Differentiating Online Products using Customer Perceptions of Sustainability

Nasreddine El Dehaibi
Mechanical Engineering,
Stanford University,
Stanford, CA 94305,
e-mail: ndehaibi@stanford.edu

Erin F. MacDonald
Assistant Professor,
Mechanical Engineering,
Stanford University,
Stanford, CA 94305
e-mail: erinmacd@stanford.edu

ABSTRACT

Customers make quick judgments when shopping online based on how they perceive product design features. These features can be visual such as material or can be descriptive like "nice gift". Relying on feature perceptions can save customers time but can also mislead them to make uninformed purchase decisions, for example, related to sustainability. In a previous study we developed a method to extract product design features perceived as sustainable from Amazon reviews, identifying that customer perceptions of product sustainability may differ from engineered sustainability. We previously crowdsourced annotations of French press reviews and used a natural language processing algorithm to extract the features. While these features may not contribute to engineered sustainability, customers identify the features as sustainable enabling them to make informed purchase decisions. In this study, we validate how our previously developed method can generalize by testing it with electric scooters and baby glass bottles. We first extracted features perceived as sustainable for both products and second, tested how participants interpret the features using a novel collage approach. Participants placed products on a set of two axes and selected features from a list. Based on our results we confirm that our proposed method is effective for identifying features perceived as sustainable, and that it can generalize for different products with limitations. We found that positively biased Amazon reviews can limit the natural language processing performance. We recommend that designers carefully select products with balanced Amazon ratings and use our method to enable customers making informed purchasing decisions.

Keywords: Customer perceptions, sustainable design, natural language processing, online reviews

1. Introduction

The growth of e-commerce has changed the way customers make purchasing decisions. With an abundance of products available, customers rely on perceptions to make quick judgments between options (Du and MacDonald, 2016). These perceptions are derived from prior experiences and available information, acting as mental shortcuts for customers to simplify decision making (MacDonald and She, 2015). For example, customers tend to judge how absorbent paper towels are based on the presence of quilted lines (Erin Faith MacDonald, Gonzalez and Papalambros, 2009). Customers can simplify their decision making based on how product features align with their perceptions.

While relying on perceptions can help customers simplify decisions, it can also mislead customers to make uninformed decisions (MacDonald and She, 2015). This is often seen with sustainable products where features perceived as sustainable may not contribute to engineered sustainability. For example, customers may perceive a stainless-steel coffee maker is more sustainable than a plastic one, but it is the energy efficiency that has the largest environmental impact. Despite this, designers tend to focus on engineered sustainability requirements while neglecting perceived sustainability (MacDonald and She, 2015). This is validated by a lack of market success for sustainable products despite market research indicating customers are willing to pay more for them (*The Sustainability Imperative: New Insights on Consumer Expectations*, 2015). Moreover, customers have grown skeptical of green marketing strategies like eco-labels (Kim and Lyon, 2015).

A robust literature exists on customer perceptions of features when purchasing products (see section 2 for an overview). MacDonald et al. identified a relationship between perceived and engineered requirements using a discrete choice analysis survey with paper towels (Erin F. MacDonald, Gonzalez and Papalambros, 2009). The results showed that customers constructed their perceptions on an as-needed basis and are not inherently found in people. Simple design features can therefore help shape perceptions and influence decision making. In a subsequent study, She and

Macdonald demonstrated how perceived sustainable features led participants to prioritize engineered sustainability concerns in a decision scenario with toasters (She and MacDonald, 2017). For example, an embossed leaf pattern on a toaster led participants to prioritize energy and shipping concerns of the product. While the embossed leaf pattern does not contribute to sustainability, it communicates information to customers that helps bridge the gap between perceived and engineered sustainability. In doing so, customers are better informed to align their intent with their purchase decisions.

Previous literature supported designing-in features based on perceptions, but a method was lacking for identifying features as perceived by the customer. Building on this, we previously developed a method to identify features perceived as sustainable from online reviews using crowdsourced annotations and natural language processing (El Dehaibi, Goodman and MacDonald, 2019) (refer to section 2.2.2 for a deeper overview). We extracted features perceived as sustainable using French presses as a case study and demonstrated a gap between engineered sustainability. In a subsequent study, we confirmed that participants identified the extracted French press features as sustainable using a novel collage activity (El-Dehaibi, Liao and MacDonald, 2021). Participants placed products on a set of axes and selected features from a list. We found that they more often selected features perceived as sustainable when evaluating product sustainability on the collage. The results validated that participants identified the French press features as sustainable even if the features may not contribute directly to engineered sustainability.

In this study, we test the generalizability of our previous findings by recreating the methods using different products and assessing the similarities in results. Our goal is to provide designers a robust method to identify product feature perceptions from online reviews so that they may differentiate their products and drive purchase decisions. The rest of the paper is organized as follows: Section 2 presents a background on the role of customer perceptions in decision making, the research propositions and hypotheses are in Section 3, Section 4 presents our method, the results and analysis are in Section 5, Section 6 presents our discussion, and we conclude our paper in Section 7.

2. Related work

As more purchases occur online, several papers have explored the changing context in which customers form perceptions, using tools like machine learning and collage activities to extract perceptions from online reviews. We provide an overview of this work in this section. In addition, we provide details on our previous studies as we build off them in this study.

2.1 Customer perceptions in online decision making

In this section we present literature on how customer perceptions shape online decision making. Wang et al. investigated the impact of online reviews embedded in product descriptions on purchasing decisions (Wang *et al.*, 2016). The authors simulated a shopping experience based on Taobao, a Chinese e-commerce website that automatically bundles online review fragments into descriptions for certain products. The authors recruited participants to explore the website while wearing an eye-tracking device and investigated how the participants interacted with pages that had and did not have online reviews in the descriptions. The results showed that product pages with online reviews in descriptions had longer fixation time on average, suggesting these descriptions aligned closer with customer perceptions. To determine the influence of purchase decisions, the authors then collected historical data from Taobao for two products, a shaving gel and an electric shaver. The data included sales, reputation, price, and whether the products had online reviews embedded in the descriptions. Using a hierarchical multiple regression model, the authors found that descriptions embedded in online reviews positively influenced purchase decisions. This finding demonstrates how perceptions of product features from online reviews can drive purchasing decisions.

Maslowska et al. studied the influence of product price and customer perceptions of reviews on online purchase decisions (Maslowska, Malthouse and Viswanathan, 2017). The authors used shopping data provided by two online retailers, one that sells unique and high-priced items while the other sells health and beauty products. There were 2.5 – 3 million observations from each retailer. For each observation the authors had access to the number of reviews for a product, the average number of stars, whether the customer clicked on the “review tab”, product price, and purchase decision. The authors used a logistic regression model with the purchase decision as a dependent variable and found that the product price plays an important role on how ratings and reviews influence purchase decision. For lower-priced products, average ratings can have a large influence with fewer reviews while for higher-priced products, more reviews are needed for the average rating to have an influence. These findings illustrate how price can influence of how customers perceive product reviews.

Helversen et al. investigated the relationship between customer age and the influence of perceptions of product attributes and reviews on purchasing decisions (von Helversen *et al.*, 2018). The authors designed three between-participant conjoint analysis surveys where they presented pairs of positively rated household products to participants. A mixture of highly positive and negative reviews was shown with a mixture of low and high ratings. The authors found that younger customers relied more on average ratings when product attributes were similar between paired choices, while older customers were quickly influenced by negative reviews. These results show the importance of factoring in age for how customers develop perceptions and make purchasing decisions.

Nysveen and Pederson explored how interactive features like content personalization and customer communities influence perceptions of customer experience on a shopping website (Nysveen and Pedersen, 2004). The authors designed six websites for two made-up companies, an airline and a restaurant. Each company had one website with email functionality only, a second website with email and personalization features, and the third website with email and customer community features. Participants interacted with the websites and then responded to a survey on the ease of use, usefulness, and attitude towards the websites. The results showed that the interactive features had a moderate influence on perceptions of customer experience and emphasize the effect of the e-commerce platform to influence customer perceptions. This study demonstrates how website content not related to the product may still influence purchasing decisions.

Li et al. investigated how return policies influence customer perceptions of products and decision making depending on the market stage of a business (Li *et al.*, 2019). The authors propose a multi-stage hidden Markov model which models randomly changing systems. They test it on 50,000 purchase records spanning three years from Taobao including returns, discounts, and total sales. The results showed that promotions and return policies had varying influence of repurchase behavior across different stages of market growth. For example, a company in the growth stage could benefit from flexible return policies and frequent promotion while a company in the introduction stage would not. Therefore, role of return policies on customer perceptions is crucial depending on the market stage of the seller.

2.2 Extracting customer perceptions from e-commerce websites

Literature discussed thus far looks at factors like age, reviews, price, and website features to influence customer perceptions and decision making, but it neglects a key component which is how customers perceive features of the product itself. This presents an opportunity for designers to determine how certain product features align with customer perceptions and can drive purchasing decisions. The development of e-commerce and social media provides a wealth of information that designers can tap in to online. In this section we present literature on methods to extract customer perceptions from online content including machine learning and collage approaches.

2.2.1 Machine Learning Approaches

Zhang et al. study the influence of self-descriptions of Airbnb host on customer trust and how they can influence booking behaviors (Zhang, Yan and Zhang, 2020). The authors annotated 4179 host descriptions from Airbnb listings based on perceived trustworthiness of the hosts. The authors then used a deep learning model to predict perceived host trustworthiness for 75,000 host descriptions. Using this data, they extracted textual features including readability, sentiment intensity, and semantic content. Semantic content included personal information about the host such as family and work. From regression analyses the authors showed that readability of the self-description had a positive influence on perceived trust, while semantic intensity had a U-shaped relationship with trust. Moreover, semantic content had a positive influence on trust if the content was related to sociability. When looking at Airbnb booking decisions, the results showed that higher perceived trust of host led to more booking decisions. These results point to the importance of language when describing products to drive purchasing decisions.

Liu et al. use natural language processing to identify product competitive advantages from social media content (Liu, Jiang and Zhao, 2019). The authors collected reviews of a Volkswagen Passat from two Chinese auto websites and identified competitors from the reviews based on comparative language. An example review could include: “the sound system in the Passat sounds better than the one in my old Camry”. The authors first preprocessed the reviews by removing stop words and performing named-entity recognition. They then performed a sentiment analysis using logistic regression and a domain specific lexicon to assess customer sentiments towards features of the Volkswagen Passat compared to its competitors. With their method the authors demonstrated how customer perceptions can drive competitor analyses to inform design decisions for next iteration products.

2.2.2 Combining Machine Learning with Collage Approaches

Previous literature identified methods to extract customer perceptions from online content but did not identify specific product features that can inform design decisions. Moreover, previous literature does not test how users interpret product features in terms of liking and evaluating products. This gap is particularly crucial for sustainable products where designers often focus on engineered requirements while neglecting perceived requirements. Motivated by this gap, we previously conducted two studies where we first developed a natural language processing approach to extract product features perceived as sustainable from online reviews (El Dehaibi, Goodman and MacDonald, 2019), and second developed a novel collage approach to test those features in terms of how users like and evaluate products (El-Dehaibi, Liao and MacDonald, 2021). We previously selected French presses as a case study. In this study we aim to validate the generalizability of our previous approaches by testing them with different products. We provide details on our previous work below since we build heavily off them for this study.

In the first study, we extracted features perceived as sustainable using crowdsourced annotations of online reviews and a natural language processing algorithm (El Dehaibi, Goodman and MacDonald, 2019). The approach combined research from identifying sustainability perceptions, rating design ideas, and natural language processing (Fig. 1) and is outlined in four steps (Fig. 2). We collected product reviews for a target product type from Amazon, annotated the reviews using a crowdsourcing platform based on criteria related to the perceptions, modeled the reviews and annotations using natural language processing, and extracted features perceived as sustainable from the model.

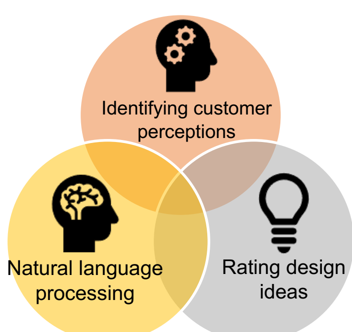

Figure 1: Interdisciplinary approach

Collect	Collect product reviews from Amazon
Annotate	Annotate reviews via crowdsourcing
Model	Model reviews and annotations using NLP
Identify	Identify perceived sustainable product features

Figure 2: Extracting customer perceptions method flow

We tested the method with 1474 reviews of French presses from Amazon and recruited 900 respondents from Amazon Mechanical Turk to annotate the reviews based on the three sustainability pillars: social, environmental, and economic. For a product to be truly sustainable it needs to account for each pillar. We previously conducted a pilot study that showed participants had more clarity when focusing on pillar, so we studied each one individually. For this study we assigned respondents to one of three versions of the survey to focus on one sustainability pillar and trained them on their assigned pillar using basic guidelines. Respondents then highlighted parts of reviews relevant to their pillar and rated the emotions in their highlights.

We modeled the annotations using a logistic classifier for each sustainability pillar and extracted French press features perceived as sustainable based on the beta parameters of the model. The precision, recall, and F1 scores for the model are shown in Table 1 (see section 4.1.3 for more on these metrics). With scores ranging from 0.83 to 0.95 for positive sentiment and 0.42 to 0.72 for negative sentiment, we were confident in the model performance while noting possibilities of noise for negative sentiment predictions.

Table 1: Precision, recall and F1 scores for French press features perceived as sustainable

	Social sustainability			**Environmental sustainability**			**Economic sustainability**		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Positive Sentiment	0.85	0.87	0.86	0.83	0.86	0.85	0.85	0.95	0.90
Negative Sentiment	0.70	0.66	0.68	0.51	0.72	0.66	0.53	0.42	0.72

We identified salient positive features based on the largest positive beta parameters in the model and identified salient negative features based on the largest negative beta parameters in the model. We then identified engineered sustainability requirements of a French press using a life cycle analysis and found that crucial engineered sustainability requirements like energy and water consumption were not salient perceived sustainable features. This demonstrated the gap between engineered and perceived sustainability and the importance for designers to account for both when creating sustainable products.

In the second study, we developed a novel collage approach to test the extracted features perceived as sustainable with users in terms of how they like products and evaluate sustainability (El-Dehaibi, Liao and MacDonald, 2021). Using the collage, we identified the relationship between features perceived as sustainable and user emotions in an engaging way without drawing attention to the features. We created a webapp collage activity with two axes: sustainability on the vertical axis (customized to one of the three pillars depending on the version of the collage) and likeability on the horizontal axis. An example of an environmental sustainability collage activity is shown in Fig. 3. We recruited 1200 participants from Amazon Mechanical Turk, assigned them to one of three sustainability pillars, and asked them to evaluate six French press products on a collage. They placed images on the collage according to the two axes and selected product features from a dropdown list. The list included features perceived as sustainable that we extracted previously as well as features "not perceived as sustainable" that we identified for the collage study.

Based on participants' placement of the products and selection of the features on the collage, we showed that they actively chose features perceived as sustainable (as outlined by the method explained in Section 4.2.2) for products that they placed higher on the sustainability axis, indicating that these features stood out to them as sustainable despite not contributing to engineered sustainability. We also found a low correlation between perceived sustainability and likeability, validating that the collage is an effective approach for measuring these two attributes separately. The results validated our previously extracted features perceived as sustainable as well as validated the collage tool as an effective tool to test features perceived as sustainable with users.

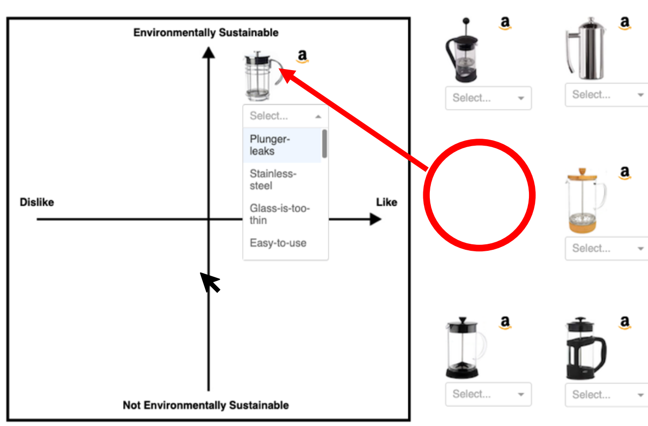

Figure 3: Dragging and dropping products on collage and selecting at least one phrase to describe each product

A limitation to the findings is that the approach has been tested on French presses only. We aim to address this limitation in this study by testing the generalizability of our approach across different product types.

3. Research Proposition and Hypotheses

This work aims to validate the generalizability of our previously developed approaches for extracting and testing features perceived as sustainable from online reviews. To validate our approaches, we extracted features perceived as sustainable for different products using annotations and a logistic classifier, and then used a collage tool to test the features with users in terms of how they like and evaluate the products. We asked participants to place products along the two axes of the collage, sustainability and likeability, and to label the products using a list of the extracted features from the logistic classifier. In our previous studies we tested and confirmed the following propositions and hypotheses using French press products as a case study (Table 2).

Table 2: Propositions and hypotheses from our previous studies

Proposition & Hypotheses	Status
P1: Phrases in product reviews perceived as sustainable contain semantic and syntactic characteristics that can be modeled (El Dehaibi, Goodman and MacDonald, 2019)	Confirmed with French presses
P2: Designing-in perceptions can help customers create an alignment between perceived sustainability and sustainable products. Based on this, we propose that customers will evaluate perceived sustainable features as being sustainable (El-Dehaibi, Liao and MacDonald, 2021) H1: participants evaluating product sustainability on a collage will select features perceived as sustainable for products that they place higher on the "sustainability" axis of the collage (El-Dehaibi, Liao and MacDonald, 2021)	Confirmed with French presses
P3: Customers tend to like products that create cognitive alignment for them, and perceptions can help them achieve that. We therefore propose that perceptions of product sustainability contribute to how much customers like a sustainable product. (El-Dehaibi, Liao and MacDonald, 2021) H2: A statistically significant relationship exists between the placement of a product on the "sustainability" axis of the collage, and the "like axis of the collage (El-Dehaibi, Liao and MacDonald, 2021)	Confirmed with French presses

Our goal for this study is to test if the same propositions and hypotheses hold when tested with multiple product types.

4. Methods

The method in this paper is based on our work from two previous papers where we used a French press as the focal product (El Dehaibi, Goodman and MacDonald, 2019; El-Dehaibi, Liao and MacDonald, 2021). In this study we validate how the method generalizes when applied to different product types. Figure 4 provides an overview of the method. Specific steps are explained below.

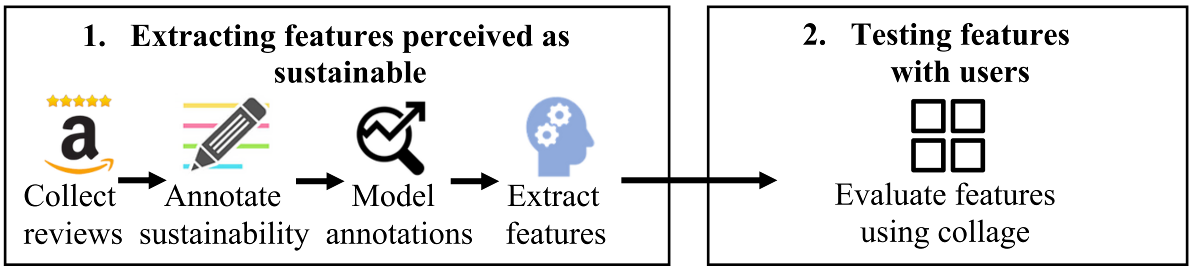

Figure 4: Method overview

4.1 Extracting Features Perceived as Sustainable from Online Reviews

The methods outlined in this section aim to test proposition 1. We extracted features perceived as sustainable for electric scooters and baby glass bottles using steps in Fig. 2.

4.1.1 Collecting Reviews

We selected electric scooters and baby glass bottles as the focal products for this study because they (1) are different in design and function from the original French press product, (2) have varying aesthetic design features available, (3) regularly receive several hundred reviews on Amazon, and (4) likely have sustainability-related concerns for customers. We wanted to select products that are different from a French press to effectively evaluate how the method can be generalized. Products like kettles or other coffee makers would have been too similar. We also selected products that have a large variety of features that reviewers can mention (paper plates, for example, would have been too simple) as well as products that have large amounts of reviews available for us to collect (there were limited reviews for electric bicycles, for example). Finally, since we are interested in extracting features that are perceived as sustainable, we wanted products where sustainability concerns are likely to be prominent in the reviews.

We scraped 1500 Amazon reviews from four electric scooters and 1444 Amazon reviews from eight baby glass bottles. We selected the four products and eight products, respectively, from Amazon so that they (1) have varying aesthetic features from one another, (2) are in a similar price range, (3) have less than 500 reviews per product, and (4) have at least 80% estimated authentic reviews according to a data analytics tool (fakespot.com) for each product. We selected products that have varying features to better test different features with users. Moreover, we selected products in a similar price range so that their quality and capabilities are like each other. Each product had less than 500 reviews so that we have a variety of products instead of one product dominating the reviews. The motivation was to have a variety of features to test. Finally, we selected products that are estimated to have a high number of authentic reviews so that we collect real customer opinions and perceptions. All the reviews scraped came from the United States to limit the number of reviews written in a foreign language. Moreover, we filtered reviews that were less than 10 words as they tended to be generic, for example, “this is a great product, I highly recommend it”.

4.1.2 Annotating Reviews

We recruited respondents from MTurk (referred to as “annotators”, see section 4.1.2.3) to annotate the scraped reviews based on sustainability criteria. These annotations are then fed into a logistic classifier to extract features perceived as sustainable.

4.1.2.1 Survey Design

To guide the annotators, we designed a survey that trains and tests them on the sustainability criteria, and then shows them a set of 15 random reviews to annotate before answering a set of demographic questions. In total we had six different versions of the survey to account for the three sustainability aspects (social, environmental, and economic) and for each of the two product types (electric scooters and baby glass bottles), shown in Fig. 5.

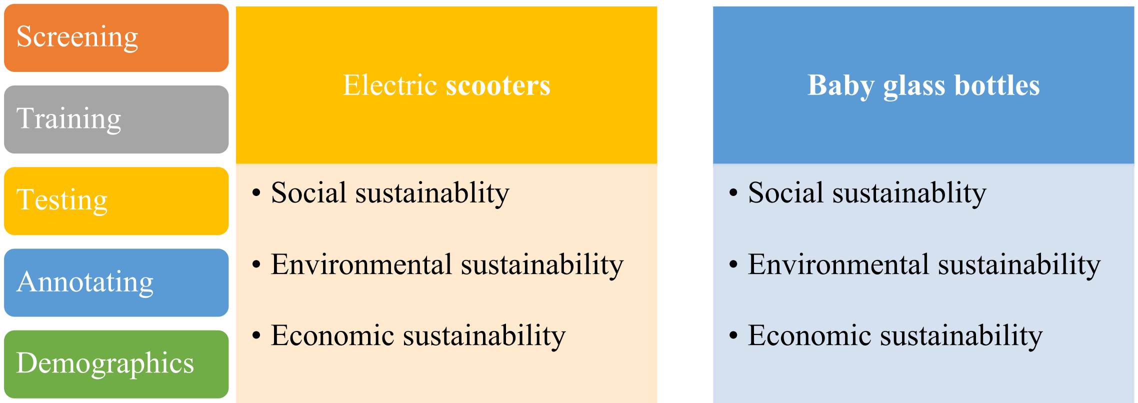

Figure 5: Three annotation survey versions per product

In the training portion of the survey, we displayed sustainability criteria to the annotators and showed them examples of annotated reviews according to their assigned sustainability aspect. We then tested them to confirm that they understood the training. After passing the test annotators began annotating the 15 reviews (see section 4.1.2.2) according to one of the sustainability aspects criteria.

4.1.2.2 Data Collection

To annotate the reviews scraped in 4.1.1, we stored the reviews on a server so that they can be pulled live during the survey. We used a biased-random algorithm for selecting the reviews from the server to ensure that each review was presented to three different annotators. Each participant saw 15 reviews in total, one at a time. For each review, we asked annotators to highlight up to five parts of the review that they found relevant to their assigned sustainability criteria. If they highlighted parts of the review as relevant, we asked annotators to type-in the specific feature that is mentioned in the highlighted part, and to label the emotion on a 5-point Likert scale ranging from negative to positive.

4.1.2.3 Annotators

We recruited a total of 1800 annotators from Amazon Mechanical Turk to complete one of the six surveys. 900 annotators annotated reviews for electric scooters and 900 annotators annotated reviews for baby glass bottles. Within each product type, 300 annotators annotated the reviews for each of the three sustainability aspects. On average annotators took 20 minutes to complete their survey and we compensated them $4 each. We recruited Amazon Mechanical Turk respondents instead of in-person annotators as it allowed us to collect many annotations in a short amount of time. Moreover, this online approach is timely due to the COVID-19 pandemic which is when we conducted this experiment.

To ensure high quality responses, we required respondents to have at least a 97% approval rating and to be based in the US. We set these requirements in the MTurk platform and confirmed them using screening questions in the survey. Moreover, we included a simple checkpoint question to gauge if annotators are paying attention. If annotators completed the survey faster than the average time by at least one standard deviation and incorrectly answered the checkpoint question, we assumed their response was low quality and did not include it in the analysis. These criteria are like what was used in our previous work (El Dehaibi, Goodman and MacDonald, 2019). Based on these criteria we approved 1702 out of the 1800 responses.

4.1.3 Machine Learning Model

We used a binary logistic classification model to extract features perceived as sustainable from the annotated reviews. We chose a logistic classification model because it has proven to be highly effective for natural language processing applications while remaining interpretable in terms of its beta parameters (James *et al.*, 2013). This enabled us to extract salient product features directly from the classifier, unlike deep learning approaches.

The logistic classification model is represented in Eq. 1. Our input "X" consists of the (1) phrases highlighted as "relevant" to sustainability and (2) the product features typed in by the annotators, while the output "Y" is binary representing the emotion in each phrase. We binarized the output Y such that 0 represented negative or neutral emotion while 1 represented positive energy. We opted for a binary output instead of a multi-class model due to the limited explanatory power from our dataset for a multi-class model.

$$p(Y = 1|X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} \tag{1}$$

The beta fitting parameters are optimized with a maximum likelihood shown in Eq. 2.

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p(x_i)^{y_i} \left(1 - p(x_i)\right)^{1-y_i} \tag{2}$$

We first pre-processed the inputs to remove potential noise in the model. This included lowercasing all text, stemming words, removing stop words such as "and" or "is", and removing punctuation. We then processed these inputs so that they can be quantified in a matrix and fed into a classifier. For the highlighted phrases we used bag of words, bigrams, and trigrams. For the typed-in features we summarized them into a set number of "topics" using Latent Dirichlet Allocation (LDA) and then hot-encoded them for each highlighted phrase. LDA is a topic modeling approach that identifies a set number of "topics" from text based on the model shown in Eq. 3:

$$P(t_i|d) = \sum_{j=1}^{|Z|} P(t_i|z_i = j) * P(z_i = j|d) \tag{3}$$

where z_i represents a product feature, d represents a review from a collection reviews D, and |Z| is a pre-set total number of product features.

We split the data into a 70% training and 30% test set and implemented the logistic classifier model in Python using the Scikit package. We used five-fold cross validation on the training set and penalty terms to shrink fitting parameters based on Ridge regularization to address potential overfitting from high dimensionality. As an external validity check on the models, we used precision, recall, and F1 as these are often more robust measures than accuracy (James *et al.*, 2013).

4.1.4 Extracting Perceptions

We extracted salient product features perceived as sustainable from the machine learning model that drove positive and negative sentiment. The magnitude of the beta parameters indicates the influence of a given feature on the model. These features come from reviews of multiple products to capture a variety of different features. After extracting the product features perceived as sustainable, we conducted a collage experiment to validate that participants identified these features as sustainable.

4.2 Testing perceived features extracted from online reviews with participants

The method outlined in this section aims to test hypotheses 1 and 2. We recruited 300 additional respondents (referred to as "participants" for this portion of the method) from MTurk to evaluate the products and features using the collage activity explained in Section 2.2.2. Based on the placement of products on the collage and the location of selected features we determined if participants identified the extracted features as sustainable. To guide participants through the activity, we designed three versions of a survey (accounting for each of the sustainability pillars) and assigned participants to one of the versions (see Fig. 6). Like Section 4.1.2.1, we asked participants to evaluate products for only one of the sustainability pillars based on pilot studies that demonstrated this led to more usable responses.

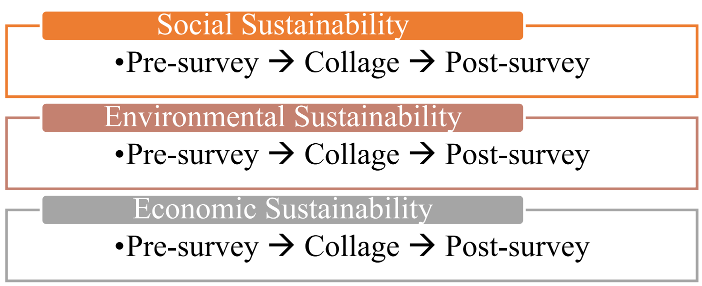

Figure 6: Three collage activity versions

4.2.1 Pre-survey

In the pre-survey we familiarized participants with their assigned sustainability criteria as well as the products that they will evaluate (Table 3). We selected the products according to the criteria explained in Section 4.1.1.

Table 3: Products in Collage Activity

						
Product Name	Gotrax	Razor E300S	Mongoose	Razor EcoSmart	Segway	SKRT

During the pre-survey we trained participants on their assigned sustainability pillar and led them to Amazon pages of the products in Table 3 to familiarize themselves before evaluating. They had to open each of the Amazon pages and spend a certain amount of time on them to proceed with the activity. We required this to ensure that participants understood the characteristics of each product before evaluating them. Participants could also access the Amazon pages later when evaluating the products on the collage.

4.2.2 Collage Activity

After completing the pre-survey participants accessed a link to a collage webapp using the same interface shown in Fig. 3. Products were presented on the right side with buttons to access their Amazon pages for a refresher about each product if needed. On the left was a button to access the sustainability criteria for a given pillar from the pre-survey. The collage consisted of two axes ranging from “Not Sustainable” to “Sustainable” vertically and “Dislike” to “Like” horizontally. The sustainability axis was named social, environmental, or economic depending on the version. Participants dragged and dropped each product on the collage and then selected features from a dropdown menu for each product as shown in Fig. 3. We test hypothesis 1 based on the placement of the features on the collage, and tested hypothesis 2 based on the placement of products on the collage.

In the dropdown menu we provided the features we extracted from the machine learning models in Section 4.1.4. Each collage version included a list of 20 features that participants could select from. Ten of these features were the most positive salient features from the machine learning model and the other ten were the most negative salient features from the machine learning model. These features are derived from reviews of multiple products but are specific to a certain sustainability pillar. Each sustainability version of the collage had its own set of 20 features. The order of the features was randomized between participants. We present these features in Section 5 as part of the results.

To further test hypothesis 1, we conducted a fourth collage activity for environmental sustainability but with a more challenging set of features. These features included ten positive features perceived as sustainable from the original environmental collage activity and ten new features not perceived as sustainable. For the features not perceived as sustainable, we derived phrases from the unhighlighted parts of the annotated reviews collected using the method in Section 4.1.2.2. Since they were unhighlighted, we assumed that they were not perceived as sustainable. We combined the unhighlighted parts and identified ten random adjectives and ten nouns using named-entity recognition, and then randomly combined them to create descriptive features. The motivation was to identify these features in a fully automated way and avoid potential bias. This set of features is more challenging because the sentiments are closer together, and we cannot be sure if the perceptions are indeed not perceived as sustainable. The derived features are presented in Section 5. For this collage activity, we recruited an additional 100 participants using the procedures outlined in Section 4.2.4.

After evaluating each product, participants rated each feature that they selected based on how relevant to sustainability they think it is using a 5-point Likert scale. We included this in the activity so that we can filter out from the participants' selection the features that they did not select due to sustainability. After rating the features participants completed a post-survey.

4.2.3 Post-survey

In the post-survey we asked participants to rate on a 5-point Likert scale the quality of images, product descriptions, and the overall product quality for each of the products they evaluated on the collage. Finally, we asked participants basic demographic questions.

4.2.4 Participants

We recruited 300 participants from MTurk to complete the collage activity using the same recruiting criteria as in Section 4.1.2.3, in addition to requiring participants to use a screen size of 10 inches or larger. This was to ensure compatibility with the collage interface. Participants self-reported their screen size in the screening question. They completed their task in 21 minutes on average and we compensated them $5 each. We did not analyze responses if they fell under one of the following: (1) participants completed the survey faster than the average time by at least one standard deviation or (2) they incorrectly answered a simple checkpoint question designed to gauge attention. Based on these criteria we analyzed 224 responses out of 300.

5. Results

We first present the results that test the generalizability of proposition 1 related to the features extracted from online reviews. Second, we present the results that test the generalizability of hypotheses 1 and 2 based on the placement of products and extracted features on the collage.

5.1 Features Perceived as Sustainable

This section presents the extracted features perceived as sustainable for electric scooters and baby glass bottles and tests the generalizability of proposition 1: Phrases in product reviews perceived as sustainable contain semantic and syntactic characteristics that can be modeled.

5.1.1 Electric scooters model evaluation

The precision, recall, and F1 scores for each of the sustainability pillars for electric scooters are shown in Tables 4.

Table 4: Precision, recall and F1 scores for electric scooter features perceived as sustainable

	Social			**Environmental**			**Economic**		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Positive	0.85	0.80	0.82	0.86	0.85	0.85	0.80	0.97	0.88
Negative	0.33	0.41	0.36	0.51	0.52	0.51	0.60	0.16	0.25

The positive sentiment ranged between 0.80 and 0.97 across all three metrics and all three sustainability pillars, indicating that we can have high confidence in the quality of the model output for positive sentiment. The negative sentiment had a lower range however between 0.16 and 0.52, indicating that we are likely to see some level of noise in the model output and is important to keep in mind while analyzing the most salient negative features. These metrics are like findings from our previous study with French presses (see table 1) which support the generalizability of proposition 1, although the negative sentiment scores fared worse with electric scooters here suggesting that there is a greater imbalance between positive and negative highlighted reviews.

5.1.2 Electric scooters model output

Figures 7-9 show the most salient 20 positive and negative features of electric scooters based on the parameters of the logistic classifier for social, environmental, and economic pillars, respectively. These features are derived from reviews of multiple products and have the largest positive and negative parameters in the model, indicating that they are the most salient features that annotators identified as sustainable. Note that the features shown in this graph are stemmed as part of preprocessing, which is why words like “warranty” appear as “warranti”. The models were able to output specific product features perceived as sustainable for electric scooters, therefore supporting the generalizability of proposition 1.

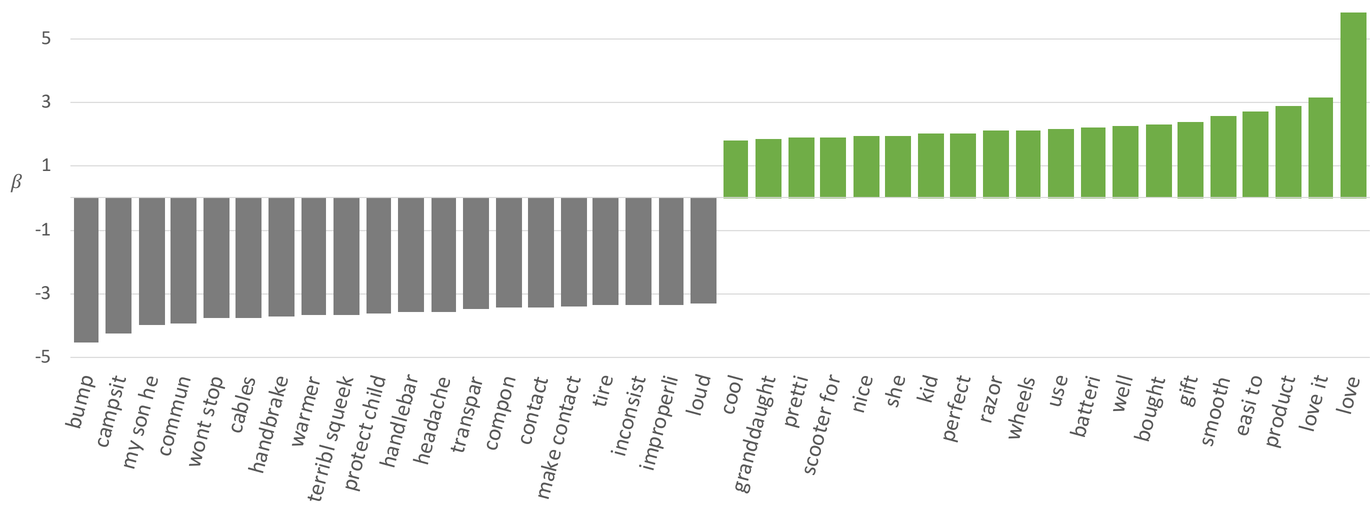

Figure 7: Most salient 20 positive and negative features of electric scooters perceived as socially sustainable

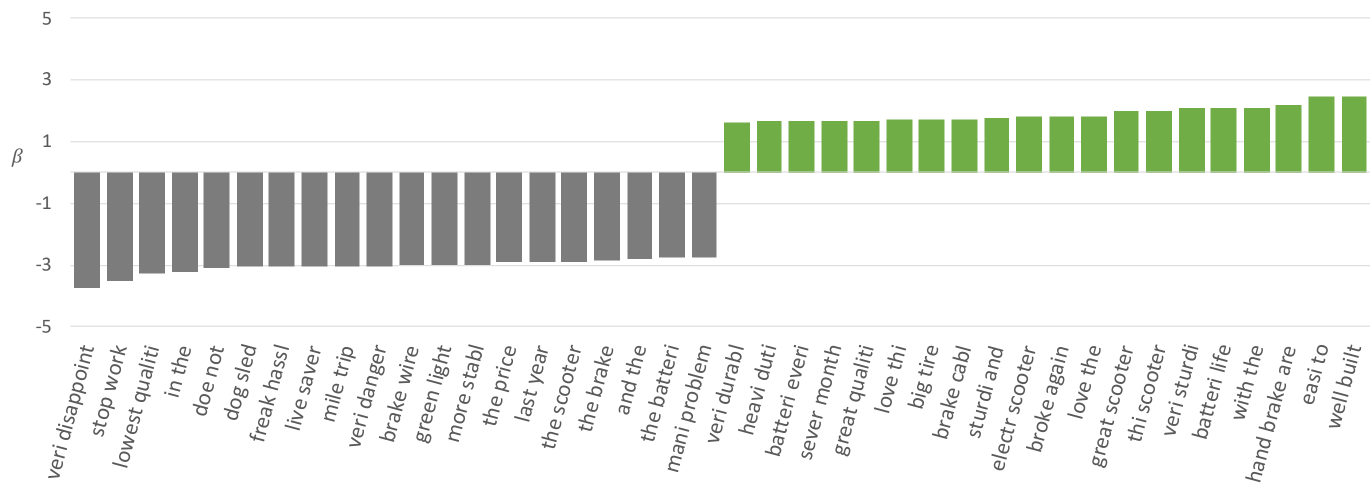

Figure 8: Most salient 20 positive and negative features of electric scooters perceived as environmentally sustainable

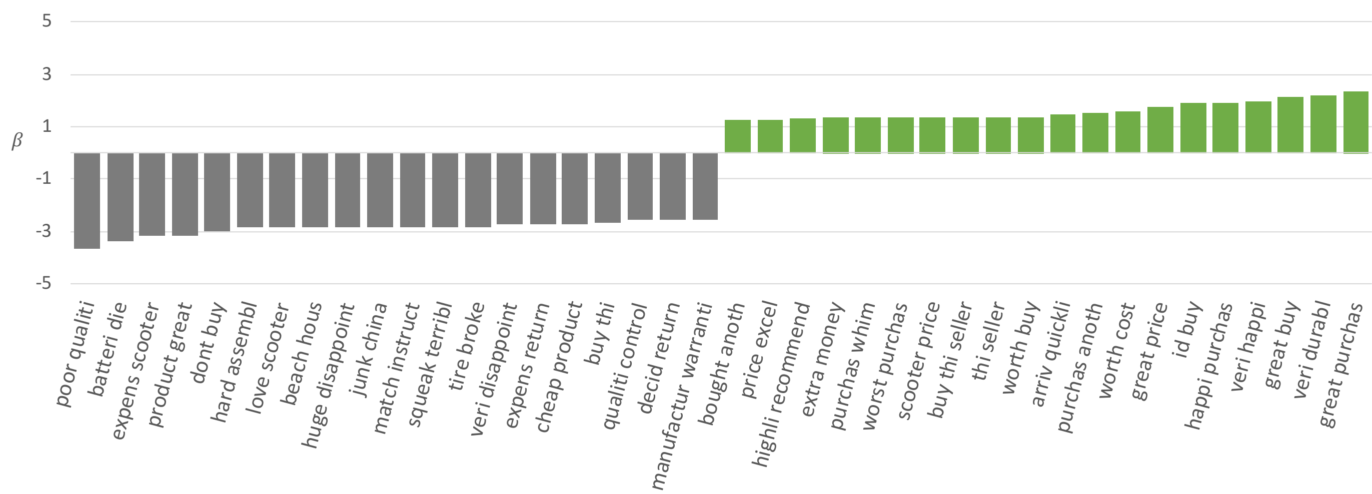

Figure 9: Most salient 20 positive and negative features of electric scooters perceived as sustainable for economic sustainability

5.1.3 Baby glass bottles model evaluation

The precision, recall, and F1 scores for each of the sustainability pillars for baby glass bottles are shown in Tables 5.

Table 5: Precision, recall and F1 scores for baby glass bottle features perceived as sustainable

	Social			Environmental			Economic		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Positive	0.87	0.86	0.86	0.84	0.94	0.89	0.87	0.99	0.93
Negative	0.28	0.29	0.28	0.36	0.16	0.22	0.47	0.06	0.11

Like our previous findings (Table 1), scores for positive sentiment are high ranging from 0.84 to 0.99. Scores for negative sentiment are exceptionally lower, ranging from 0.06 to 0.29. This suggests that there may be considerable noise in the model output. This emphasizes the importance of data balance for using this approach to extract features. Therefore, while proposition 1 may generalize for different products there are limitations in terms of selecting products with balanced reviews.

5.1.4 Baby glass bottles model output

Figures 10-12 show the most salient 20 positive and negative features of baby glass bottles based on the parameters of the logistic classifier for social, environmental, and economic pillars, respectively. There is less consistency in the extracted features for baby glass bottles. For example, many of the top negative features contain little meaning or are unintuitive, such as “bit” for social sustainability or “pretty durabl” for environmental sustainability. These are likely due to the low metrics identified in Table 5. Therefore, while proposition 1 generalized with electric scooters, it could not generalize with baby glass bottles due to the severe imbalance in product review sentiments.

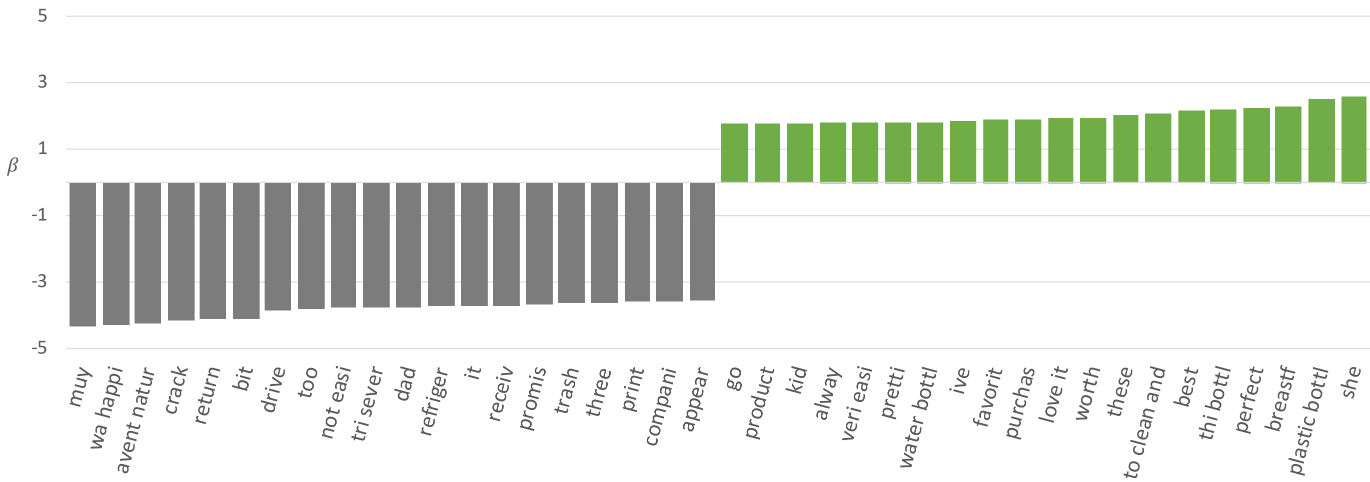

Figure 10: Most salient 20 positive and negative features of baby glass bottles perceived as socially sustainable

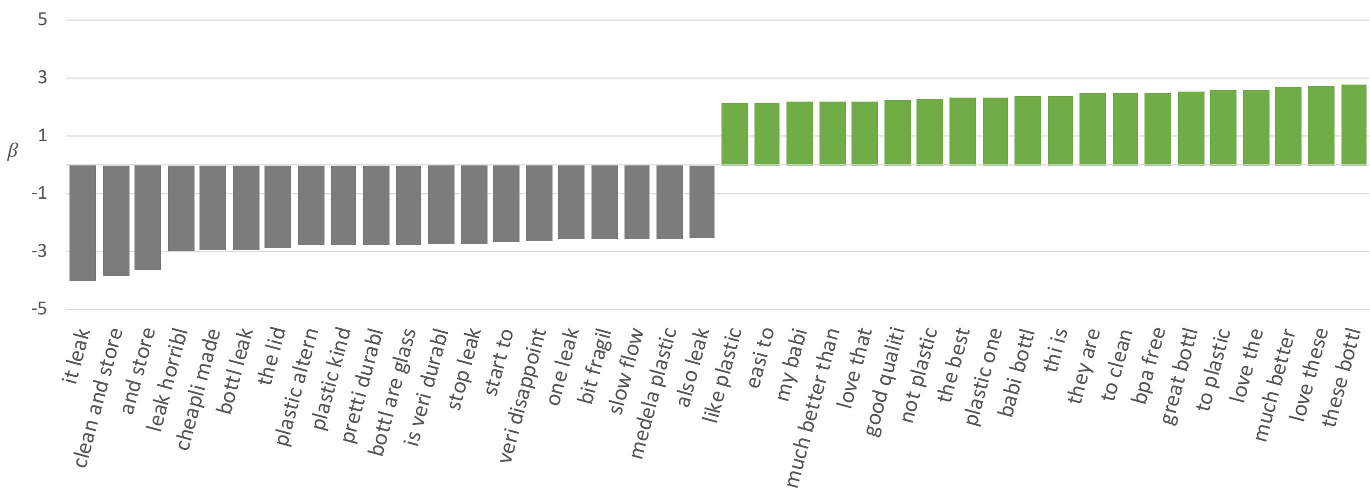

Figure 11: Most salient 20 positive and negative features of baby glass bottles perceived as environmentally sustainable

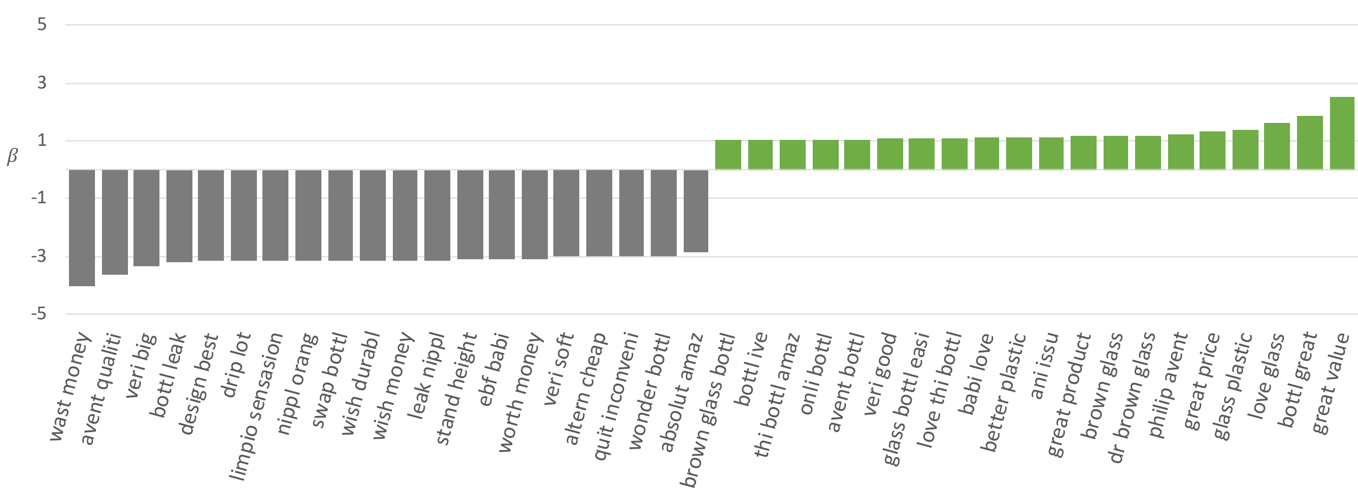

Figure 12: Most salient 20 positive and negative features of baby glass bottles perceived as economically sustainable

Based on the findings from Section 5.1.3 and the low model evaluation scores for baby glass bottles, we opted to conduct the collage activity using the features extracted for the electric scooters only.

5.2 Collage Results

This section is split into two parts, first we analyze the location of electric scooter features on the collage which test hypothesis 1 and second, we analyze the placement of the products which test hypothesis 2. We excluded 294 datapoints for products that were not moved from their starting location (starting locations are outside the collage boundaries, see Fig. 3) from of a total of 1834 recorded datapoints.

5.2.1 Feature Analysis

The analysis below tests the generalizability of hypothesis 1: participants evaluating product sustainability on a collage will select features perceived as sustainable for products that they place higher on the "sustainability" axis of the collage.

5.2.1.1 Positive and Negative Features Perceived as Sustainable

Based on Figs. 10-12, we identified 10 positive and 10 negative features to provide to participants during the collage activity for each of the sustainability pillars. These features are shown in Tables 6 and 7.

Table 6: Positive perceptions of electric scooter sustainability

Social Aspects	**Environmental Aspects**	**Economic Aspects**
Love it	Well built	Great purchase
Easy to use	Easy to use	Very durable
Great gift	Electric-powered	Arrived quickly
Perfect for kids	Long battery range	Want more than one
Smooth ride	Very sturdy	Comprehensive warranty
Looks pretty	Heavy duty	Happy purchase
Looks cool	Very durable	Excellent price
Want this for my child	Quick charge	Highly recommend
Life saver	Big tires	Buy from this seller
Stable ride	Love this	Very durable

Table 7: Negative perceptions of electric scooter sustainability

Social Aspects	Environmental Aspects	Economic Aspects
Loud motor	Very disappointed	Decided to return
Inconsistent power	Stopped working	Useless warranty coverage
Terrible squeak	Difficult to assemble	Too expensive
Very dangerous	Battery died	Won't buy this
Not stable	Poor battery range	Huge disappointment
Difficult to use handbrake	Many problems	Expensive return
Poor ride quality	Brakes failed	Cheap product
Brake cables get tangled	Long charge time	Needs quality control
Tire broke	Low quality	Long wait time
Pure headache	Battery disposal	Piece of junk

Table 8 shows a summary of the features selected during the collage activity.

Table 8: Summary of features selected in collage

	Social		Environmental		Economic		Combined	
	Positive Features	Negative Features	Positive Features	Negative Features	Positive Features	Negative Features	Positive Features	Negative Features
# participants	96		93		97		286	
Observations	355	222	384	154	371	242	1110	618
Avg. features per participant	3.70	2.31	4.13	1.66	3.82	2.49	3.88	2.16
Avg. features per product	59.17	37.00	64.00	25.67	61.83	40.33	185.00	103.00
Most common feature selectd	Great gift	Poor ride quality	Electric powered	Low quality	Excellent price	Too expensive	Electric powered	Poor ride quality

The most selected positive feature across the three sustainability criteria was "electric powered" while the most common negative feature was "poor ride quality". Figs. 13-16 show the average placement of features on the collage.

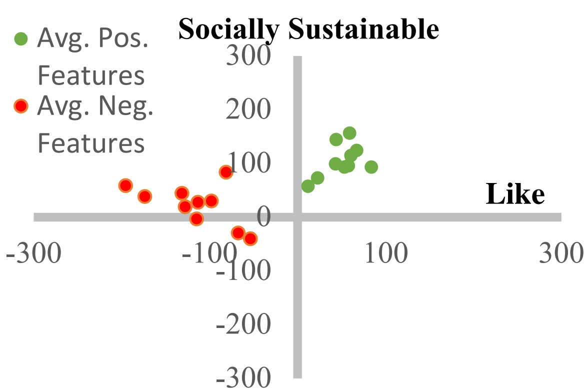

Figure 13: Avg. placement of positive and negative electric scooter features perceived as socially sustainable

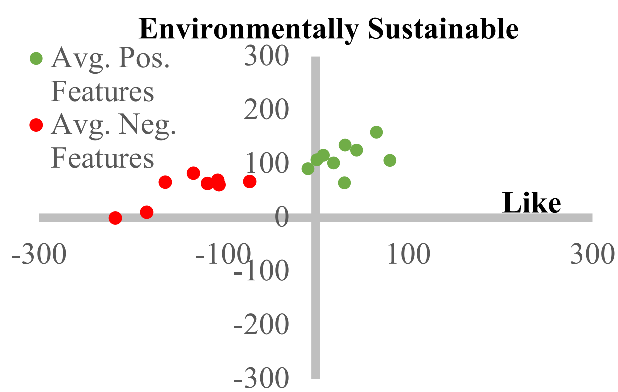

Figure 14: Avg. placement of positive and negative electric scooter features perceived as environmentally sustainable

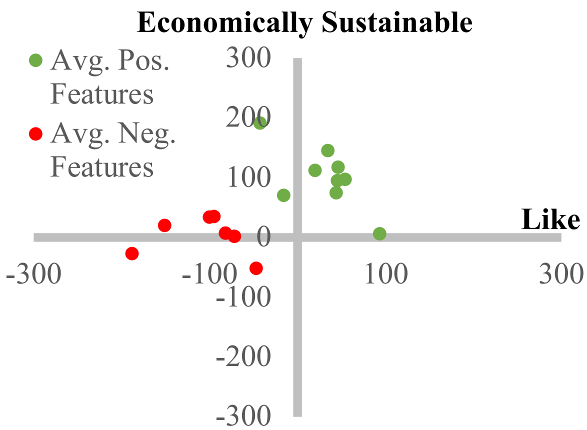

Figure 15: Avg. placement of positive and negative electric scooter features perceived as economically sustainable

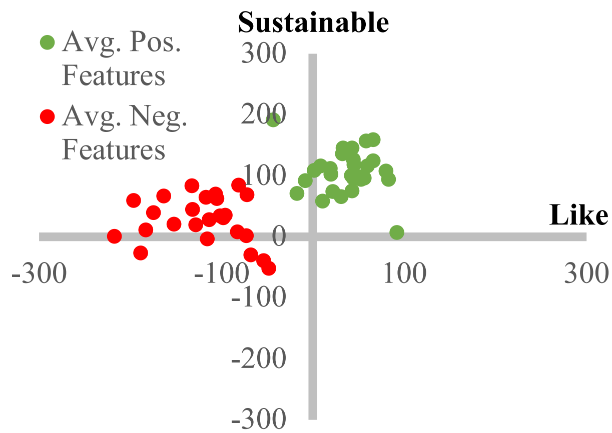

Figure 16: Average placement of positive and negative electric scooter features perceived as sustainable for all criteria

The figures show distinct clusters between the positive and negative features, which supports the generalizability of hypothesis 1. We performed a t-test on the y-coordinates between positive and negative clusters to determine if they are statistically different along the sustainability axis for each of the sustainability pillars (Table 9).

Table 9: Two-sample t-test between positive and negative features perceived as sustainable
*: significant at p = 0.05, **: significant at p = 0.01, ***: significant at p = 0.001

	Social		Environmental		Economic		Combined	
	Positive features	Negative features	Positive features	Negative features	Positive features	Negative features	Positive features	Negative features
Mean Y	103	21	118	56	106	7	110	23.9
Observation	246	137	278	88	209	140	733	365
P(T<=t)	<0.001***		0.004**		<0.001***		<0.001***	
t Critical	1.97		1.98		1.97		1.96	

Like our previous findings with the French press, there was a significant difference along the vertical axis across all sustainability aspects which supports the generalizability of hypothesis 1.

For a more rigorous test that considers repeated measures, we performed a multivariate analysis (MANOVA) using the "x" and "y" coordinates from the collage as dependent variables, and the rest of available information as independent variables (Table 10). We chose the Pillai criterion for its robustness when linearity assumptions are not met (HD Delaney, 2003). Across all sustainability criteria the features were statistically significant, like our findings with features extracted for French presses. Thus, our findings fail to reject hypothesis 1 for electric scooters and validates the generalizability of the hypothesis.

Table 10: MANOVA output with positive and negative features perceived as sustainable
*: significant at p = 0.05, **: significant at p = 0.01, ***: significant at p = 0.001

	Social			Environmental			Economic			Combined		
	Pillai	~F	**Pr(>F)**	Pillai	~F	**Pr(>F)**	Pillai	~F	**Pr(>F)**	Pillai	~F	**Pr(>F)**
Product	0.14	5.35	$<0.001^{***}$	0.21	8.51	$<0.001^{***}$	0.23	8.36	$<0.001^{***}$	0.12	14.0	$<0.001^{***}$
Criteria	-	-	-	-	-	-	-	-	-	0.02	4.79	$<0.001^{***}$
FeatureType	0.29	73.2	$<0.001^{***}$	0.21	48.0	$<0.001^{***}$	0.19	37.5	$<0.001^{***}$	0.24	171	$<0.001^{***}$

5.2.1.2 Positive Features Perceived as Sustainable and Features not Perceived as Sustainable

The features not perceived as sustainable that we derived from the unhighlighted parts of the Amazon reviews are shown in Table 13. Figure 17 shows the placement of the new set of features.

Table 11: Phrases not containing perceptions of electric scooter sustainability

Great assembly	Cheap basket	Awesome brake	Small fit	Front product
Significant cables	Easy picture	Typical work	Good brand	Extra torque

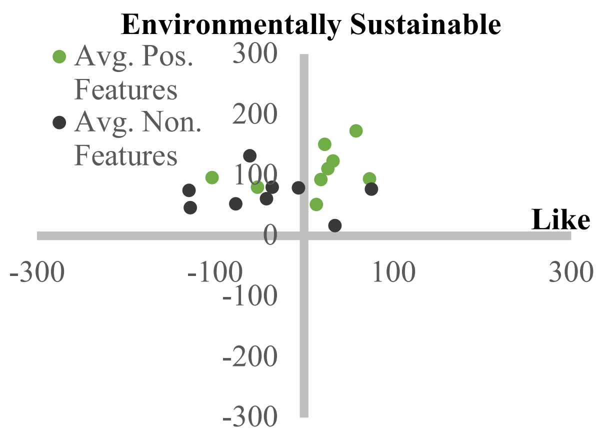

Figure 17: Average placement of positive features perceived as sustainable and features not related to sustainability

A t-test using the y-coordinates of the two sets of features is shown in Table 14.

Table 12: Two-sample t-test between positive features perceived as environmentally sustainable and features not related to sustainability
*: significant at p = 0.05, **: significant at p = 0.01, ***: significant at p = 0.001

	Positive Features	**Features not related to sustainability**
Mean	109	71
Observations	206	182
Number of participants	72	
Avg. features per participant	2.86	2.52
Avg. features per product	34.33	30.33
P(T<=t) two-tail	0.026^{*}	
t Critical two-tail	1.96	

The t-test shows a statistically significant difference, supporting the generalizability of hypothesis 1. A MANOVA analysis with repeated measures is shown in Table 15.

Table 13: MANOVA output with positive features perceived as sustainable and features not related to sustainability
*: significant at p = 0.05, **: significant at p = 0.01, ***: significant at p = 0.001

	Pillai	~F	num Df	den Df	Pr(>F)
Product	0.192	8.13	10	768	<0.001***
FeatureType	0.056	11.3	2	383	<0.001***

The electric scooter features are statistically significant even with the more challenging list, like our previous findings with the French press. Thus, our findings support the generalizability of hypothesis 1 when using positive features perceived as sustainable and features not perceived as sustainable.

5.2.2 Product Analysis

In this section we present analyses for testing the generalizability of hypothesis 2. We used a repeated measured correlation to determine the relation between the "like" axis and "sustainability" axis based on where participants placed the products during the collage activity. The repeated measures correlation controls for between-participant variance (Bakdash and Marusich, 2017). The results are shown in Table 16.

Table 14: Repeated measures correlation between perceived sustainability of a product and liking the product

	Social	Environmental	Economic	Combined
Repeated Measure Correlation	**0.18**	**0.09**	**0.08**	**0.11**
P-value	**0.006**	**0.042**	**0.034**	**0.001**

There is a statistically significant relationship between liking a product and perceiving it as socially, environmentally, or economically sustainable. Moreover, there is a statistically significant relationship between liking a product and perceiving it as sustainable in general. These findings support the generalizability of hypothesis 2 and our previous findings when using a French press.

The correlations are low across the board, ranging from 0.08 to 0.18, suggesting that sustainability and liking a product can be measured separately and demonstrates the usefulness of the collage tool for assessing sustainability perceptions.

6 Discussion

Our findings support the generalizability of proposition 1 that phrases perceived as sustainable in reviews contain semantic and syntactic characteristics that can be modeled. Looking at the machine learning model metrics in tables 4 and 5 for electric scooters and baby glass bottles, respectively, we see that they are like the ones in our previous study with French presses (table 1). The metrics for negative sentiment faired poorer in this study, suggesting potential limitations on the generalizability (see below for more on limitations).

We found similarities and differences between features perceived as sustainable for electric scooters and baby glass bottles. For social sustainability in Fig. 7, many of the positive features for electric scooters were intangible, such as relating to family or gift-giving. This is like what we found when we previously tested the method using French presses. For baby glass bottles in Fig. 10, the positive features focused more on the bottle itself. As for the negative features for social sustainability, electric scooters had mainly tangible features relating to convenience, safety, and comfort while baby glass bottles had both intangible features such as “dad” or “promise”, and tangible features such as “crack”.

For environmental sustainability, the positive features for both electric scooters and baby glass bottles are mainly tangible. In the case of baby glass bottles in Fig. 11 this mainly related to the material such as “not plastic” and “bpa free”. Our previous results with French presses also showed that positive features for environmental sustainability focused on material. For electric scooters in Fig. 8, the positive features included many components such as the battery life, brakes, and tires. The same pattern appeared with negative features where baby glass bottles mainly focused on material while electric scooters included a range of features.

For economic sustainability, features for electric scooter in Fig. 9 related to how great of a value the product is, such as “great purchas” or “poor qualiti”. This is like what we found with our previous study with French presses. For baby glass bottles in Fig. 12, positive features included different brands, as well as tangible features like “plastic” while for negative features they included tangible features like “bottl leak”.

The results from the collage activity demonstrate a strong support for the generalizability of hypotheses 1 and 2. Tables 10 and 15 show that participants consistently placed positive electric scooter features perceived as sustainable higher on the sustainability axis than they did for other features across all sustainability pillars. This is like our results with French presses and supports the generalizability of hypothesis 1. It demonstrates that the method in this paper can be generalized to different products to extract perceived sustainable features that customers identify as sustainable. Designers can therefore use this method to identify and include features in sustainable products that align with customer perceptions of sustainability.

We also identified significant relationships between participants perceiving sustainability and liking a product in Table 16. The results indicate that the way customers perceive sustainability in products plays a role in how they like products, like our previous findings when we used French press products. Our findings therefore support the generalizability of hypothesis 2 that a significant relationship exists between evaluating sustainability and likeability of different products. Moreover, Table 16 shows low correlations for the different sustainability criteria, as we found in our previous work. The correlations were lower with electric scooters however, with social sustainability having the highest correlation at 0.18. This contrasts our results with French presses where environmental sustainability had the highest correlation at 38%. This suggests that the role that perceived sustainability plays in customers liking a product can differ between product types. The low correlations for both electric scooters and French presses, however, support that perceived sustainability can be measured separately from liking a product and demonstrate the effectiveness of the collage tool for designers to test perceived sustainable features with participants.

Our findings reveal crucial implications for guiding customers to making informed purchase decisions. We recommend that designers use the method in this study so that they may bridge the gap between perceived and engineered sustainability in their products and drive purchase decisions. The findings do come with limitations. Amazon products tend to have more positive than negative reviews by design, as they would not thrive on the platform if it were the other way around. One limitation therefore is that the machine learning model performance may suffer due to an imbalance in the dataset (Tables 4-5). We saw this in our results, for example, “love scooter” appears as a salient negative economic sustainability feature for electric scooters in Fig. 9. We found a greater imbalance in the annotated reviews for baby glass bottles, possibly because reviews for them tend to be exceptionally high to survive on Amazon. Designers, therefore, need to carefully assess potential data imbalance before using this method. Possible workarounds include collecting enough negative reviews to have a neutral overall rating score when annotating, or to collect reviews from a different platform that may have more balanced ratings.

7 Conclusion

This study validates the generalizability of our previously developed method to extract and test features perceived as sustainable from online reviews for different products. The insights from our results can help shape customer decisions to make informed purchases. To demonstrate this, we used two focal products, electric scooters and baby glass bottles, and recreated our previous work where we used French presses (El Dehaibi, Goodman and MacDonald, 2019; El-Dehaibi, Liao and MacDonald, 2021). We collected Amazon reviews for the focal products and recruited Amazon Mechanical Turks (MTurks) to annotate fragments of the reviews that are relevant to one of the sustainability pillars – social, environmental, and economic. We then modeled the annotations using a logistic classifier and extracted features perceived as sustainable based on the parameters of the model. We confirmed that the perceived features were identified as sustainable using a novel collage activity. We tasked participants with placing products along the two axes of the collage, sustainability and like, and to select from a dropdown menu features that we extracted from the machine learning model.

Based on the results we found that our previously proposed method does generalize with limitations. We shared crucial design insights that can help customers make informed sustainability purchase decisions. Designers can use the method in this study on different products to identify the gaps between perceived and engineered sustainability and create sustainable products that can drive purchasing decisions. We confirmed that this method can be applied to identify salient sustainable features based on customer perceptions. We recommend that designers use the collage tool to test and better understand customer perceptions of sustainability. We demonstrated how perceived sustainability and liking a product can be measured separately based on their low correlation. Moreover, we recommend that designers consider the influence of demographics on specific pillars of sustainability.

A limitation to our findings is that the method can be ineffective if there is an imbalance of positive and negative annotations from the reviews. Products should therefore be carefully selected to ensure a balanced dataset. Moreover, we recommend conducting a more thorough demographic analysis to better understand the relationship between demographics and customer perceptions. Finally, our analyses do not include real purchase decisions. For next steps, we aim to address the limitations in this study by exploring modifications to the method for imbalanced data and investigating how features perceived as sustainable in products influence purchase decisions.

Financial support

We would like to thank Qatar National Research Fund (QNRF) for supporting this work. This work was funded by QNRF under the Qatar Research Leadership Program (QRLP).

References

Bakdash, J. Z. and Marusich, L. R. (2017) 'Repeated Measures Correlation', *Frontiers in Psychology*, 8, p. 456. doi: 10.3389/fpsyg.2017.00456.

Du, P. and MacDonald, E. F. (2016) 'Product Body Shapes, Not Features, Provide Fast and Frugal Cues for Environmental Friendliness', in *Volume 7: 28th International Conference on Design Theory and Methodology. ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Charlotte, North Carolina, USA: American Society of Mechanical Engineers, p. V007T06A046. doi: 10.1115/DETC2016-60283.

El Dehaibi, N., Goodman, N. D. and MacDonald, E. F. (2019) 'Extracting Customer Perceptions of Product Sustainability From Online Reviews', *Journal of Mechanical Design*, 141(12), p. 121103. doi: 10.1115/1.4044522.

El-Dehaibi, N., Liao, T. and MacDonald, E. F. (2021) 'Validating Perceived Sustainable Design Features using a Novel Collage Approach', *2021 ASME International Design Engineering Technical Conferences, American Society of Mechanical Engineers, Online*, p. 14.

HD Delaney (2003) 'Higher Order Designs WIth Within-Subjects Factors: The Multivariate Approach', in *Designing Experiments and Analyzing Data*.

von Helversen, B. *et al.* (2018) 'Influence of consumer reviews on online purchasing decisions in older and younger adults', *Decision Support Systems*, 113, pp. 1–10. doi: 10.1016/j.dss.2018.05.006.

James, G. *et al.* (2013) *An Introduction to Statistical Learning*. New York, NY: Springer New York (Springer Texts in Statistics). doi: 10.1007/978-1-4614-7138-7.

Kim, E.-H. and Lyon, T. P. (2015) 'Greenwash vs. Brownwash: Exaggeration and Undue Modesty in Corporate Sustainability Disclosure', *Organization Science*, 26(3), pp. 705–723. doi: 10.1287/orsc.2014.0949.

Li, X. *et al.* (2019) 'A multi-stage hidden Markov model of customer repurchase motivation in online shopping', *Decision Support Systems*, 120, pp. 72–80. doi: 10.1016/j.dss.2019.03.012.

Liu, Y., Jiang, C. and Zhao, H. (2019) 'Assessing product competitive advantages from the perspective of customers by mining user-generated content on social media', *Decision Support Systems*, 123, p. 113079. doi: 10.1016/j.dss.2019.113079.

MacDonald, Erin Faith, Gonzalez, R. and Papalambros, P. (2009) 'The construction of preferences for crux and sentinel product attributes', *Journal of Engineering Design*, 20(6), pp. 609–626. doi: 10.1080/09544820802132428.

MacDonald, Erin F., Gonzalez, R. and Papalambros, P. Y. (2009) 'Preference Inconsistency in Multidisciplinary Design Decision Making', *Journal of Mechanical Design*, 131(3), p. 031009. doi: 10.1115/1.3066526.

MacDonald, E. F. and She, J. (2015) 'Seven cognitive concepts for successful eco-design', *Journal of Cleaner Production*, 92, pp. 23–36. doi: 10.1016/j.jclepro.2014.12.096.

Maslowska, E., Malthouse, E. C. and Viswanathan, V. (2017) ‘Do customer reviews drive purchase decisions? The moderating roles of review exposure and price’, *Decision Support Systems*, 98, pp. 1–9. doi: 10.1016/j.dss.2017.03.010.

Nysveen, H. and Pedersen, P. E. (2004) ‘An exploratory study of customers’ perception of company web sites offering various interactive applications: moderating effects of customers’ Internet experience’, *Decision Support Systems*, 37(1), pp. 137–150. doi: 10.1016/S0167-9236(02)00212-9.

She, J. and MacDonald, E. F. (2017) ‘Exploring the Effects of a Product’s Sustainability Triggers on Pro-Environmental Decision-Making’, *Journal of Mechanical Design*, 140(1), p. 011102. doi: 10.1115/1.4038252.

The Sustainability Imperative: New Insights on Consumer Expectations (2015). Nielsen, pp. 1–19.

Wang, Z. *et al.* (2016) ‘Saliency effects of online reviews embedded in the description on sales: Moderating role of reputation’, *Decision Support Systems*, 87, pp. 50–58. doi: 10.1016/j.dss.2016.04.008.

Zhang, Le, Yan, Q. and Zhang, Leihan (2020) ‘A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb’, *Decision Support Systems*, 133, p. 113288. doi: 10.1016/j.dss.2020.113288.