Article

AMERICAN MARKETING ASSOCIATION

Journal of Marketing Research
2024, Vol. 61(1) 5-26

DOI: 10.1177/00222437231169711
journals.sagepub.com/home/mrj
Sage

# Words Meet Photos: When and Why Photos Increase Review Helpfulness

**Gizem Ceylan , Kristin Diehl, and Davide Proserpio**

## Abstract

Are reviews with photos more helpful? If so, do consumers find reviews more helpful when photos and text convey similar or different information? This article examines the effect of content similarity between photos and text on review helpfulness and its underlying mechanism. Using a data set of 7.4 million reviews associated with 3.5 million photos from Yelp, and applying machine learning algorithms, the authors quantify the similarity of the content between text and photos. They find that, overall, photos increase the helpfulness of a review. More importantly, though, greater similarity between photos and text heightens review helpfulness more. The authors then validate algorithm-based similarity assessments with similarity perceptions of human judges. Using real-world reviews from Yelp and carefully designed stimuli, they replicate the core findings in five laboratory experiments. Further, testing the underlying mechanism, they find that greater similarity facilitates the ease with which consumers can process the review, which, in turn, increases that review's helpfulness to consumers. Finally, they show that factors that impede the ease of processing (e.g., language difficulty or poor image quality) can reduce the effect of similarity on helpfulness. These findings provide novel insights into the value of user-generated content that includes text and photos and its underlying mechanism.

## Keywords

photos, similarity, natural language processing, reviews, helpfulness, visuals, user-generated content

Online supplement: https://doi.org/10.1177/00222437231169711

People learn from others about a variety of things, such as places to visit and products to buy. The ability to access reviews online has systematically changed this learning process. Instead of asking a friend or an expert agent, today many people consult review platforms before eating at a restaurant or traveling to a new city. Prior research in marketing and computer science has examined the effect of structured features (e.g., star ratings) and unstructured features (e.g., the valence of the text) of a review as well as reviewers' characteristics (e.g., prior experience) on a review's helpfulness. Helpful reviews are not just important in shaping readers' decisions but also critical to firms' economic outcomes. For instance, helpful reviews are more likely to affect consumers' attitudes and behaviors, driving economic outcomes such as product sales (Ghose and Ipeirotis 2011; Topaloglu and Dass 2021) and box office success (Lee and Choeh 2020). In addition to text, review writers increasingly include photos in their reviews. Indeed, in surveys, readers state that they value user-generated photos in reviews (BazaarVoice 2021). In this research, we investigate how and when the relationship of photos and text may impact review helpfulness.

Our research question is centered around the interplay between what people communicate in text versus photos and its consequences on the helpfulness of that review. Review platforms (e.g., Amazon) often suggest that adding a photo to a review can increase that review's helpfulness (Schwartz 2019). Outside the review context, the mere presence of a photo increases engagement with the platform (Li and Xie 2020). Further, Hartmann et al. (2021) distinguish between the type of image that generates more engagement with the post versus the brand. They find that consumer selfies receive more post-related engagement, such as likes and comments, whereas brand selfies result in more brand engagement, such as purchase intentions. On platforms such as Airbnb, platform-provided (high-quality) photos of the property can increase demand (Zhang et al. 2022). While important, these papers focus only on the effect of photos without taking text into account.

Gizem Ceylan is Postdoctoral Researcher, Yale School of Management, Yale University, USA (email: gizem.ceylan@yale.edu). Kristin Diehl is Professor of Marketing, Marshall School of Business, University of Southern California, USA (email: kdiehl@marshll.usc.edu). Davide Proserpio is Associate Professor of Marketing, Marshall School of Business, University of Southern California, USA (email: proserpi@marshall.usc.edu).

In our research, we establish that adding photos to a review increases review helpfulness. More importantly, though, we focus on the relationship between photos and text and examine whether consumers find a review more helpful when the photo and text convey similar (vs. dissimilar) content, using a combination of secondary data and lab experiments. We test our predictions in a large-scale data set of 7.4 million restaurant reviews from two different metropolitan areas, of which about 1.42 million were associated with at least one photo. Applying a representation learning algorithm, Doc2Vec (Le and Mikolov 2014), we assess the extent to which the content of the photo is similar to the content of the text. Controlling for review and reviewer characteristics, we find that greater similarity between photo and text content heightens helpfulness. We replicate these findings in five laboratory experiments that enable us to establish the causal link between photo–text similarity and review helpfulness. In addition, we establish one reason that similar (rather than dissimilar) content in photos and text is helpful: similar content in photos and text makes the review easier to process for the reader, which subsequently heightens the review's perceived helpfulness.

Jointly, our research examines a previously unexplored aspect of online reviews and user-generated content that includes both visual and verbal information and makes several important contributions. We identify similarity of the content between photos and text as a novel antecedent that impacts a review's helpfulness. The interplay between visual and verbal information has been predominantly studied in the advertising literature (e.g., Edell and Staelin 1983). These studies often used highly stylized, strategically designed stimuli and small-sample laboratory experiments. We extend the investigation of visual–verbal interplay to organic conversations in user-generated content. User-generated content is distinct from firm-generated advertisements because consumers generate qualitatively different content (e.g., in their review, they admit purchase mistakes; Reich and Maglio 2020) and use different styles (e.g., they include figurative speech; Kronrod and Danziger 2013) compared with how managers communicate in firm-generated advertisements. Examining the text and pictures in millions of real-world consumer reviews using advances in machine learning, we demonstrate the effect of photo–text similarity on helpfulness. We also shed light on the process underlying this effect, contributing an important but previously unidentified psychological mechanism: ease of processing. Using a carefully screened subset of real-world reviews as well as systematically designed stimuli, we conduct laboratory experiments ($N = 3,879$) and causally test the mechanism underlying our effect and its boundary condition.

Our findings rest on a multimethod approach that combines advances in machine learning with human validation and experimental evidence. This multimethod approach yields externally valid results that provide robust, causal evidence for our predictions. In doing so, we shed light on the emerging literature on visual–verbal user-generated content and identify novel insights into what makes user-generated content helpful in general.

Further, understanding the antecedents of helpfulness in the context of visual–verbal content is of practical importance for platforms, influencers, and everyday consumers. Platforms continuously search for ways to recognize helpful reviews even before a critical mass of consumers can identify them (e.g., through likes). They also strive to guide consumers to create high-value reviews for others. Professional influencers and everyday consumers search for ways to create helpful content and gain influence. Our findings can inform all these efforts and provide implications for platforms, influencers, and consumers.

## Theoretical Development

### *Reviews and Review Helpfulness*

Consumers often consult reviews to reduce uncertainty and aid their decision-making processes. Previous work suggests that reviews can significantly impact firms' economic outcomes. For example, research has demonstrated an association between review ratings and sales (Chevalier and Mayzlin 2006) and between review volume and sales (Duan, Gu, and Whinston 2008).

Helpful reviews are more likely to influence consumers' attitudes and behavior. As a result, helpful reviews have a larger economic impact on firms' sales (Ghose and Ipeirotis 2011; Lee and Choeh 2020). But what makes a review helpful? Prior research has identified several reviewer characteristics and features of the review that can improve helpfulness: the identity of the reviewer (Forman, Ghose, and Wiesenfeld 2008), star ratings (Kim et al. 2006), rating extremity (Mudambi and Schuff 2010), semantic and stylistic aspects of the review text (Ghose and Ipeirotis 2011), and text readability and informativeness (Ghose and Ipeirotis 2011). Our research adds to this literature by examining a novel antecedent of helpfulness: the similarity between the content of the photo and the text.

### *The Impact of Photos on Review Helpfulness*

Whereas words are important to convey experiences, with the widespread adoption of smartphones, it has become easier than ever for people to take and share photos of their experiences. Indeed, people share 6.9 billion photos daily on WhatsApp (Broz 2023) and create more than 4 billion snaps daily on Snapchat (Aslam 2023). Further, many review platforms increasingly encourage and even require reviewers to include photos with their review text. For instance, Amazon now reminds reviewers to add a photo to their review text, claiming that shoppers find reviews that include images more helpful than text alone.

Prior research offers several reasons why photos may be helpful. First, photos can carry important information that is hard to convey with words alone (e.g., the restaurant's decor). Second, photos grab attention and are processed first (Pieters and Wedel 2004), increasing consumers' likelihood to pay attention and read the review. Third, visuals evoke more intense emotional reactions (Rossiter and Percy 1980), which may be particularly important for hedonic purchases. Whereas we focus on photos in the context of user-generated reviews, the advertising literature has also examined photos' effect on comprehension, recall, and attitudes and found that photos indeed improve these outcomes (Edell and Staelin 1983). For these reasons, we posit that the mere presence of a photo (or multiple photos) in a review would increase its perceived helpfulness.

## *The Interplay Between Photos and Review Text Can Affect Review Helpfulness*

Whereas the mere presence of a photo may increase a review's helpfulness, in this research, we uncover *when and why* this would be the case. In the context of visual–verbal communication, both the photo and the text convey information, and readers integrate this information to construct an overall meaning. In this research, we investigate how the similarity between photo and text content affects ease of processing and its downstream consequences.

*Similarity between photo and text.* Similarity assessments are a fundamental human process (Markman and Gentner 1993; Tversky 1977), and similarity between different pieces of information is central to a wide range of outcomes, such as attitude formation and persuasion (Edell and Staelin 1983), learning (Harp and Mayer 1998), and memory (Markman and Gentner 1996). Whereas prior research has largely focused on the similarity between entities of the same modality (e.g., photo–photo; Markman and Gentner 1996; Tversky 1977), similarity judgment can also involve entities of different modalities (e.g., photo–text; Clark and Mayer 2016), which is the focus of our research.

Similarity assessments involve a comparison between two entities. One can compare either the features of the entities (feature-based similarity; Tversky 1977) or the structural relations within each entity (relational similarity; Markman and Gentner 1993). Whereas humans employ both forms of similarity assessments independently (Markman and Gentner 1993), feature-based similarity is often more prevalent. Feature-based (vs. relational) similarity assessment is developed earlier in life (Pierce and Gholson 1994) and underlies many fundamental human processes, such as categorization and inference-making (Tversky 1977). Thus, we focus our investigation on feature-based similarity between the photo and the text content.

We define feature-based content similarity as alignment between what the picture depicts and what the text describes. For instance, imagine a reviewer writes about coffee at a coffee shop and includes a photo of the coffee mentioned in the text. This is an example of a review in which the reviewer conveys *similar* content (i.e., coffee) in each modality. However, if the reviewer writes about coffee but includes a photo of a croissant, then the reviewer conveys *dissimilar* content (i.e., coffee vs. food) in the text and photo.

Because photos are limited in what they can represent compared with words (Kieras 1978), we limit our investigation to representations and properties that can be conveyed visually and verbally. We leave it to future research to investigate other dimensions on which text and photos can align, such as valence, level of abstraction, or intended call to action (Amit, Algom, and Trope 2009; Villarroel Ordenes et al. 2019). We also leave it to future research to investigate cases in which text and photo align content-wise, but this information is incongruent with the context in which they are shared (e.g., posting a review text and picture of a washing machine [high content similarity] on a restaurant review site such as Yelp). In the literature, alignment on higher-order dimensions such as valence, level of abstraction, or context has been referred to as "coherence." Notably, coherence is independent of structural aspects such feature-based similarity and does not necessitate greater processing ease (Winkielman et al. 2012). Although such situations are important, they are beyond the scope of this research; we will return to them in the "General Discussion" section.

*The effect of content similarity on perceived processing ease and helpfulness.* The question we ask is: how does similarity affect helpfulness? When the content of the review text differs from the content of the photo (e.g., coffee in the text and food in the photo), the review overall conveys a larger amount of information compared with when the text and photo relay the same content (e.g., coffee in the text and in the photo). Reviews with more (vs. less) information can be more helpful (Ghose and Ipeirotis 2011) because they reduce uncertainty to a greater extent (Driscoll and Lanzetta 1964). However, we argue that when the content of the photo and the text convey similar (not dissimilar) information, the review as a whole becomes more helpful due to greater perceived processing ease.

Our prediction is based on prior work on processing fluency (i.e., the subjective feeling of ease with which people process information; Reber, Schwarz, and Winkielman 2004). This research argues that "consistency," a content match between cognitive elements, may result in greater fluency (Winkielman et al. 2012). Whether individuals process information visually or verbally (e.g., seeing a photo of a coffee cup or reading about coffee), a particular concept (e.g., coffee) is activated in their working memory. Prior research shows that earlier exposure to an object (Hershenson and Haber 1965; Jacoby and Dallas 1981), to some of its attributes (Reber, Winkielman, and Schwarz 1998), or to related semantic concepts (Winkielman et al. 2003) makes subsequent processing of similar information easier, increasing both objective and subjective processing fluency (Winkielman et al. 2003). However, when different modalities activate different concepts, people may experience conflict and frustration (Meyers-Levy and Tybout 1989), that is, a lack of objective and subjective processing fluency. Thus, we expect greater content similarity to increase perceived processing ease.

Perceived processing ease elicits positive affect (Winkielman et al. 2003). When asked for an evaluation of an entity, people draw on this positive subjective experience and evaluate the fluently processed entity more favorably (for reviews, see Reber, Schwarz, and Winkielman [2004]). Hence, we expect greater similarity between the photo and text content to enhance perceived processing ease and, as a result, increase the helpfulness of the review. In summary, we predict:

**$H_1$:** Greater similarity between the content in the photo and in the text of a review increases the perceived helpfulness of a review.

**$H_2$:** Greater similarity between the content in the photo and in the text of a review increases perceived processing ease.

**$H_3$:** The effect of content similarity on helpfulness is due to the positive effect of similarity on perceived processing ease.

### *Limits of Similarity-Driven Effects on Helpfulness*

We predict that reviews become more helpful when they convey similar (vs. dissimilar) content in photo and text because feature-based content similarity increases perceived processing ease. Prior research suggests that when a particular factor (e.g., familiarity) affects a focal judgment (e.g., truth) due to greater ease of processing, other unrelated factors that independently hamper perceived processing ease (e.g., text readability) can attenuate or even eliminate that original effect (e.g., of familiarity on truth judgments; Reber and Schwarz 1999). In our context, we expect that such unrelated elements (i.e., unrelated to feature-based content similarity) that hamper processing ease will weaken the positive effect of feature-based content similarity on helpfulness.

One reason for this attenuating effect is that greater difficulty with which a stimulus is processed disrupts naturally occurring processes (Alter et al. 2007). For example, people are less likely to answer easy trivia questions correctly when they are presented in a hard-to-read font (Song and Schwarz 2008). Hence, we predict that when ease of processing of a review is hampered, it will disrupt or even eliminate the positive effect of content similarity on helpfulness.

Ease of processing can be altered in ways that are incidental or integral to the problem at hand (Alter and Oppenheimer 2009). Prior research suggests that incidental factors such as font usage (Novemsky et al. 2007), figure–ground contrast (Reber and Schwarz 1999), and semantic priming (Kelley and Lindsay 1993) can all alter ease of processing. Further, factors integral to the focal problem, such as linguistic ease (Alter and Oppenheimer 2009) or prototypicality of the target (Winkielman et al. 2006), or congruity with one's lay beliefs (Philipp-Muller, Costello, and Reczek 2022), can alter perceived processing ease. In the context of reviews, we expect two integral factors to be particularly relevant to reducing processing ease and hence to attenuate or even eliminate the effect of content similarity on helpfulness: review text difficulty and photo quality. A harder-to-read text (Shulman et al. 2020) can make the review text more difficult to parse, and low-quality images impair visual assessment (Ryu, Park, and Park 2022). In summary, we expect factors that hamper processing ease to attenuate the effect of content similarity on helpfulness (compared with the absence of such impediments). Formally:

**$H_4$:** The positive effect of content similarity on helpfulness is attenuated when processing ease is low.

## Overview of Studies

To ensure external and internal validity, we test our predictions in a data set that includes organically created consumer reviews (from Yelp) and in five experiments. Study 1 provides real-world evidence for our core prediction: greater photo–text similarity is associated with greater helpfulness. Next, we validate algorithm-based similarity judgments with similarity assessments of human judges (Study 2a), establishing that humans and computers align on similarity perceptions. We then replicate our core finding across several experimental studies using both Yelp reviews (Studies 2b, 3a, and 4) and more controlled stimuli (Studies 3b and 3c). We examine the psychological mechanism of ease of processing using mediation (Studies 3a–3c). Finally, Study 4 provides additional evidence for the underlying mechanism by manipulating ease of processing. We provide an overview of studies and findings in Table 1.

## Study 1: Greater Photo–Text Similarity Increases Helpfulness in Yelp Reviews

In Study 1, we use field data to investigate the relationship between photo–text similarity and helpfulness. We also provide initial insights into the proposed ease of processing-based mechanism by examining text readability and photo quality of Yelp reviews.

### *Data*

To test our predictions, we collected a panel data of Yelp reviews at the restaurant level for the complete set of restaurants located in two U.S. locations: Los Angeles County in California and the Boston area in Massachusetts.[1] The data set contained about 24,964 restaurants listed on Yelp, 7.4 million reviews (including the 3.5 million photos associated with them) written by 2.1 million reviewers during 2004–2020 (2021 for Boston). For every review, we obtained its text, star rating, number of useful votes, and all the photos that the reviewer uploaded along with the review, if any. For every reviewer, we obtained their location (at the city level that Yelp provides), whether the reviewer had Yelp elite status, and the total number of reviews the reviewer had written so far.

[1] Surrounding cities such as Cambridge and Brookline are part of the Boston area data set.

**Table 1.** Overview of Studies.

| | Key Design Factors | Main Findings | Alternative Explanations Addressed |
|---|---|---|---|
| Study 1 | Yelp data set (correlational) in Los Angeles and Boston areas | Key effect: similarity between review text and photo is positively related to useful votes in Yelp reviews<br>Process evidence: interactive effects of similarity and text readability and similarity and photo quality; processing impairments alter the effect of similarity on useful votes | — |
| Study 2a | 5 × 100 reviews from Yelp data set (correlational) | Human validation: provides evidence that human-judged and algorithm-based similarity assessments are aligned using a subset of Yelp reviews | Algorithm-based similarity assessments are indeed related to human assessments. |
| Study 2b | Main independent variable: similarity between text and photo (similar vs. dissimilar)<br>Within-subject factor: four replicate sets | Experimental demonstration of the effect of similarity: assesses the effect of similarity on helpfulness across four different stimuli sets selected from Yelp reviews and pretested with human judges | Direction of causality (similarity → helpfulness) is established. |
| Studies 3a, 3b, 3c | Main independent variable: similarity between text and photo (similar vs. dissimilar) | Generalization of the effect of similarity and process evidence (mediation):<br>• Study 3a: Yelp reviews<br>• Study 3b: each photo/text replicate appears in both similar and dissimilar conditions<br>• Study 3c: manipulated levels of similarity (high vs. low) holding conceptual commonality constant<br>Mediation of ease of processing: consumers process similar review text with greater ease than dissimilar review text, and ease of processing mediates the relationship between similarity and helpfulness | Direction of the effect (similarity → helpfulness) is established.<br>Stimulus idiosyncrasies do not drive results.<br>Self-selection is ruled out.<br>Results are not due to differences in conceptual commonality. |
| Study 4 | Main independent variable: similarity between text-photo (similar vs. dissimilar)<br>Moderator: fluency of text and photo (fluent vs. disfluent) | Boundary condition or process by moderation: provides evidence that similarity enhances helpfulness when the text and the photo are fluent to the reader but not when they are disfluent | Direction of causality (similarity → helpfulness) is established.<br>Direction of causality (ease of processing → helpfulness) is established. |

## *Measuring Photo–Text Similarity*

We created a measure of similarity between the content of the text and the photos in two steps. First, we used Google Cloud Platform Vision API and the "Detect Labels" function. The Vision API can detect and extract information about entities in an image across a broad group of categories. Labels can identify general objects, locations, activities, animal species, products, and more.[2] We provide a few examples of images and labels extracted in Figure 1.

Next, we applied Doc2Vec—a representation learning algorithm that converts text documents to low-dimensional vectors—to obtain vectors for both reviews and photo labels (Le and Mikolov 2014). An important property of these document vectors is that documents close together in the vector space have similar meanings and documents distant from each other in the vector space differ in meaning. In line with prior research, we measured semantic similarity between reviews and photo labels by computing the cosine similarity between their vectors. The maximum similarity between the vector of the review text and the vector of the photo label is indicated by $\cos\theta = 1$, and maximum dissimilarity is indicated by $\cos\theta = -1$.

We trained Doc2Vec using 80% of the entire review corpus of both Yelp reviews and image labels extracted from the photos associated with each review.[3] Since a review can be associated with multiple photos, we created one "photos document" for each review by concatenating the labels extracted from each photo associated with the same review. Additionally, a few parameters need to be set in this type of analysis: the size of the vectors, the number of dimensions of the vectors, the window size (i.e., the maximum distance [in words] between the current and the predicted word within a document), and the number of iterations carried out over the training corpus. After testing different configurations and assessing each model by computing its ability to find similar documents, we set these values to 64 vector dimensions, a window size of eight words, and 120 iterations because this configuration behaved slightly better

[2] For more information, see https://cloud.google.com/vision/docs/labels.

[3] We preprocess reviews and photo documents to tokenize and remove words shorter than 2 characters and longer than 15 characters.


**Figure 1.** Examples of Photos and Labels Extracted Using Google Vision API.

than all others.[4] We provide a visual description of our approach in Figure 2 and examples of reviews with high and low photo–text similarity in Figure 3.

### Measuring Reading Difficulty

We measured reading difficulty using the Flesch–Kincaid index (FKI; Kincaid et al. 1975) as a readability metric. Higher FKI values imply greater difficulty in comprehension of the written text.

Any advantage of photo–text similarity on ease of processing and helpfulness should diminish when the text is difficult to read. We test this prediction ($H_4$) by including an interaction between similarity and reading difficulty in the model. If $H_4$ is supported, this interaction should be negative, such that more difficult language should dampen the effect of photo–text similarity.

### Measuring Image Quality

We measured image quality using the Neural Image Assessment model described in Talebi and Milanfar (2018), which uses convolutional neural networks to compute a measure of technical image quality.[5] Higher values of this variable imply higher image quality.

Any advantage of photo–text similarity on ease of processing and subsequently helpfulness should diminish when the image is hard to grasp. We test this prediction ($H_4$) by including an interaction between similarity and image quality in the model. We expect higher values of the image quality measure to be associated with easier processing of images (e.g., clearer). If $H_4$ is supported, this interaction should be positive, such that easier-to-process photos facilitate the effect of photo–text similarity.

### Descriptive Statistics

Uploading photos in addition to the review text became popular in the last decade, likely due to the diffusion of smartphones. This is visible in Figure 4, where we plot, in Panel A, the monthly number of reviews and, in Panel B, the monthly number of photos posted on Yelp.[6] Out of the 7.4 million reviews, about 1.4 million are associated with at least one photo, and among these reviews, the average number of images per review is 2.46.

As Figure 5 shows, photos are more likely to be associated with reviews with positive ratings (three stars or above) rather than negative ratings (one or two stars). Turning to the similarity between review text and photos, in Figure 6, we plot the distribution of similarity. The mean similarity is .21 (SD = .11), suggesting that, on average, there is some overlap between the content of the review text and what is displayed in the photos. Finally, the average number of useful votes per review is 1.1, and this number doubles when considering only reviews with at least one image. Table 2 reports the summary statistics for the data set containing all reviews and for the data set containing only reviews with at least one photo.

### Model

We started by assessing the effect of the presence of photos in a review on the number of helpful votes the review received. We estimate the following model:

$$\log \text{Helpful Votes}_{ijt} = \beta\, \text{Has Photos}_{ijt} + X'_{ijt}\gamma + \alpha_j + \tau_t + \epsilon_{ijt}, \quad (1)$$

where the dependent variable is the log of the number of helpful votes of review i of restaurant j received at year-month t.[7] Has Photos, the variable of interest, is a binary indicator of whether the review i is associated with any photos (1) or not (0); $X'_{ijt}$ is a

[4] Across all models we tested, we found that inferred documents are found to be most similar to themselves in more than 95% of the cases, suggesting that the models behaved in a consistent manner. In assessing Doc2Vec models we followed procedures outlined at https://radimrehurek.com/gensim/auto_examples/tutorials/run_doc2vec_lee.html.

[5] We use the implementation and pretrained model available at https://idealo.github.io/image-quality-assessment/#datasets.

[6] In addition to showing the popularity of photos, Figure 4 shows the effect that the COVID-19 pandemic had on the number of reviews and photos uploaded to Yelp. All results presented in the empirical analysis section are robust to the exclusion of the year 2020.

[7] Every time we take the log of a variable, we add 1 to avoid taking the log of zero.

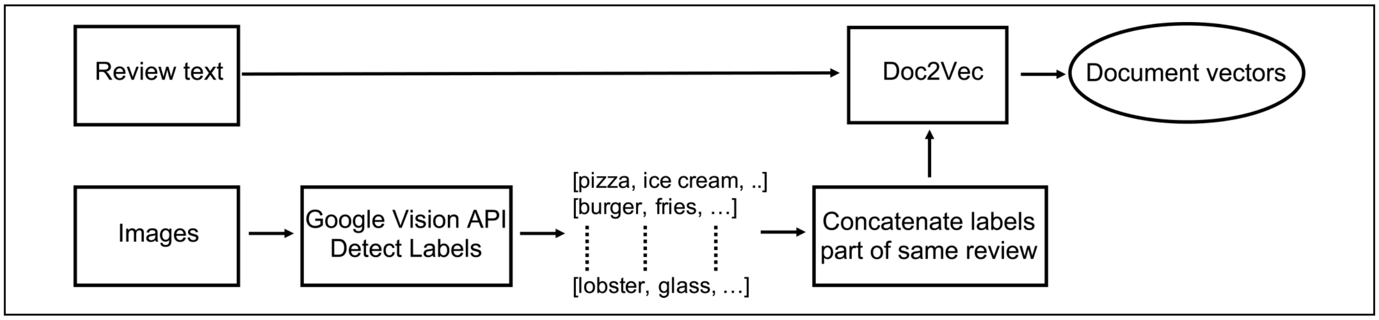


**Figure 2.** Visual Depiction of Similarity Assessment Between Review Text and Images in a Review.

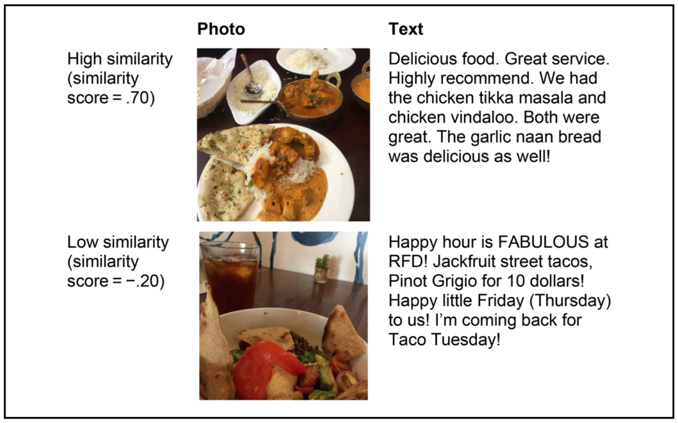


**Figure 3.** Examples of Reviews with High and Low Similarity Between the Photo and the Text.

vector of time-varying controls in which we included review rating, the log of review length (in characters), whether the reviewer is local (i.e., whether the reviewer wrote a review for a restaurant that is in the same city as the reviewer's location), whether the reviewer had elite status, and the log of the number of reviews written by the reviewer of review i. Finally, we included restaurant and year-month fixed effects to account for time-invariant unobservable restaurant characteristics and time-varying shocks (e.g., the COVID-19 pandemic) common to all restaurants that can affect the number of helpful votes a review receives. Because we include restaurant fixed effects, our specification exploits within-restaurant variation to estimate the effect of a review including at least one photo on the helpful votes it receives. We estimated Equation 1 using ordinary least squares, clustering standard errors at the restaurant level, following standard practice (Bertrand, Duflo, and Mullainathan 2004). We report the estimates in Table 3. Without any time-varying controls (Column 1), the coefficient of interest Has Photos was positive and significant, suggesting that adding photos to reviews increases the review's helpful votes. The results were similar when we included the wide array of controls discussed previously (Column 2).

We next measure the effect of similarity on helpfulness. For this purpose and because similarity exists only for reviews with photos, we focus on the set of reviews with at least one photo.

We estimate the following model:

$$\log \text{Helpful Votes}_{ijt} = \beta\ \text{Similarity}_{ijt} + X'_{ijt}\gamma + \alpha_j + \tau_t + \epsilon_{ijt}, \qquad (2)$$

where the dependent variable is the log of the number of useful votes of review i of restaurant j received at year-month t. $\text{Similarity}_{ijt}$, the variable of interest, is the similarity score discussed previously. The rest of the variables are as in Equation 1 with the only addition of the log of the number of photos in $X'_{ijt}$. We estimated Equation 2 using ordinary least squares, clustering standard errors at the restaurant level.

### *Effect of Photo–Text Similarity on Useful Votes*

We report the estimates of Equation 2 in Column 1 of Table 4. The coefficient of interest, β, is positive and significant and suggests that increasing similarity also increases useful votes. The interpretation of the estimate is that a 1 percentage point increase in similarity leads to a .11% increase in helpful votes. If similarity changes by one standard deviation (.12), then the effect on helpfulness is roughly 1.3%. While it may seem a small effect, review rankings on platforms like Amazon or Yelp likely depend on how helpful a review is. Therefore, even small changes in helpful votes can potentially affect the rank—and thus the visibility—of a review, and more helpful reviews can positively influence sales (Ghose and Ipeirotis 2011).

In addition, and consistent with the results presented in Table 3, we find that the number of photos associated with a review has a positive and significant effect. In this analysis, we control for the rating of the review; in the Web Appendix, we show that the positive effect of similarity holds at each level of star ratings, but the magnitude of the effect increases as ratings increase.

### *Robustness Checks*

*Reviewer fixed effects.* So far, we compared usefulness votes across reviews written for the same restaurants, controlling for unobserved restaurant heterogeneity. However, one could argue that reviewer heterogeneity could be driving

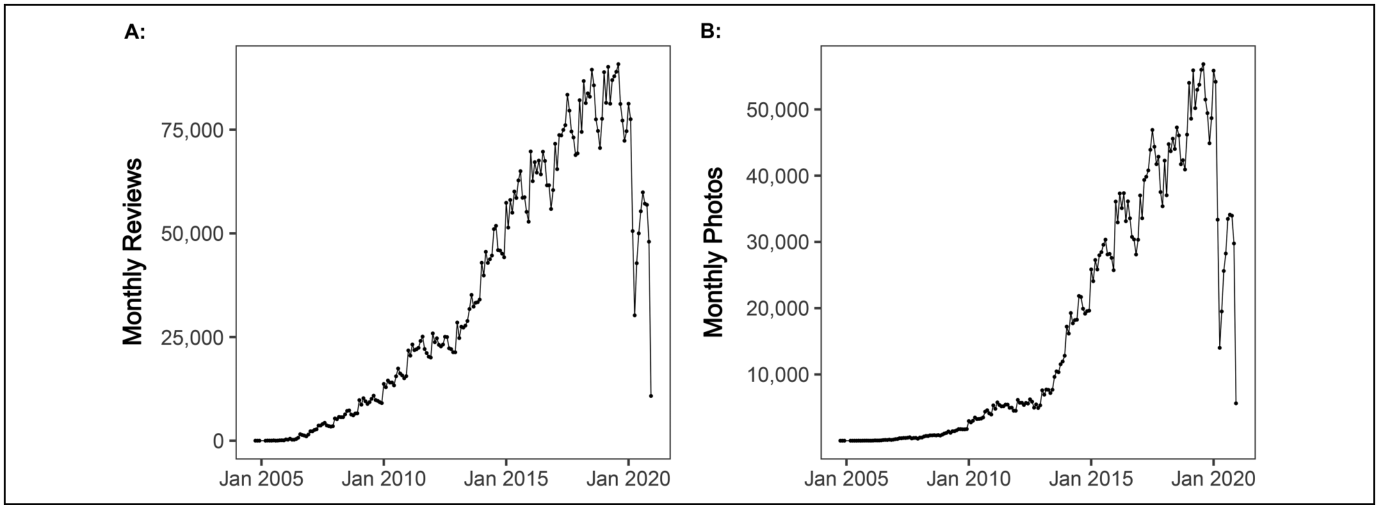


**Figure 4.** Number of Monthly Yelp Reviews (Panel A) and Monthly Number of Photos Posted (Panel B).

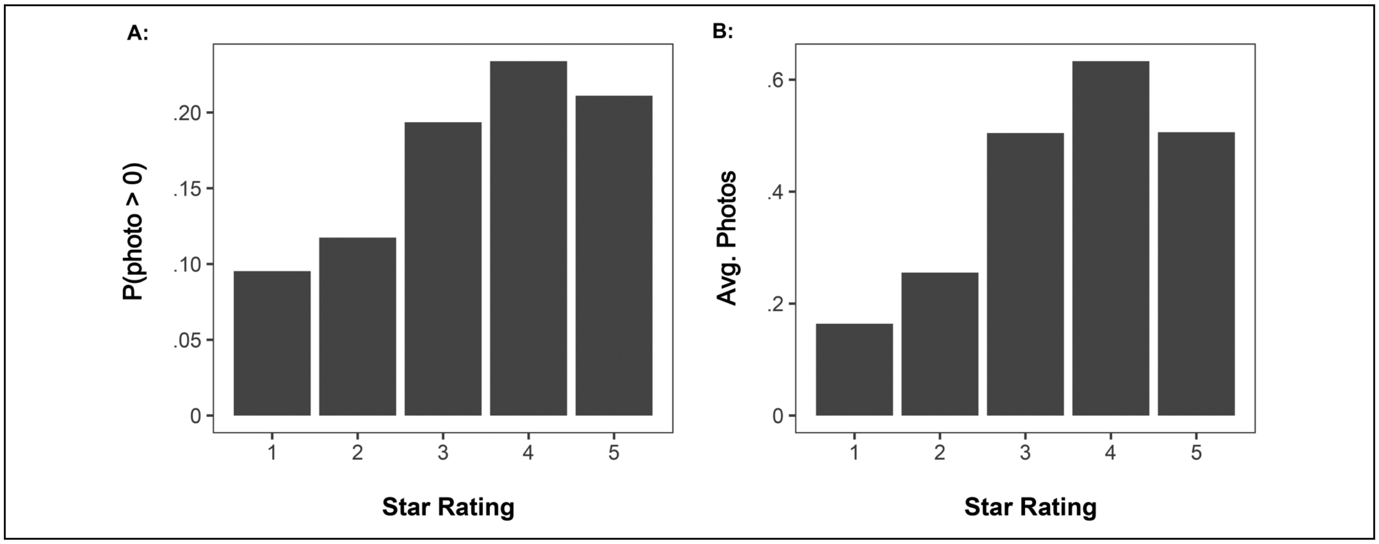


**Figure 5.** Fraction of Reviews with at Least One Photo by Star Rating (Panel A) and Average Number of Photos per Review by Star Rating (Panel B).

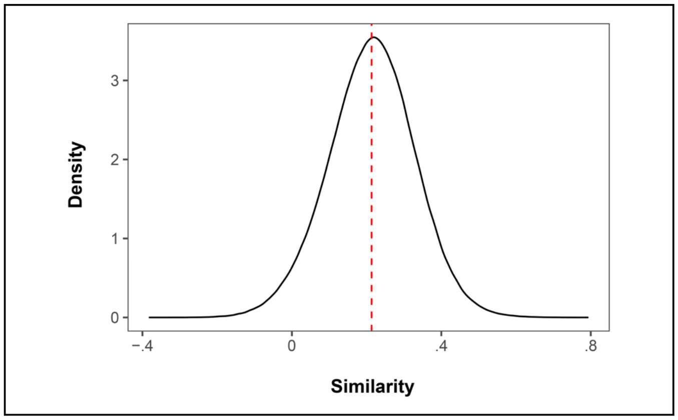


**Figure 6.** Density of the Similarity Scores Between Review Text and Photos.

the results discussed previously. Whereas we control for reviewers' experience by including in the model whether the reviewer is an elite reviewer, the number of reviews written by the reviewer, and whether the reviewer is local, it is still possible that some reviewers are just better than others at creating useful reviews and that these reviews are more likely to use similar text and photos. To reduce this concern, we reestimated Equation 2 but replaced the restaurant fixed effect with the reviewer fixed effect. Doing so, we estimate the effect of similarity by exploiting within-reviewer variation. In addition, we cluster standard errors at the reviewer level to account for potential serial correlation among useful votes of reviews written by the same reviewer. We report the estimates in Column 2 of Table 4. Overall, the results are consistent with those previously reported in

**Table 2.** Summary Statistics.

| | All Reviews | | With Photos | |
|---|---|---|---|---|
| | **M** | **SD** | **M** | **SD** |
| Similarity | — | — | .214 | .115 |
| Helpful votes | 1.073 | 3.347 | 2.137 | 6.088 |
| Number of photos | .477 | 1.365 | 2.462 | 2.176 |
| Star rating | 3.897 | 1.354 | 4.165 | 1.111 |
| Review length | 532.051 | 503.156 | 667.765 | 601.849 |
| User is local | .199 | .399 | .172 | .377 |
| User is elite | .115 | .319 | .263 | .440 |
| User reviews | 128.919 | 327.374 | 213.896 | 474.306 |
| Avg. photo quality | — | — | 4.604 | .540 |
| FKI | 11.080 | 16.279 | 10.907 | 17.240 |

**Table 3.** The Effect of Photos on Useful Votes.

| | (1) Without Controls | (2) With Controls |
|---|---|---|
| Has photos | .323*** | .172*** |
| | (.001) | (.001) |
| Star rating | | −.050*** |
| | | (.001) |
| Log review length | | .189*** |
| | | (.001) |
| User is local | | .018*** |
| | | (.001) |
| User is elite | | .223*** |
| | | (.001) |
| Log user reviews | | .078*** |
| | | (.0004) |
| Restaurant fixed effects | Yes | Yes |
| Year-month fixed effects | Yes | Yes |
| Observations | 7,375,820 | 7,375,820 |
| $R^2$ | .114 | .277 |

*$p < .1$.
**$p < .05$.
***$p < .01$.
*Notes:* The dependent variable is the log of useful votes of each review. All specifications include restaurant and year-month fixed effects. Cluster-robust standard errors at the restaurant level are in parentheses.

Column 1 of Table 4, suggesting that unobserved reviewer heterogeneity is unlikely to drive our results.

*Review and image label topics.* One may also wonder whether some reviews are more helpful as a function of the topic the review addresses or the photo depicts. To reduce this concern, we assessed the robustness of the results by including both review and image label topics in our model. We estimated review and image label topics using the latent Dirichlet allocation (LDA) algorithm discussed in Blei, Ng, and Jordan (2003). We relied on the parallelized LDA algorithm provided by the Python Gensim library. We set the parameter α (which represents the document-topic density) to be either symmetric or asymmetric, and the parameter η (which represents the topic-word

**Table 4.** The Effect of Similarity on Useful Votes.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Similarity | .107*** | .041*** | .124*** | .109*** |
| | (.005) | (.005) | (.005) | (.005) |
| Log number of photos | .142*** | .061*** | .144*** | .142*** |
| | (.002) | (.002) | (.002) | (.002) |
| Star rating | −.050*** | −.024*** | −.037*** | −.049*** |
| | (.001) | (.001) | (.001) | (.001) |
| Log review length | .239*** | .180*** | .234*** | .239*** |
| | (.001) | (.002) | (.001) | (.001) |
| User is local | .026*** | | .024*** | .026*** |
| | (.002) | | (.002) | (.002) |
| User is elite | .161*** | | .165*** | .161*** |
| | (.002) | | (.002) | (.002) |
| Log user reviews | .145*** | | .148*** | .145*** |
| | (.001) | | (.001) | (.001) |
| Review-to-review similarity | | | | −.024* |
| | | | | (.014) |
| Restaurant fixed effects | Yes | No | Yes | Yes |
| Reviewer fixed effects | No | Yes | No | No |
| Year-month fixed effects | Yes | Yes | Yes | Yes |
| LDA topics | No | No | Yes | No |
| Observations | 1,428,587 | 1,428,587 | 1,428,587 | 1,428,587 |
| $R^2$ | .344 | .720 | .347 | .344 |

*$p < .1$.
**$p < .05$.
***$p < .01$.
*Notes:* The dependent variable is the log of useful votes of each review. All specifications include year-month fixed effects. Cluster-robust standard errors (at the restaurant level in Columns 1, 3, and 4, and at the reviewer level in Column 2) are in parentheses.

density) to "auto" (i.e., the model learns the asymmetric prior from the data).[8] We varied the number of topics k between 5 and 20 and plotted the coherence score (Syed and Spruit 2017) as a function of the number of topics (see Figure 7). First, we noticed that using asymmetric α produces slightly higher coherence scores. Second, we used the so-called elbow technique to select the optimal number of topics. The idea behind this method is that we want to choose a point after which the increase of the coherence score is no longer worth the additional increase of the number of topics. Using this method, we selected 11 review topics and 10 image label topics.[9] Extracting topics from review text, we identified topics related to the quality of the service received and the type of food items served (e.g., burger, sushi). Extracting topics from photo labels, we identified topics related to the restaurant menu, drinks, indoor and outdoor elements of the restaurant, and the type of food items served (e.g., pizza, sandwich). We then included the topic weights of each

[8] In the symmetric case, the model uses a fixed symmetric prior for the document-topic distribution, whereas in the asymmetric case, the model uses a fixed normalized asymmetric prior.
[9] We also tested 18 label topics because this number produces the highest coherence score, and we obtained similar results.

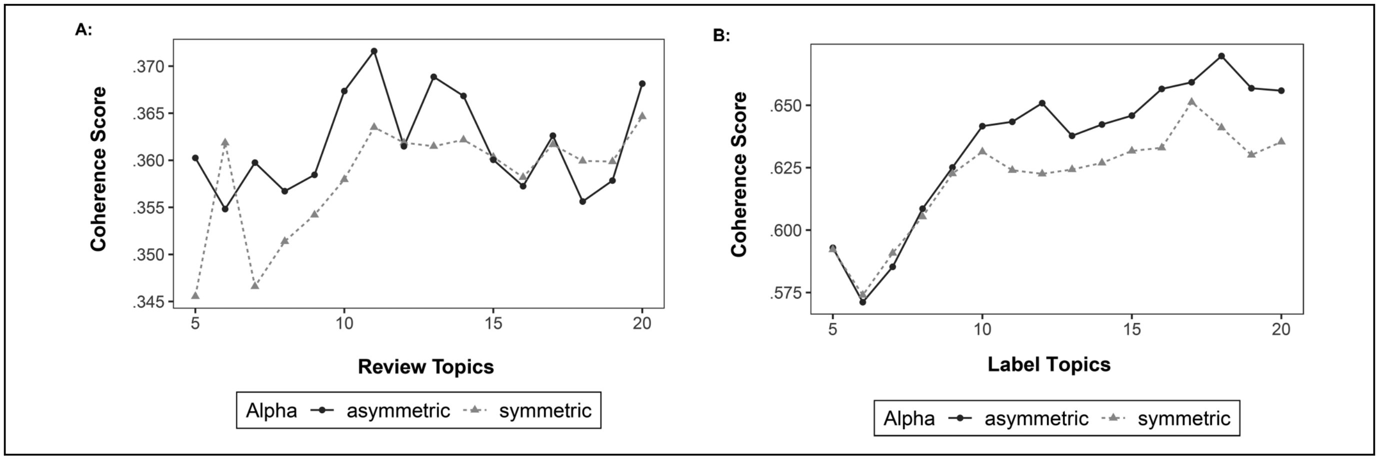


**Figure 7.** Coherence Score as a Function of the Number of Topics for Reviews (Panel A) and Image Labels (Panel B).

review and image labels as controls in Equation 2. We report these results in Column 3 of Table 4. We continue to observe positive and significant effects of similarity, suggesting that the topics discussed in the review or depicted in the photo do not drive the results reported in Column 1 of Table 4.

*Review-to-review similarity.* As a final robustness check, we controlled for how similar a given review text was to other review texts written for the same restaurant. To do so, we relied again on Doc2Vec and computed the pairwise text similarity (using cosine similarity) across all reviews belonging to the same restaurant. Then, for each review, we computed the average similarity to the other reviews. We included this review-to-review similarity as an additional control in Equation 2 and report results in Column 4 of Table 4. Again, the focal effect of photo–text similarity continues to be positive and significant.

## *Interactive Effects on Useful Votes*

To provide insights into our proposed mechanism, we estimated two additional models in which we separately interacted similarity with review text difficulty (i.e., FKI) and image quality. We report these results in Table 5. In Column 1, we first report the results of Equation 2, including both FKI and image quality, and show that including these variables does not affect the effect of similarity on useful votes.[10] More importantly, in Column 2, we reported the results with the interaction between FKI and similarity in the model. As predicted, this interaction was negative and significant, suggesting that increasing text difficulty dampens the positive effect of photo–text similarity on useful votes. Further, in Column 3, we report the results when we included the interaction between image quality and similarity. As predicted, this interaction was positive and significant, suggesting that when images are easier to process, the similarity effect on useful votes is enhanced. These results are consistent with $H_4$.

**Table 5.** The Interactive Effects on Useful Votes.

| | (1) | (2) | (3) |
|---|---|---|---|
| Similarity | .104*** | .130*** | −.227*** |
| | (.005) | (.006) | (.040) |
| Similarity × FKI | | −.002*** | |
| | | (.0004) | |
| Similarity × avg. photo quality | | | .072*** |
| | | | (.008) |
| Log number of photos | .143*** | .143*** | .142*** |
| | (.002) | (.002) | (.002) |
| Avg. photo quality | .011*** | .011*** | −.004** |
| | (.001) | (.001) | (.002) |
| Star rating | −.050*** | −.050*** | −.050*** |
| | (.001) | (.001) | (.001) |
| Log review length | .227*** | .228*** | .227*** |
| | (.001) | (.001) | (.001) |
| FKI | .002*** | .003*** | .002*** |
| | (.00005) | (.0001) | (.00005) |
| User is local | .027*** | .027*** | .027*** |
| | (.002) | (.002) | (.002) |
| User is elite | .163*** | .163*** | .163*** |
| | (.002) | (.002) | (.002) |
| Log user reviews | .144*** | .144*** | .144*** |
| | (.001) | (.001) | (.001) |
| Restaurant fixed effects | Yes | Yes | Yes |
| Year-month fixed effects | Yes | Yes | Yes |
| Observations | 1,428,587 | 1,428,587 | 1,428,587 |
| $R^2$ | .345 | .346 | .346 |

*$p < .1$.
**$p < .05$.
***$p < .01$.
*Notes:* The dependent variable is the log of useful votes of each review. All specifications include restaurant and year-month fixed effects. Cluster-robust standard errors at the restaurant level are in parentheses.

[10] Note that FKI on its own has a positive effect on helpfulness, suggesting that more difficult text is associated with greater helpfulness. Ghose and Ipeirotis (2011) find a similar effect and suggest that reviews written in more sophisticated language may enhance the credibility and informativeness of such reviews. Importantly, our prediction is about the interaction between readability and similarity as discussed previously.

### Discussion

Overall, the results presented so far suggest that photos that more closely relate to the content of the review (and vice versa) increase the review's helpfulness ($H_1$). Further, we provide initial evidence suggesting that factors that hamper or facilitate ease of processing alter the effect of photo–text similarity. In line with $H_4$, we find that greater reading difficulty reduces the effect of photo–text similarity on helpful votes, whereas greater image quality increases the effect.

By its very nature, this study is subject to self-selection concerns and cannot establish causality or the direction of the relationship between similarity and helpfulness. In the next section, we conduct a series of lab experiments with four objectives. First, we provide evidence that algorithm-based similarity measures capture how humans judge similarity (Study 2a). Second, we address endogeneity concerns and establish a causal link between similarity and helpfulness (Studies 2b, 3, and 4). Third, we address reverse causality concerns that reviewers who write more helpful reviewers are more likely to create higher similarity reviews (Studies 3b, 3c, and 4). Fourth, we replicate these findings and provide evidence for the underlying mechanism by measuring (Studies 3a–3c) and manipulating (Study 4) ease of processing.

## Study 2a: Do Humans Perceive Similarity Akin to Algorithms?

The field data suggest that greater similarity between the text and the photo is associated with greater helpfulness. However, this result presumes that the algorithm-based measure of similarity we created in Study 1 aligns with how humans perceive similarity. Our theorizing is based on feature-based content similarity, defined as alignment between what the text describes and what the photo depicts. This approach is akin to how the representation learning algorithm assessed similarity in Study 1, that is, by quantifying the semantic similarity between photo labels extracted by Google API and the review text. Thus, we expect algorithm-based metrics to align with human-based judgments systematically. To test whether the two approaches align, we asked human judges to assess the similarity between review photos and text and compared human-judged similarity with the algorithm-based similarity from Study 1.

### Procedure and Participants

Since it is not possible to have human judges rate all reviews from Study 1, we drew five random samples from the pool of all reviews that included a single photo. Each sample consisted of 100 reviews. The overall pool was characterized by an average similarity score of .21 and a standard deviation of .11. As expected, each random sample aligned with the overall sample ($M_1 = .21$, $SD_1 = .11$; $M_2 = .21$, $SD_2 = .11$; $M_3 = .21$, $SD_3 = .11$; $M_4 = .20$, $SD_4 = .12$; $M_5 = .23$, $SD_5 = .10$).

For each sample, we recruited 100 Amazon Mechanical Turk (MTurk) workers to rate the reviews for a total of 500 raters. To ensure that human judges rated a similar distribution of photo–text similarity as assessed by the algorithm, each judge evaluated a stratified random sample of 10 randomly presented reviews from one of the five random samples of 100 reviews. Ten judges assessed each review.

Previous findings show that similarity judgments are asymmetric, and the direction of similarity comparison is determined by the relative salience of the stimuli (Tversky 1977). Because photos are generally more salient than text, we phrased our questions such that photo content was the subject of comparison and text content was the referent. To assess similarity holistically, participants first rated all assigned reviews on the question "Overall, how similar is the information conveyed in the photo to the information conveyed in the text?" (1 = "not at all," and 9 = "very"), adopted from Gentner and Markman (1994). Subsequently, participants rated feature-based similarity by answering, "With regards to *concrete features and aspects*, how similar is the information conveyed in the photo to the information conveyed in the text?" (1 = "not at all," and 9 = "very"). As suggested by prior research, both measures of similarity correlated strongly ($r_{\text{overall-feature}}(500) = .92$, $p < .001$).[11]

### Results

We found a positive and significant correlation between the algorithm-based similarity scores and overall similarity rated by human judges ($r(500) = .13$, $p = .005$). The magnitude and the direction of the relationship was similar between algorithm-based similarity and human-judged feature-based similarity ($r(500) = .13$, $p = .003$).

### Discussion

Our findings suggest that algorithm-based similarity and human-perceived similarity systematically align. The relatively small size of the correlation between algorithm-based and human-perceived similarity is expected. Machine learning approaches calculate similarity in a bottom-up approach on an attribute-by-attribute basis (Stahl 2002). However, humans use a more top-down approach and integrate their domain knowledge to generate a more holistic similarity assessment. Despite these differences, the critical implication of this study is that both methods converge: we find a positive and significant correlation indicating that the algorithm does capture meaningful aspects of human perceptions of similarity.

## Study 2b: Does Greater Similarity as Assessed by Humans Lead to Greater Usefulness?

In Study 2b, we replicate the findings from Study 1 and provide causal evidence for the proposed relationship. To provide a close

[11] For completeness, we also measured relational similarity (order counterbalanced with feature-based similarity). Empirically, this measure was highly related with holistically assessed similarity ($r_{\text{overall-relational}}(500) = .91$, $p < .001$) and feature-based similarity ($r_{\text{feature-relational}}(500) = .90$, $p < .001$).

connection with the field study, we used a set of eight reviews from Study 1 and manipulated photo–text similarity using these reviews. In line with our prediction ($H_1$) and the findings in Study 1, we expect reviews to be more helpful when humans perceive the content in the text and photo as more similar.

### Method

*Participants and exclusions.* We recruited 440 MTurk workers ($M_{age} = 39.08$ years; 54% male, 45% female, 1% other) in a 2 (similarity: similar vs. dissimilar) by 4 (review replicates) within-subjects design. We did not exclude any participants. Here and in subsequent studies, we planned for 200 participants per between-subject condition. This sample size is based on the effect size of the between-subject effect in our preliminary studies ($d = .30$) and at least 90% power. In the first two experimental studies (i.e., Studies 2b, 3a), we recruited about 10% more participants than that target sample to account for potential exclusions. Because we ended up not excluding any participants in those studies, we aimed for 200 participants per between-subject condition in subsequent studies.

*Procedure.* The 500 reviews used in Study 2a that were assessed by human judges served as our starting point for stimulus selection. As star rating influences helpfulness judgments (Mudambi and Schuff 2010), we restricted our stimulus pool to reviews with a five-star rating. From this set of 263 five-star reviews (53%), we randomly selected four reviews that were above average and four that were below average on the nine-point similarity scale used in Study 2a. The average similarity rating in the overall set of 263 reviews was $M = 6.20$ ($SD = 1.37$). To manipulate similarity, the rating of each high similarity review we chose for this study was around one standard deviation above the mean, and the rating of each low similarity review was around one standard deviation below the mean. The eight reviews and additional descriptive statistics (e.g., review length) can be found in the Web Appendix.

Participants saw all eight reviews in random order, a total of four times. Each time, participants rated the reviews on a different dimension. First, they rated all reviews on usefulness (1 = "useful," and 0 = "not useful"), replicating Yelp's usefulness votes. Then, they rated all reviews on the same similarity questions we used in Study 2a, rating overall similarity first ($r_{overall\text{-}feature}(440) = .76$, $p < .001$).[12]

### Results

*Similarity perceptions (manipulation check).* First, we conducted two separate mixed-effects models and regressed each similarity rating (overall and feature-based) on the similarity condition (1 = similar, −1 = dissimilar) with random intercepts for participants. As expected, reviews in the similar (vs. dissimilar) condition scored significantly higher on overall similarity ($M_{similar} = 7.71$, $M_{dissimilar} = 3.37$; $b = 4.34$, $SE = .09$, $t(1,319) = 46.70$, $p < .001$) and feature-based similarity ($M_{similar} = 7.55$, $M_{dissimilar} = 3.25$; $b = 4.30$, $SE = .09$, $t(1,319) = 46.31$, $p < .001$). These results suggest that our similarity manipulation was successful.

*Perceived usefulness.* Next, we conducted a mixed-effects logit model with usefulness judgment (1 = "useful," and 0 = "not useful") as the dependent variable, similarity condition as the independent variable (1 = similar, −1 = dissimilar), and random intercepts for participants. As predicted, a larger percentage (90%) of participants in the similar condition indicated the review was useful than in the dissimilar condition (67%; $b = 1.52$, $SE = .13$, $z = 11.07$, $p < .001$). Using participants' continuous similarity judgments instead of the experimental condition, we also found that both overall and feature-based similarity are positively related to usefulness votes ($r_{useful\text{-}overall}(440) = .34$, $p < .001$; $r_{useful\text{-}feature}(440) = .27$, $p < .001$).

### Discussion

This study provides initial causal evidence for our prediction that individuals find reviews more helpful when the information in the review photo and text are similar (vs. dissimilar). Using organic reviews, we manipulated similarity and provided causal evidence of its effect on helpfulness. While we carefully screened reviews on several dimensions (e.g., review length) and held these dimensions as constant as possible, a consequence of using organic reviews is that review topics differ between conditions. To ensure that idiosyncrasies of any given stimulus do not drive our results, we used multiple replicates per condition. We continue using real-life reviews in the next study (Study 3a) and provide further evidence for the proposed mechanism of ease of processing. We complement this investigation using carefully designed stimuli in Study 3b to further address any concerns about stimulus idiosyncrasies.

## Study 3a: Are Similar Reviews More Helpful Due to Greater Processing Ease?

Study 3a had two objectives. The first one was to replicate the main effect of similarity on review helpfulness ($H_1$) using two additional Yelp reviews that differed from those in Study 2b. The second one was to test $H_2$ and $H_3$, using self-reported processing ease.

### Method

*Participants and exclusions.* In a preregistered study (https://aspredicted.org/u3cs2.pdf), we recruited 439 MTurk workers ($M_{age} = 41.48$ years; 53% female, 47% male). We did not exclude any participants.

*Procedure.* Study 3a followed a two-cell (similarity: similar vs. dissimilar) between-subjects design. All participants examined

[12] As in Study 2a, following the measure of overall similarity, we measured both feature-based and relational similarity in randomized order. The relational measure was strongly correlated with both overall and feature-based similarity ($r_{overall\text{-}relational}(440) = .70$, $p < .001$; $r_{feature\text{-}relational}(440) = .73$, $p < .001$).

one of the two reviews that we selected from the subset of Yelp reviews tested in Study 2a and that were not part of Study 2b. We followed the same selection process as before and further ensured that review texts were equally readable and the photo quality was equally high between the reviews (for details and stimuli, see the Web Appendix).

After examining the review, participants rated the extent to which the review was helpful, useful, and valuable on a nine-point scale (1 = "not at all," and 9 = "very"; a = .97). We averaged them into a composite score of helpfulness. We measured the mediator—ease of processing—using a three-item self-reported measure on a seven-point scale (1 = "not at all," and 7 = "very much"; a = .93) adapted from previous work (Graf, Mayer, and Landwehr 2018): (1) "This review was easy to process," (2) "Understanding this review felt effortless," and (3) "I comprehended this review without difficulty."[13] Finally, as a manipulation check, participants rated the photo–text similarity using the same holistic similarity measure from prior studies (1 = "not at all," and 9 = "very"). To reduce common-method bias, we used scales of different lengths (nine-point, seven-point) in measuring the mediator and the dependent variable. Throughout the research, we assess and find evidence for discriminant validity between mediator and dependent variable using the criterion suggested by Fornell and Larcker (1981). These analyses (for all relevant studies we report in this article) can be found in the Web Appendix.

### Results

*Similarity perceptions (manipulation check).* Participants perceived greater similarity between review text and photo in the similar condition ($M = 7.28$, $SD = 1.61$) than in the dissimilar condition ($M = 3.65$, $SD = 2.21$; $b = 3.63$, $SE = .18$, $t(438) = 19.69$, $p < .001$).

*Perceived helpfulness.* Supporting $H_1$, participants in the similar condition ($M = 7.22$, $SD = 1.70$) rated the review as more helpful than those in the dissimilar condition ($M = 6.85$, $SD = 1.78$; $b = .37$, $SE = .16$, $t(438) = 2.25$, $p = .025$).

*Ease of processing.* Supporting $H_2$, participants in the similar condition ($M = 6.55$, $SD = .73$) rated the review easier to process than participants in the dissimilar condition ($M = 6.24$, $SD = 1.12$; $b = .31$, $SE = .09$, $t(438) = 3.49$, $p < .001$).

*Mediation.* To test $H_3$, we estimated a mediation model assessing whether the effect of review similarity on helpfulness is mediated by ease of processing. We estimated Hayes's Model 4 (using 10,000 bootstrap samples; Hayes 2017) with photo–text similarity as the independent variable (similar = 1 vs. dissimilar = −1), perceived helpfulness as the dependent variable, and self-reported ease of processing as the mediator (see Figure 8). Supporting $H_3$, we found a significant indirect effect of photo–text similarity on perceived helpfulness through ease of processing ($b_{indirect} = .28$, $SE = .08$, 95% $CI = [.13, .45]$).

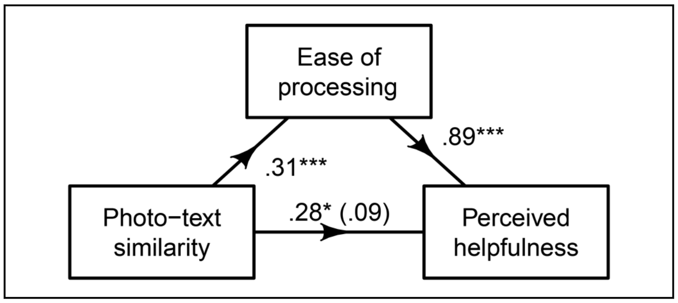


**Figure 8.** Study 3a Mediation Model.
*Notes:* The coefficient in parentheses indicates the direct effect of similarity on helpfulness.

### Discussion

Using yet another set of ecologically valid set of stimuli, Study 3a provides further causal support for the prediction in $H_1$. In addition, Study 3a provides evidence for the proposed mechanism: individuals feel greater ease of processing when the content of text and photo is similar (vs. dissimilar), and greater ease is associated with greater review helpfulness.

## Study 3b: Ruling Out Idiosyncratic Effects of Stimuli

Whereas Study 3a provides causal evidence for our predictions, one limitation is that the stimuli were vastly different across conditions (similar vs. dissimilar). Even though we screened and pretested these stimuli extensively, in Study 3a we traded off experimental control for realism. While it is unlikely that differences between any set of stimuli can account for the pattern of results in Study 3a, Study 3b rules out this alternative explanation (i.e., that text–photo differences between conditions drive the result) using more controlled stimuli.

### Method

*Participants and exclusions.* We recruited 800 MTurk workers ($M_{age} = 40.66$ years; 59% female, 40% male, 1% other). We did not exclude any participants.

*Procedure.* Study 3b followed a 2 (similarity: similar vs. dissimilar) by 2 (review replicates: coffee review, avocado toast review) between-subjects design. All participants read a review about Meno's coffee shop. Each review text discussed one of two items the coffee shop offered (coffee or avocado toast). Importantly, we paired each review text with one of two photos (coffee or avocado toast; for stimuli, see the Web Appendix). In line with our definition of feature-based

[13] We also captured the timing of the response but did not find any differences between conditions, likely because the timing measure captured both processing the review and providing the ratings.

content similarity (i.e., an alignment between what the text describes and what the picture depicts), in the similar condition, we paired the text describing the coffee's latte art with a photo showing the latte art design and the avocado toast text with a photo showing the avocado toast. In the dissimilar condition, we switched the photos and paired the latte art text with the avocado toast photo and vice versa. We used the same photo in the similar and dissimilar conditions, such that it aligned or did not align with the review text. This design rules out an alternative account that our proposed effect is driven by a particular text or a photo in the stimulus.

After examining the review, participants answered the same questions on helpfulness, ease of processing, and similarity (manipulation check) that we used in Study 3a. Items for helpfulness ($a = .98$) and ease of processing ($a = .92$) were averaged into their respective composite scores.

### Results

For all analyses in this study, we estimated linear regression models for the continuous dependent variable with similarity (dissimilar = −1, similar = 1), replicates (replicate 1 = −1, replicate 2 = 1) and their interaction as independent variables.

*Similarity perceptions (manipulation check).* Participants perceived greater similarity between review text and review photo in the similar condition ($M = 8.15$, $SD = 1.19$) than in the dissimilar condition ($M = 2.69$, $SD = 2.47$; $b = 5.55$, $SE = .20$, $t(796) = 28.41$, $p < .001$). No other effect was significant.

*Perceived helpfulness.* Supporting $H_1$, participants in the similar condition ($M = 6.17$, $SD = 2.33$) rated the review as more helpful than those in the dissimilar condition ($M = 5.43$, $SD = 2.32$; $b = .96$, $SE = .23$, $t(796) = 4.10$, $p < .001$). Neither the effect of replicates ($p > .13$) nor the interaction between similarity and replicates ($p > .18$) was significant.

*Ease of processing.* Finally, supporting $H_2$, participants in the similar condition ($M = 6.43$, $SD = .82$) rated the review easier to process than participants in the dissimilar condition ($M = 6.02$, $SD = 1.22$; $b = .50$, $SE = .10$, $t(796) = 4.77$, $p < .001$). Neither the effect of replicates ($p > .20$) nor the interaction between similarity and replicates ($p > .25$) was significant.

*Mediation.* To test $H_3$, we estimated a mediation model assessing whether the effect of review similarity on helpfulness is mediated by ease of processing. As replicates did not interact with the similarity manipulation, we collapsed across the stimulus replicates and estimated Hayes's Model 4 (using 10,000 bootstrap samples; Hayes 2017) with photo–text similarity as the independent variable (similar = 1 vs. dissimilar = −1), perceived helpfulness as the dependent variable, and self-reported ease of processing as the mediator (see Figure 9). Supporting $H_3$, we found a significant indirect effect of photo–text similarity on perceived helpfulness through ease of processing ($b_{indirect} = .11$, $SE = .02$, 95% $CI = [.06, .16]$).

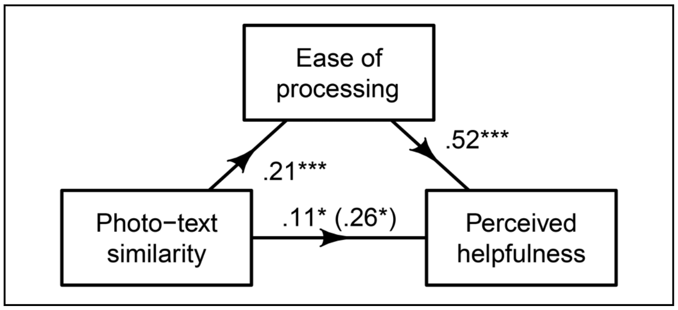


**Figure 9.** Mediation Model in Study 3b.

*Notes:* The coefficient in parentheses indicates the direct effect of similarity on helpfulness.

### Discussion

Study 3b provided further evidence of the proposed mechanism by using two stimulus replicates and using the same photo and text in both similar and dissimilar conditions. These findings suggest that it is not the text or the photo per se that drives our findings, but, rather, the similarity between the text and the photo facilitates processing and increases helpfulness. Further, using carefully designed stimulus replicates, this study rules out concerns about reviewer self-selection and reverse causality (i.e., the concern that more helpful reviewers use similarity more). Still, one could argue that because the text and the photo in the dissimilar condition referred to entities from different categories and provided such divergent information, readers may have rejected the review outright or may have considered it a mistake to be ignored. To rule out this potential concern, in the next study, we created reviews with photos and text that aligned along the conceptual category of the focal review object in both similarity conditions.

## Study 3c: Feature-Based Similarity, Not Conceptual Commonality, Affects Processing Ease and Helpfulness

So far, in most Yelp reviews and the stimuli from Study 3b, the dissimilarity between the text (e.g., describing the coffee) and the photo (e.g., showing the avocado toast) arose from the lack of commonality at the feature level as well as at the conceptual level (i.e., the conceptual category differed between modalities: drink vs. food). Prior research suggests that when the information in photo and text differs in content but is part of the same conceptual category, ease of processing increases (Newman et al. 2015). In this study, we kept the conceptual category constant (coffee) and only varied the extent to which the photo and the text aligned on features (latte art).

### Method

*Participants and exclusions.* In a preregistered study (https://aspredicted.org/9gw4j.pdf) we recruited 401 MTurk workers

($M_{age} = 42.66$ years; 50% female, 49% male, 1% other). We did not exclude any participants.

*Procedure.* Study 3c followed a two-cell (similarity: high vs. low) between-subjects design. We randomly assigned participants to one of two similarity conditions (high vs. low similarity). All participants read a review text that described the colorful latte art design at a coffee shop called Meno's. The photos in both conditions featured latte art designs and thus depicted the same conceptual category. However, photos varied by how much they aligned with the latte art described in the text. In the high similarity condition, the photo depicted a cup of coffee with a colorful latte art design exactly as described in the text. In the low similarity condition, the photo depicted a regular latte art design without colors. We expected that readers would perceive the combination of the photo showing regular latte art and the review text as less similar (vs. the combination of the photo showing colorful latte art and the same review text) and, therefore, find the review less helpful (for stimuli, see the Web Appendix). After reading the review, participants provided the same measures of perceived helpfulness ($a = .98$) and ease of processing ($a = .94$) and the manipulation check of similarity perceptions as in Studies 3a and 3b.

## Results

*Similarity perceptions (manipulation check).* As intended, participants perceived greater similarity between review text and review photo in the high similarity condition ($M = 8.33$, $SD = 1.12$) compared with the low similarity condition ($M = 5.14$, $SD = 2.77$; $b = 3.19$, $SE = .21$, $t(399) = 15.14$, $p < .001$).

*Perceived helpfulness.* Supporting $H_1$, participants in the high similarity condition ($M = 6.22$, $SD = 2.31$) perceived the review as more helpful than those in the low similarity condition ($M = 5.14$, $SD = 2.48$; $b = 1.07$, $SE = .24$, $t(399) = 4.48$, $p < .001$).

*Ease of processing.* Supporting $H_2$, participants in the high similarity condition ($M = 6.31$, $SD = .96$) rated the review easier to process than those in the low similarity condition ($M = 5.89$, $SD = 1.33$; $b = .42$, $SE = .12$, $t(399) = 3.60$, $p < .001$).

*Mediation.* To test $H_3$, we estimated a mediation model assessing whether the effect of review similarity on helpfulness is mediated by ease of processing. We estimated Hayes's Model 4 (using 10,000 bootstrap samples; Hayes 2017) with photo–text similarity as the independent variable (high similarity = 1 vs. low similarity = −1), perceived helpfulness as the dependent variable, and self-reported ease of processing as the mediator (see Figure 10). Supporting $H_3$, we found a significant indirect effect of photo–text similarity on perceived helpfulness through ease of processing ($b_{indirect} = .13$, $SE = .04$, 95% $CI = [.06, .21]$).

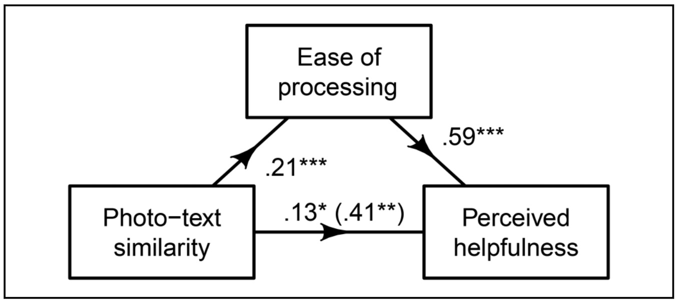


**Figure 10.** Mediation Model in Study 3c.
*Notes:* The coefficient in parentheses indicates the direct effect of similarity on helpfulness.

## Discussion

This study supports the effect of content similarity beyond conceptual commonality on review helpfulness and provides evidence for the ease of processing-based mechanism. We find that even when the photo and text are conceptually related, greater feature-based similarity between the content conveyed in the text and the photo enhances the ease with which readers process the review ($H_2$), which then increases the review's helpfulness ($H_3$).

# Study 4: Testing Boundaries of the Photo–Text Similarity Effect on Helpfulness

In Study 4, we provide further process evidence via moderation. Since similarity heightens helpfulness through ease of processing, we expect the positive effect of similarity to be attenuated when the review is difficult to process ($H_4$). Study 1 provided initial support for this moderation, focusing on each modality separately and on factors integral to the review: when the review text was linguistically difficult to process or when picture quality was low, the positive effect of photo–text similarity on helpfulness was reduced. In Study 4, we incidentally manipulate processing difficulty in a way that affects the review as a whole (i.e., affecting the processing of both the review text and the review photo). We borrow from prior research and vary processing fluency perceptually, holding the content of the review constant (Novemsky et al. 2007).

## Method

*Participants and exclusions.* This study followed a 2 (similarity: similar vs. dissimilar) by 2 (perceptual fluency: fluent vs. disfluent) between-subjects. Based on the observed effect sizes in an earlier study with a similar design but a stronger manipulation, we first collected responses from 801 participants. Given these results, following Sommet et al. (2022), we conducted additional power estimations for interactions of different forms. With close examination of the simple effects, these analyses suggested that a total sample size of $N = 1,393$ was required to detect the predicted interaction effect with .8 power. Subsequently, we conducted a second

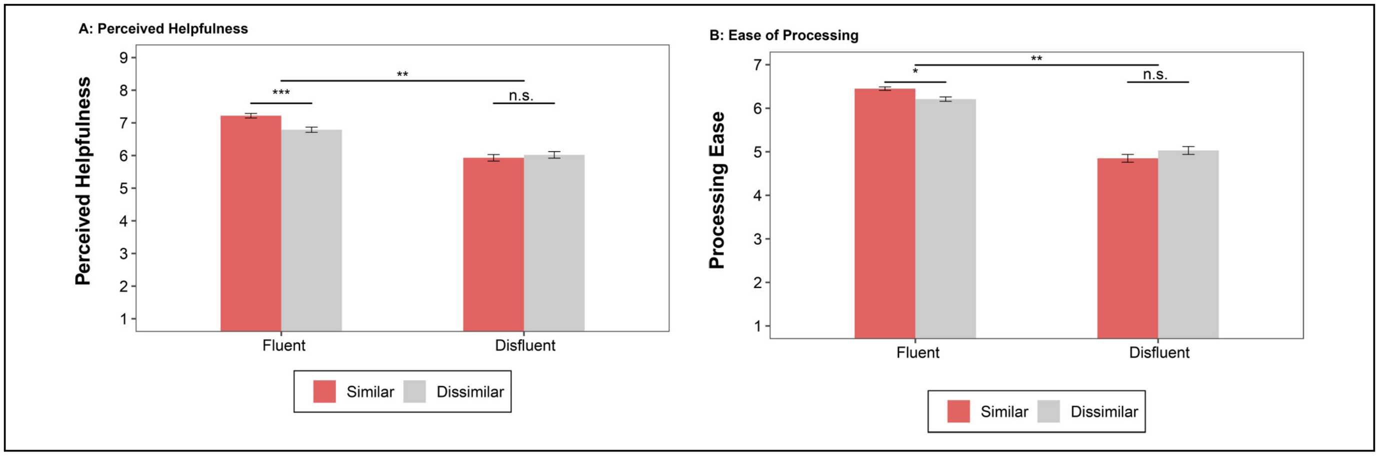


**Figure 11.** Interactive Effect of Similarity and Processing Fluency on Helpfulness (Panel A) and Ease of Processing (Panel B) in Study 4.

wave targeting 1,000 participants. We combined data from completed participants in both waves for a total 1,798 MTurk workers ($M_{age}$ = 41.70 years; 54% female, 46% male). We did not exclude any participants. When we controlled for wave (1 or 2) in the analysis, the wave had no effect on the results, and we subsequently report the analyses without this control.

*Procedure.* All participants imagined searching for a lunch place on Yelp and read the reviews used in Study 3a. Half of the participants in each similarity condition saw the review fluently: the text was written in Arial font and the photo was clear and high quality. The other half saw the same review disfluently: the text was written in Juice ITC font (Oppenheimer 2006), and the picture was 3% pixelated using Paint.net (see the Web Appendix for stimuli).

After examining the review, participants rated the review's helpfulness (a = .98) and the ease of processing using the same three items as before (a = .96). The items of each measure were averaged into a composite score of helpfulness and ease of processing, respectively.

## Results

For all analyses in this study, we estimated linear regression models for the dependent variable (helpfulness) with fluency (disfluent = −1, fluent = 1), similarity (dissimilar = −1, similar = 1), and their interaction as independent variables.

*Perceived helpfulness.* Participants in the fluent condition (M = 7.00, SD = 1.65) rated the review as more helpful than those in the disfluent condition (M = 5.97, SD = 2.11; b = .77, SE = .13, t(1,795) = 6.12, $p$ < .001). Importantly, the predicted interaction of similarity and fluency was significant (b = .51, SE = .18, t(1,795) = 2.87, $p$ = .004). When fluency was high, a similar review (M = 7.21, SD = 1.53) was more helpful than a dissimilar review (M = 6.78, SD = 1.73; b = .43, SE = .13, t(1,795) = 3.38, $p$ < .001), supporting $H_1$. However, when participants could not process the review fluently, the effect attenuated, and similar and dissimilar reviews were perceived as equally helpful ($p$ > .49; see Figure 11, Panel A).

*Ease of processing.* As expected, we found a main effect of fluency, such that participants in the fluent condition (M = 6.32, SD = .96) rated the review easier to process than those in the disfluent condition (M = 4.93, SD = 1.95; b = 1.17, SE = .10, t(1,795) = 11.51, $p$ < .001). Importantly, the interaction of similarity and fluency was also significant (b = .43, SE = .14, t(1,795) = 2.95, $p$ = .003). When fluency was high, the similar review (M = 6.45, SD = .81) was easier to process than the dissimilar review (M = 6.21, SD = 1.08; b = .23, SE = .10, t(1,795) = 2.34, $p$ = .019), supporting $H_2$. However, when processing was impaired due to disfluency, relative advantage of the similar review was reduced significantly and, unexpectedly, even directionally reversed ($M_{dissimilar}$ = 5.03, SD = 1.92, $M_{similar}$ = 4.85, SD = 1.99; $p$ = .07; see Figure 11, Panel B). We would not read too much into this reversal, as deviations in the disfluent condition were larger than in the other conditions, reflecting the heterogeneity and complexity in assessing ease of processing.

*Moderated mediation.* Though our design is one of "process by moderation," for completeness, we also estimated a moderated mediation model assessing whether the effect of review similarity on helpfulness is mediated by ease of processing only when the reviews were fluent but not when they were disfluent. We estimated Hayes's Model 8 (using 10,000 bootstrap samples, Hayes 2017) with perceived helpfulness as the dependent variable, photo–text similarity as the independent variable (dissimilar = −1
vs. similar = 1), fluency as the moderator (disfluent = −1, fluent = 1), and self-reported ease of processing as the mediator (see Figure 12). The index of moderated mediation was significant (index = .14, SE = .05, 95% CI = [.05, .24]). When the reviews were fluent, we found a significant indirect effect of photo–text similarity on perceived helpfulness through ease of processing ($b_{indirect}$ = .09, SE = .03, 95% CI = [.04, .15]).

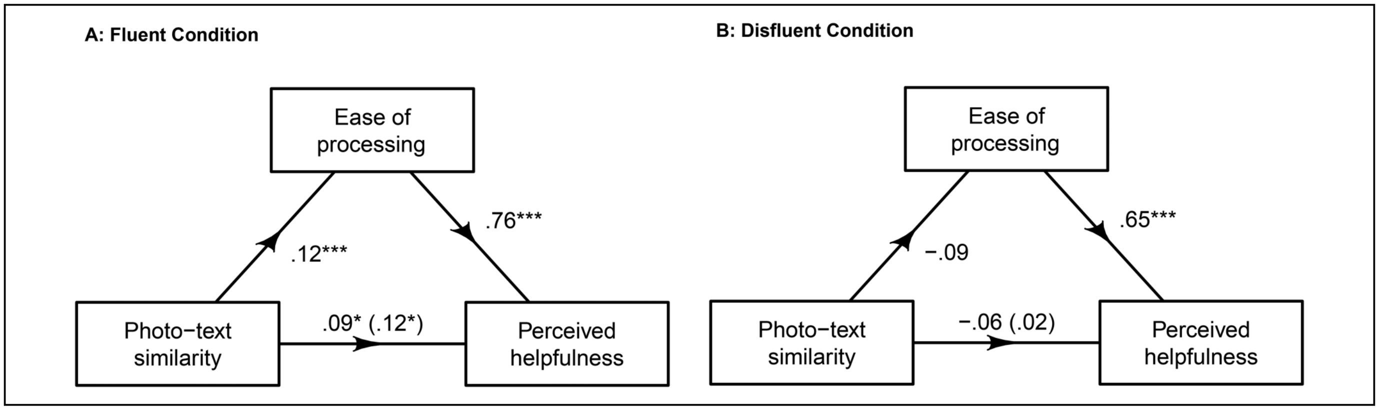


**Figure 12.** Mediation Models for Fluent (Panel A) and Disfluent (Panel B) Conditions in Study 4.
*Notes:* The coefficient in parentheses indicates the direct effect of similarity on perceived helpfulness.

The indirect effect was not significant when the reviews were disfluent ($b_{indirect} = -.06$, $SE = .04$, 95% $CI = [-.14, .02]$). Similar patterns held when we estimated Hayes's Model 7.

### Discussion

This study provides additional causal evidence for the proposed mechanism by manipulating ease of processing and identifies a boundary condition. As predicted, greater similarity between the review text and photo leads to greater helpfulness when the review is generally easy to process. However, when the review is difficult to process, this is no longer the case. Note that certain attributes of a stimulus may hamper perceptual processes (perceptual fluency), whereas others may hamper conceptual processes (conceptual fluency). Study 1 showed evidence for reduction in helpfulness due to conceptual disfluency (i.e., reading difficulty), and both Study 1 and our manipulation in this study (i.e., image quality and text font) hampered perceptual fluency and showed similar effects on helpfulness. Prior research has shown that both types of fluency have similar effects on stimulus evaluations and that perceivers often misread the influence of perceptual manipulations (e.g., readability of a print font) as difficulty in engaging in a conceptual operation (Reber, Schwarz, and Winkielman 2004). Collectively, we find that when the review is harder to process, similarity no longer has a positive effect on helpfulness.

## General Discussion

This research examines how text and photos jointly communicate information in the context of user-generated content, namely reviews of experiences. Our findings from organic reviews on Yelp suggest that greater similarity between photos and text heightens helpfulness perceptions. Since the content of reviews in our secondary data and, more importantly, the indicators of helpfulness (i.e., useful votes on Yelp) are self-selected, we validate this finding in five experiments and establish a causal link between photo–text similarity and review helpfulness. In addition, we identify the underlying mechanism and boundary condition of this effect. We find that greater similarity facilitates ease of processing of the review content; consequently, factors that impede processing ease reduce or even eliminate the effect of photo–text similarity on helpfulness.

Given the prevalence and importance of photos in online reviews, these findings provide several novel theoretical contributions. We add to the growing literature on user-generated content by examining visual–verbal content that consumers create and share extensively with others. Prior literature has focused on structured (e.g., star rating) and unstructured features (e.g., textual content) of online reviews. However, consumers frequently use visuals along with verbal information in their communication. Hence, expanding our understanding of user-generated content that includes text and photos is both timely and important. Whereas most papers that examine visuals focus on particular characteristics of the photo, we focus on the interplay between the text and the photos and examine how they jointly affect a review's perceived helpfulness. Notably, we identify photo–text similarity as an important determinant of review helpfulness that was previously unexplored.

We also contribute to the literature examining visuals in marketing communication. In the advertising literature, the effect of visual information is predominantly studied in small-sample laboratory experiments (e.g., Edell and Staelin 1983). However, this literature lacks analyses of visual–verbal information in naturally occurring settings (e.g., on online platforms). Further, the focus is generally on a single aspect of the visuals, such as shape (Lutz and Lutz 1976) or size (Rossiter and Percy 1980), rather than the interplay between words and visuals. In the future, our approach using machine learning to parse and relate visual and verbal content can be applied to large-scale data sets in advertising (e.g., the Hartman Center digitized collections at Duke University). Such examinations and our findings could provide insights into how to improve advertisement effectiveness and conduct more externally valid and theoretically grounded randomized controlled trials (Gordon et al. 2019).

Further, we identify and document an important psychological mechanism (ease of processing) that impacts the helpfulness of visual–verbal user-generated content. Our identification of this process contributes to the sizable literature identifying factors that drive review helpfulness (for a meta-analytic review, see Hong et al. [2017]). That literature has examined many review-related factors but has focused entirely on the review text. We extend this literature by examining the effect of photo–text similarity, suggesting new avenues that take visual aspects of the review into account.

By examining photo–text similarity, we also contribute to the existing literature on information processing, learning, and linguistics. Prior research in linguistics shows that repeating verbal information (e.g., providing different verbal accounts of the same event) increases comprehension and learning (Berlyne 1970). We add to this literature by suggesting that repeating the similar information in different modalities (i.e., in words and photos) makes the information easier to process and heightens its helpfulness to the reader. In addition, prior research in information processing has shown that message repetition heightens the evaluation of the focal object (Petty and Cacioppo 1986). Similarly, we show that in messages that include text and photos, the repetition of the same content in words and photos heightens the evaluation of that communication (i.e., helpfulness of the review).

From a methodological perspective, we offer a multimethod study that integrates insights from the field data with lab experiments. Recently, marketing journals have called for uniting tribes and building bridges (Grewal, Gupta, and Hamilton 2020; Peracchio, Luce, and McGill 2014) between different methodologies for more relevant (Schmitt et al. 2022) and groundbreaking research (MacInnis et al. 2020). Our research contributes to bridging this gap in the literature. We combined machine learning, human validation, and experimental design to increase the relevance and realism of our research. Further, we put significant effort into integrating organic stimuli from field data into our lab experiments to increase internal and external validity of our experimental findings and establish causality. Our method not only offers validation of the approach but also provides robust guidance for future stimuli validation efforts.

## *Managerial Implications*

Our research makes important theoretical contributions in a managerially relevant context with meaningful implications for business practice. Reviews are one key source of information for consumers. Further, businesses, large and small, put extensive effort into using and managing the information consumers convey in reviews. Academics and practitioners have focused heavily on natural language processing to gain insights from the written text (Fedewa and Holder 2022). However, photos are critical tools that consumers extensively use to communicate their experiences. Most consumers today state that they rely on visual content when making decisions and find user-generated visual content more valuable than professional content (Power Reviews 2021). Thus, we believe harvesting information jointly from both photos and text is the next frontier of gaining unique and novel customer insights. Our investigation is a first step in this direction.

We focus on review helpfulness as the focal outcome because of its critical impact on downstream consequences, including sales. In this context, our findings suggest that the interplay between visual and verbal content matters. These findings allow review sites to guide consumers on the *type* of photo and the *structure* of the review text that accompanies a photo to increase a review's value to others. Notably, review sites can nudge consumers to convey similar content in the text and photo. Our Yelp data suggest that, on average, consumers already create reviews with text and photos conveying somewhat similar content. What may be counterintuitive to consumers and managers alike is that a photo is *not* "worth a thousand words"; that is, a photo does not substitute for the text. Rather, readers of a review find that review more helpful when both visual and verbal content conveys similar information.

Further, we find that ease of processing causally affects review helpfulness. Factors such as photo quality or text readability that many retail and review sites can control impact the effect of similarity on helpfulness. Thus, these sites may nudge consumers to create easy-to-process reviews, by encouraging them to upload high-quality, easy-to-grasp photos and use easy-to-understand language. Sites may be able to improve photo quality automatically when consumers upload a photo (e.g., employing sharpening tools on photos). Further, as consumers write their reviews, review platforms may point out difficult language and suggest simpler language, which, based on our findings, would boost the effect of similarity on helpfulness.

Finally, platforms can use our findings to identify helpful reviews even before the reviews receive helpfulness votes from readers, an effort that has garnered significant interest in the past, particularly in computer science.

## *Limitations and Future Research*

We see our investigation as the first large-scale, systematic examination of visual–verbal user-generated content. As such, our research leads to many open questions for future study.

*Types of alignment between photo and text.* Our investigation focused on feature-based similarity between photo and text. However, an alignment between the photo and text may exist also on other higher-level dimensions, referred to as coherence (Winkielman et al. 2012). One such higher-level alignment that is integral to the review context is valence. For example, the text of a review could praise the beautiful decoration, but the photo may suggest otherwise (or vice versa). Such conflicting information could confuse the reader and result in processing difficulty, ultimately reducing helpfulness of the review, which future research may examine. Further research may also examine how people resolve such conflict (i.e., when valence differs between visual and verbal content). As visuals generate more concrete and memorable representations than words (Paivio 1969), one may predict that the valence of the photo gets more weight when readers evaluate the experience in the review. Or, in today's age, people may be skeptical about the validity of the photo, as photos can be easily manipulated. Hence, they may discount the photo when evaluating the experience based on conflicting information in the photo and text. Future research may disentangle these different effects.

Another higher-level alignment that may be worth investigating is between the review content and the context. Even if the photo and text align on content and valence, such a review may still be incongruent with the context (e.g., posting a washing machine review with high content and valence similarity on a restaurant review site such as Yelp). Such incongruence would unlikely be helpful because it is not congruent with the goal review readers are trying to achieve. Entirely incongruent reviews like the previous example are rare in real life. However, reviews may be incongruent with the goal of the reader in other, more subtle ways. In those cases, the effect of feature-based content similarity between photo and text may be reduced or eliminated. For example, photos that include people may help consumers shopping for more experiential purchases, but not for more material purchases (Ding et al. 2021).

*Nature of the relationship between similarity and helpfulness.* We predict and find a linear relationship between feature-based content similarity and helpfulness in the Yelp data. In part, this is due to extreme similarity/dissimilarity being sparse in our field data, and hence the best-fitting model capturing this relationship is a linear one. According to our proposed mechanism, perceived ease of processing, it is possible that similarity has a diminishing return on ease of processing. It is also possible that, when one modality can convey the experience clearly, the other modality may become redundant, and thus extreme similarity may hurt helpfulness. Future research may examine these potential boundary conditions.

*Content photos can express.* We chose restaurant reviews as our empirical context. While most reviews are centered on food, some also mention other aspects of the restaurant experience (e.g., location, decoration, ambience). We believe our findings are not restricted to food settings. They are, however, limited to what photos can express. Photos can convey anything tangible and visual very well, but they are less apt at conveying nonvisual information. For example, photos are able to communicate the taste of food only partially (e.g., they may communicate the spiciness of a dish by capturing the hot peppers that are visible) or completely unable to convey other aspects (e.g., the noisiness of a place). If it is possible to visualize the experience at least partially, our findings should still apply. For example, Latour and Deighton (2019) find that creating visuals of taste experiences (i.e., wine) allows people to be transported more into their own taste experience. Hence, the similarity effect we found in our research may extend to partially visualizable dimensions, which future research can address.

People increasingly communicate using both visual and verbal information. In 2023, people will have taken 1.81 trillion photos of their experiences (Broz 2023). According to these trends, visual communication will continue to be central to user-generated and user-shared content. Our research is one of the first to examine the interplay between photos and text in user-generated content. We hope our findings will open many novel avenues for future research in this area.

## Acknowledgments

The authors thank Poet Larsen and Kaihang Zhao for their invaluable help with Study 1. This article is based on an essay in Gizem Ceylan's dissertation conducted under the direction of the other two authors.

## Authors Contributions

All authors contributed equally, and authorship is alphabetical by last name.

## Coeditor

Maureen Morrin

## Associate Editor

Jacob Goldenberg

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Gizem Ceylan https://orcid.org/0000-0002-3876-9312

## References

Alter, Adam L. and Daniel M. Oppenheimer (2009), "Uniting the Tribes of Fluency to Form a Metacognitive Nation," *Personality and Social Psychology Review*, 13 (3), 219–35.

Alter, Adam L., Daniel M. Oppenheimer, Nicholas Epley, and Rebecca N. Eyre (2007), "Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning," *Journal of Experimental Psychology: General*, 136 (4), 569–76.

Amit, Elinor, Daniel Algom, and Yaacov Trope (2009), "Distance-Dependent Processing of Pictures and Words," *Journal of Experimental Psychology: General*, 138 (3), 400–415.

Aslam, Salman (2023), "Snapchat by the Numbers: Stats, Demographics & Fun Facts," Omnicore (March 23), https://www.omnicoreagency.com/snapchat-statistics/.

BazaarVoice (2021), "The 101 on UGC," (accessed October 2, 2022), https://www.bazaarvoice.com/resources/the-101-on-ugc/.

Berlyne, D.E. (1970), "Novelty, Complexity, and Hedonic Value," *Perception & Psychophysics*, 8 (5), 279–86.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119 (1), 249–75.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3 (4–5), 993–1022.

Broz, Matic (2023), "Number of Photos (2023): Statistics, Facts, & Predictions," (March 10), https://photutorial.com/photos-statistics/.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345–54.

Clark, Ruth C. and Richard E. Mayer (2016), *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, 4th ed. John Wiley & Sons.

Ding, Mengqi (Annie), Shirley (Shuo) Chen, Xin (Shane) Wang, and Neil Bendle (2021), "Show Me You or the Goods? Effect of Image Content on Review Helpfulness," in *ACR North American Advances*, Vol. 49, Tonya Williams Bradford, Anat Keinan, and Matthew Thomson, eds. Association for Consumer Research, 138–40.

Driscoll, James M. and John T. Lanzetta (1964), "Effects of Problem Uncertainty and Prior Arousal on Pre-Decisional Information Search," *Psychological Reports*, 14 (3), 975–88.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "The Dynamics of Online Word-of-Mouth and Product Sales—An Empirical Investigation of the Movie Industry," *Journal of Retailing*, 84 (2), 233–42.

Edell, Julie A. and Richard Staelin (1983), "The Information Processing of Pictures in Print Advertisements," *Journal of Consumer Research*, 10 (1), 45.

Fedewa, Dave and Chauncey Holder (2022), "Why Business Must Heed Customer Reviews," McKinsey (February 22), https://www.mckinsey.com/capabilities/operations/our-insights/why-business-must-heed-customer-reviews.

Forman, Chris, Anindya Ghose, and Batia Wiesenfeld (2008), "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, 19 (3), 291–313.

Fornell, Claes and David F. Larcker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (1), 39–50.

Gentner, Diedre and Arthur B. Markman (1994), "Structural Alignment in Comparison: No Difference Without Similarity," *Psychological Science*, 5 (3), 152–58.

Ghose, Anindya and Panagiotis G. Ipeirotis (2011), "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Transactions on Knowledge and Data Engineering*, 23 (10), 1498–512.

Gordon, Brett R., Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky (2019), "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook," *Marketing Science*, 38 (2), 193–225.

Graf, Laura K.M., Stefan Mayer, and Jan R. Landwehr (2018), "Measuring Processing Fluency: One Versus Five Items," *Journal of Consumer Psychology*, 28 (3), 393–411.

Grewal, Rajdeep, Sachin Gupta, and Rebecca Hamilton (2020), "The *Journal of Marketing Researc*h Today: Spanning the Domains of Marketing Scholarship," *Journal of Marketing Research*, 57 (6), 985–98.

Harp, Shannon F. and Richard E. Mayer (1998), "How Seductive Details Do Their Damage: A Theory of Cognitive Interest in Science Learning," *Journal of Educational Psychology*, 90 (3), 414–434.

Hartmann, Jochen, Mark Heitmann, Christina Schamp, and Oded Netzer (2021), "The Power of Brand Selfies," *Journal of Marketing*, 58 (6), 1159–77.

Hayes, Andrew F. (2017), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Press.

Hershenson, Maurice and Ralph Norman Haber (1965), "Role of Meaning in the Perception of Briefly Exposed Words," *Canadian Journal of Psychology*, 19 (1), 42–46.

Hong, Hong, Di Xu, G. Alan Wang, and Weiguo Fan (2017), "Understanding the Determinants of Online Review Helpfulness: A Meta-Analytic Investigation," *Decision Support Systems*, 102, 1–11.

Jacoby, Larry L. and Mark Dallas (1981), "On the Relationship Between Autobiographical Memory and Perceptual Learning," *Journal of Experimental Psychology: General*, 110 (3), 306–40.

Kelley, Colleen M. and D. Stephen Lindsay (1993), "Remembering Mistaken for Knowing: Ease of Retrieval as a Basis for Confidence in Answers to General Knowledge Questions," *Journal of Memory and Language*, 32 (1), 1–24.

Kieras, David (1978), "Beyond Pictures and Words: Alternative Information-Processing Models for Imagery Effect in Verbal Memory," *Psychological Bulletin*, 85 (3), 532–54.

Kim, Soo Min, Patrick Pantel, Tim Chklovski, and Marco Pennacchiotti (2006), "Automatically Assessing Review Helpfulness," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 423–30.

Kincaid, J. Peter, Robert P. Fishburne, Richard L. Rogers, and Brad S. Chissom (1975), "Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel," Research Branch Report 8-75, *Naval Technical Training Command*.

Kronrod, Ann and Shai Danziger (2013), "'Wii Will Rock You!' The Use and Effect of Figurative Language in Consumer Reviews of Hedonic and Utilitarian Consumption," *Journal of Consumer Research*, 40 (4), 726–39.

Latour, Kathryn A. and John A. Deighton (2019), "Learning to Become a Taste Expert," *Journal of Consumer Research*, 46 (1), 1–19.

Le, Quoc and Tomas Mikolov (2014), "Distributed Representations of Sentences and Documents," *ICML '14: Proceedings of the 31st International Conference on Machine Learning*, 32, 2931–39.

Lee, Sangjae and Joon Yeon Choeh (2020), "The Impact of Online Review Helpfulness and Word of Mouth Communication on Box Office Performance Predictions," *Humanities and Social Sciences Communications*, 7 (1), 1–12.

Li, Yiyi and Ying Xie (2020), "Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social

Media Engagement," *Journal of Marketing Research*, 57 (1), 1–19.

Lutz, Kathy A. and Richard J. Lutz (1976), "Effects of Interactive Imagery on Learning: Application to Advertising," *Journal of Applied Psychology*, 62 (4), 493–98.

MacInnis, Deborah J., Vicki G. Morwitz, Simona Botti, Donna L. Hoffman, Robert V. Kozinets, Donald R. Lehmann, John G. Lynch Jr., and Cornelia Pechmann (2020), "Creating Boundary-Breaking, Marketing-Relevant Consumer Research," *Journal of Marketing*, 84 (2), 1–23.

Markman, Arthur B. and Dedre Gentner (1993), "Splitting the Differences: A Structural Alignment View of Similarity," *Journal of Memory and Language*, 32 (4), 517–35.

Markman, Arthur B. and Dedre Gentner (1996), "Commonalities and Differences in Similarity Comparisons," *Memory & Cognition*, 24 (2), 235–49.

Meyers-Levy, Joan and Alice M. Tybout (1989), "Schema Congruity as a Basis for Product Evaluation," *Journal of Consumer Research*, 16 (1), 39–54.

Mudambi, Susan M. and David Schuff (2010), "Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com," *MIS Quarterly*, 34 (1), 185.

Newman, Eryn J., Maryanne Garry, Christian Unkelbach, Daniel M. Bernstein, D. Stephen Lindsay, and Robert A. Nash (2015), "Truthiness and Falsiness of Trivia Claims Depend on Judgmental Contexts," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41 (5), 1337–48.

Novemsky, Nathan, Ravi Dhar, Norbert Schwarz, and Itamar Simonson (2007), "Preference Fluency in Choice," *Journal of Marketing Research*, 44 (3), 347–56.

Oppenheimer, Daniel M. (2006), "Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly," *Applied Cognitive Psychology*, 20 (2), 139–56.

Paivio, Allan (1969), "Mental Imagery in Associative Learning and Memory," *Psychological Review*, 76 (3), 241–63.

Peracchio, Laura A., Mary Frances Luce, and Ann L. McGill (2014), "Building Bridges for an Interconnected Field of Consumer Research," *Journal of Consumer Research*, 40 (6), v–viii.

Petty, Richard E. and John T. Cacioppo (1986), "The Elaboration Likelihood Model of Persuasion," in *Communication and Persuasion*, Advances in Experimental Social Psychology, Vol. 19, L. Berkowitz, ed. Springer, 123–205.

Philipp-Muller, Aviva, John P. Costello, and Rebecca Walker Reczek (2022), "Get Your Science Out of Here: When Does Invoking Science in the Marketing of Consumer Products Backfire?" *Journal of Consumer Research*, 49 (5), 721–40.

Pierce, Karen A. and Barry Gholson (1994), "Surface Similarity and Relational Similarity in the Development of Analogical Problem Solving: Isomorphic and Nonisomorphic Transfer," *Developmental Psychology*, 30 (5), 724–37.

Pieters, Rik and Michel Wedel (2004), "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects," *Journal of Marketing*, 68 (2), 36–50.

Power Reviews (2021), "How User-Generated Visual Content Became Fundamental to Ecommerce," (accessed September 30, 2022), https://www.powerreviews.com/insights/how-visual-ugc-became-fundamental/.

Reber, Rolf and Norbert Schwarz (1999), "Effects of Perceptual Fluency on Judgments of Truth," *Consciousness and Cognition*, 8 (3), 338–42.

Reber, Rolf, Norbert Schwarz, and Piotr Winkielman (2004), "Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience?" *Personality and Social Psychology Review*, 8 (4), 364–82.

Reber, Rolf, Piotr Winkielman, and Norbert Schwarz (1998), "Effects of Perceptual Fluency on Affective Judgments," *Psychological Science*, 9 (1), 45–8.

Reich, Taly and Sam J. Maglio (2020), "Featuring Mistakes: The Persuasive Impact of Purchase Mistakes in Online Reviews," *Journal of Marketing*, 84 (1), 52–65.

Rossiter, John R. and Larry Percy (1980), "Attitude Change Through Visual Imagery in Advertising," *Journal of Advertising*, 9 (2), 10–16.

Ryu, Sann, Yun-na Park, and Jungkun Park (2022), "Looks Clear and Sounds Familiar: How Consumers Form Inferential Beliefs About Luxury Hotel Service Quality," *Cornell Hospitality Quarterly*, 63 (1), 48–65.

Schmitt, Bernd H., June Cotte, Markus Giesler, Andrew T. Stephen, and Stacy Wood (2022), "Relevance—Reloaded and Recoded," *Journal of Consumer Research*, 48 (5), 753–55.

Schwartz, Ali (2019), "Why It's Important to Upload Quality Photos on Yelp," Yelp (July 19), https://blog.yelp.com/businesses/businesses-upload-photos-yelp/.

Shulman, Hillary C., Graham N. Dixon, Olivia M. Bullock, and Daniel Colón Amill (2020), "The Effects of Jargon on Processing Fluency, Self-Perceptions, and Scientific Engagement," *Journal of Language and Social Psychology*, 39 (5–6), 579–97.

Sommet, Nicolas, David Weissman, Nicolas Cheutin, and Andrew J. Elliot (2022), "How Many Participants Do I Need to Test an Interaction? Conducting an Appropriate Power Analysis and Achieving Sufficient Power to Detect an Interaction," *Advances in Methods and Practices in Psychological Science* (forthcoming), https://osf.io/xhe3u/.

Song, Hyunjin and Norbert Schwarz (2008), "If It's Hard to Read, It's Hard to Do: Processing Fluency Affects Effort Prediction and Motivation," *Psychological Science*, 19 (10), 986–88.

Stahl, Armin (2002), "Defining Similarity Measures: Top-Down vs. Bottom-Up," in *Advances in Case-Based Reasoning*, Lecture Notes in Computer Science, Vol. 2416, Susan Craw and Alun Preece, eds. Springer, 406–20.

Syed, Shaheen and Marco Spruit (2017), "Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation," in *2017 IEEE International Conference on Data Science and Advanced Analytics*. IEEE, 165–74.

Talebi, Hossein and Peyman Milanfar (2018), "NIMA: Neural Image Assessment," *IEEE Transactions on Image Processing*, 27 (8), 3998–4011.

Topaloglu, Omer and Mayukh Dass (2021), "The Impact of Online Review Content and Linguistic Style Matching on New Product Sales: The Moderating Role of Review Helpfulness," *Decision Sciences*, 52 (3), 749–75.

Tversky, Amos (1977), “Features of Similarity,” *Psychological Review*, 84 (4), 327–52.

Villarroel Ordenes, Francisco, Dhruv Grewal, Stephan Ludwig, Ko De Ruyter, Dominik Mahr, and Martin Wetzels (2019), “Cutting Through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social media Brand Messages,” *Journal of Consumer Research*, 45 (5), 988–1012.

Winkielman, Piotr, Jamin Halberstadt, Tedra Fazendeiro, and Steve Catty (2006), “Prototypes Are Attractive Because They Are Easy on the Mind,” *Psychological Science*, 17 (9), 799–806.

Winkielman, Piotr, David E. Huber, Liam Kavanagh, and Norbert Schwarz (2012), “Fluency of Consistency: When Thoughts Fit Nicely and Flow Smoothly,” in *Cognitive Consistency: A Fundamental Principle in Social Cognition,* Bertram Gawronski and Fritz Strack, eds. Guilford Press, 89–111.

Winkielman, Piotr, Norbert Schwarz, Tetra Fazendeiro, and Rolf Reber (2003), “The Hedonic Marking of Processing Fluency: Implications for Evaluative Judgment,” in *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, Jochen Musch and Karl C. Klauer, *eds*., Lawrence Erlbaum Associates, 189–217.

Zhang, Shunyuan, Dokyun Lee, Param Vir Singh, and Kannan Srinivasan (2022), “What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features,” *Management Science*, 68 (8), 5644–66.