# Redundancy in Visual-Verbal Word of Mouth: How Photos Increase Visual Language Use

GIZEM CEYLAN AND KRISTIN DIEHL

**ABSTRACT**    With camera-enabled phones, consumers can share their experiences not just in words but also in photos. Drawing on Grice's conversational maxims, this article asks: Do photos substitute for words, following the maxim of quantity, as "a picture is worth a thousand words" suggests? Or following the maxim of relation, do people communicate redundantly across modalities by expressing visual aspects both visually and verbally? Across two large-scale data sets of more than 6.9 million consumer reviews from Yelp and TripAdvisor, two validation studies with human judges, and a controlled experiment, we find that reviewers generally communicate redundantly: While photos naturally convey visual aspects of experiences, reviewers also communicate these aspects in their text, creating redundancy across modalities. This tendency is heightened when reviewers begin by selecting a photo and when they are concerned about how their message is received by readers.

Consumers frequently review and share experiences with others (e.g., a meal or vacation), creating word of mouth (WOM). A large body of work on WOM has examined the verbal content of these conversations (Chevalier and Mayzlin 2006; Forman et al. 2008; Moore 2012; Kronrod et al. 2023). Yet with the widespread adoption of camera-enabled phones, consumers increasingly create multimodal reviews that combine words and photos. We examine this type of visual-verbal WOM, that is, consumer-initiated communication that includes both visual (i.e., photos) and verbal (i.e., text) elements of their experiences.

Prior communication theories suggest that conversational norms guide how people share information with others. Grice's (1975) conversational maxims highlight two fundamental yet contradictory principles: the maxim of quantity (being brief and avoiding redundancy) and the maxim of relation (being relevant and emphasizing important details). We examine whether and to what extent these norms apply to visual-verbal communication. Specifically, we ask: Do photos substitute for words, following the maxim of quantity, as the saying "a picture is worth a thousand words" suggests? Or do communicators engage in redundancy across modali-

ties, following the maxim of relation, highlighting visual aspects verbally, as suggested by the expression "show and tell"?

To examine these questions, we analyze consumer reviews from Yelp and TripAdvisor and validate our large-scale text-analyses with human evaluations of a subset of reviews. We find that reviewers generally communicate redundantly: While photos naturally convey visual aspects of one's experience, reviewers also communicate visual aspects in their text, creating redundancy across modalities. We also conduct an experiment to examine two potential moderators: an external, platform-driven factor (i.e., photo selection order), and an internal, communicator-driven factor (i.e., impression goals). When reviewers begin by selecting a photo or focus on how their message is received by readers, they engage in greater redundancy.

Our findings make several important contributions. First, we contribute to research on WOM that has focused largely on verbal communication. We examine how people adapt the verbal parts of their communication when they can convey their experiences both verbally and visually.

Second, we extend research on conversational norms by examining their role in visual-verbal WOM. Our findings

reveal that communicators prioritize redundancy when communicating visually and verbally and that this tendency is heightened by platform- and consumer-driven factors.

Third, we build on research examining the linguistic content and structure of WOM. Prior work identified antecedents of using different types of words (e.g., authentic experiences are expressed using more concrete words; Kronrod et al. 2023). We add to this literature two new drivers of linguistic variation that increase consumers' use of visual language (i.e., visual words): the presence of photos and active impression goals. Integrating theories from linguistics and psychology and using a multi-method approach (computational text analyses, controlled experiments), we provide novel insights into what shapes the creation of visual-verbal WOM.

## CREATING VISUAL-VERBAL WORD OF MOUTH

Consumers share their experiences of products and services through conversations with others. In recent years many scholars have focused on such conversations in the form of user-generated reviews, examining the effects of "what" reviewers say (e.g., type of words used) and "how" they say it (e.g., length and narrative structure) on readers (see Moore and Lafreniere 2020 for a review). While the presence of photos is important for how reviews are received (e.g., Ceylan et al. 2024), prior investigations have largely overlooked whether and how adding visuals in WOM affects verbal content. By examining *visual-verbal WOM creation*, this article sheds light on a prevalent yet under-researched form of communication.

### Conversational Norms in Word of Mouth

Prior research characterizes verbal communication as a cooperative exchange between communicators and receivers, governed by norms (Sperber and Wilson 1986; Clark 1992; Schwarz 1994). Grice's cooperative principles (1975) outline maxims (norms) by which conversational partners establish what information is mutually available, how much detail is expected, and what constitutes redundant information. We focus on two fundamental, yet contradictory norms of conversation: quantity and relation.

Grice's maxim of quantity states that communicators should be brief. They should provide just the appropriate amount of information, avoiding being overly detailed and redundant (Grice 1975). Reducing redundancy benefits receivers by preventing information overload and helps communicators minimize effort (Goodman and Frank 2016; Zipf 1949). In visual-verbal WOM, this principle suggests that communicators should use each modality to convey different aspects of an experience, minimizing overlap between modalities. For example, instead of describing the color of the sky during sunset, a communicator might rely on a photo to convey this visual aspect of the experience.

In contrast, Grice's maxim of relation posits that communicators should be relevant to the audience. One way to accomplish this is to emphasize details in different ways to minimize misinterpretation and help the receiver recognize the relevance of key details (Wilson 1993). In visual-verbal WOM, this principle may suggest that communicators might emphasize visual aspects of their experience in both photos and text. For example, they may share an image of a sunset and also describe the vivid colors of the sky in words. This type of redundancy across modalities can facilitate fluent interpretation of the review, which in turn enhances audience reception and perceived helpfulness (Ceylan et al. 2024).

Past research suggests that communicators prioritize redundancy, particularly in environments where the receivers' attention is fragmented (Partan and Marler 1999), message clarity is critical, and message transmission may be challenging. Given the dearth of attention in today's world (Davenport and Beck 2001) and the relatively low cost of using both words and photos, we expect consumers to prioritize the maxim of relation (i.e., creating redundancy) over the maxim of quantity.

> **H1:** In visual-verbal WOM, communicators offer redundant content, referencing visual aspects in their text even when photos are present.

### Impact of Platform Design

Platforms make a myriad of design decisions when soliciting reviews. They determine who can post (e.g., verified vs. unverified; Mayzlin et al. 2014), whether to display reviewer distinctions (i.e., badges; Zhang et al. 2020), and the order in which review elements are solicited (e.g., star ratings before or after writing text; Aerts et al. 2017). These decisions can impact the linguistic characteristics and overall quality of reviews.

We focus on one platform decision: the order in which photos and text are solicited. Some platforms prompt reviewers to select photos before writing text (e.g., Instagram), while other platforms prompt text first (e.g., Google Reviews). We posit that selecting a photo before writing text should increase redundancy. The mere selection of a photo can affect how people recall and reconstruct the experience

(Diehl et al. 2025). Further, because photos naturally draw attention to visual details (Pieters and Wedel 2004) and evoke emotional reactions to those details (Rossiter and Percy 1980), the photo may serve as a visual anchor during text creation. As photos make visual elements more cognitively accessible and emotionally salient, reviewers who select a photo before writing (vs. after) should reference more visual elements in their text, increasing redundancy.

> **H2:** If communicators select a photo before (vs. after) creating a verbal narrative, they will communicate more redundantly.

### Impact of Reviewers' Goals

When consumers create online WOM, they may pursue different goals. Some research suggests individuals derive hedonic benefits from self-expression in recommendation context (Yeomans 2019), while other work distinguishes between self-presentation goals, i.e., presenting oneself in the best way to impress others, and self-disclosure goals, i.e., presenting oneself in an authentic way (Schlosser 2020).

In the context of reviewing, sharing experiences with many (i.e., broadcasting; Barasch and Berger 2014) is likely to activate impression goals, particularly because online (vs. offline) WOM offers greater opportunity to self-enhance (Eisingerich et al. 2015). When people hold an impression goal, they become attuned to how they may be perceived by others. They strategically manage their self-presentation, for instance by avoiding sharing information that reflects negatively on them (Barasch and Berger 2014) such as content that may be perceived as bragging (Valsesia and Diehl 2022). This heightened sensitivity to audience impressions makes communicators more deliberate in their communication (Zhang et al. 2020).

In visual-verbal WOM, impression-concerned communicators may be especially motivated to ensure that key aspects of their experience are noticed, found relevant, and appreciated by their audience. When communicating through both photos and text, redundancy serves this goal by reinforcing important information across modalities, reducing the risk of miscommunication and increasing the likelihood that their message resonates with readers. As a result, communicators with impression goals may strategically communicate redundantly to make a favorable impression.

> **H3:** Communicators with impression goals communicate more redundantly.

Impression goals may be situationally activated (e.g., by audience size; Barasch and Berger 2014) or chronically held (e.g., due to communicator's role; Leary and Kowalski 1990). Elite reviewers on platforms like Yelp likely exemplify the latter. These reviewers receive visible recognition for their status. Recognition heightens reviewers' concerns about how their content is received, leading them to produce more deliberate and strategic contributions (Burtch et al. 2022). They write longer and more polished reviews than nonelite users (Zhang et al. 2020) and are often motivated to maintain their designation by being helpful to readers (Mathwick and Mosteller 2017). This chronic concern with impression management, reinforced by their desire to maintain their designation, should lead to more redundant communication in line with hypothesis 3.

## OVERVIEW OF STUDIES

We test our predictions in four studies using restaurant reviews. Study 1 analyzes over 6.9 million consumer reviews on Yelp and TripAdvisor, examining whether photo presence predicts greater visual language use, a proxy for redundancy. We use two complementary text analysis approaches: *word-frequency* using the LIWC visual category, and a *word-embedding* method. Controlling for review characteristics (e.g., review length, star rating, device type, restaurant type, review date), we find that photo presence predicts greater visual language use (hypothesis 1) and that elite (vs. nonelite) reviewers use more visual language overall (hypothesis 3).

Studies 2A and 2B use human judges to validate and extend these computational findings among a subset of Yelp reviews used in study 1. Study 2A finds that human judges perceive review text originally accompanied by photos as more focused on visual details (hypothesis 1), consistent with the computational findings. Humans also perceived review text written by elite reviewers was also perceived as including more visual language. Study 2B focuses exclusively on reviews that were originally accompanied by a photo and directly measures photo-text redundancy. Results show that elite reviewers exhibit greater redundancy (hypothesis 3). Study 3 tests the role of platform design and situational impression goals. The results indicate that individuals communicate more redundantly when they start by selecting a photo (vs. writing text, hypothesis 2) and when they are more concerned about how their review is received (hypothesis 3). Data, code, and the study instruments for studies 2A, 2B, and 3 can be found at https://researchbox.org/261&PEER_REVIEW_passcode=ZCAMWM.

## STUDY 1: TESTING FOR REDUNDANT COMMUNICATION IN NATURALLY OCCURRING WOM

Using two real-world data sets of online restaurant reviews, we explore the extent to which reviewers communicate redundantly, i.e., use language that focuses on visual details (hypothesis 1). As a test of hypothesis 3, we also examine whether elite users (e.g., those who review more and/or are designated via a badge) use more visual language.

### Data Sets

We examined two large, real-world data sets of restaurant reviews from Yelp and TripAdvisor. Both data sets included reviews for restaurants located in Los Angeles County, written between 2004 and 2019. The Yelp data set includes 6,713,325 reviews, and the TripAdvisor data set includes 245,569 reviews. On both platforms, when users write reviews, they can upload photos as part of their reviews. These user-submitted photos are displayed with the review text, creating integrated visual-verbal content from the same reviewer. While users can also upload standalone photos to a business's photo gallery without writing a review, our investigation focuses exclusively on individual reviews, allowing us to examine how individuals deliberately combine visual and verbal elements when sharing their experiences.

For each review, we collected the review text as well as several control variables: (*a*) photos uploaded by the reviewer, (*b*) star rating, (*c*) reviewer status (elite vs. not, available only on Yelp), (*d*) total number of reviews written by the reviewer, (*e*) device type (mobile vs. not, available only on TripAdvisor), (*f*) reviewer type (international vs. local, available only on TripAdvisor), (*g*) restaurant characteristics (e.g., cuisine), and (*h*) date of review. We included only reviews written in English. Descriptive statistics are provided in tables 1 and 2.

### Automated Text-Analyses

**Word-Frequency Method.** To measure visual language in reviews, we used the Linguistic Inquiry and Word Count program (LIWC; Tausczik and Pennebaker 2010), a dictionary-based approach widely used and extensively validated in marketing and consumer research (Berger et al. 2020; Humphreys and Wang 2018). Our analyses focused on the visual category, a subcategory under the perceptual process dimension that includes words describing visual features of objects and scenes, such as *blue, colorful, round, dark* (app. A (available online) for all words in that category). Assessing the relative frequency of these words in each review (i.e., visual word count/total word count) serves as a measure of visual language use.

**Word-Embedding Method.** While the frequency method is useful, one limitation is that it relies on exact word matches and can miss conceptually relevant words that are absent from the predefined dictionary. To address this limitation and ensure the robustness of our results, we used word embeddings, a computational linguistics approach that captures the semantic similarity between words based on the contexts in which they appear (Mikolov et al. 2013).

Specifically, we applied the distributed dictionary method (Garten et al. 2018), which leverages word embeddings to measure the semantic distance (i.e., similarity in meaning) between a review and a predefined construct characterized by a set of words (Bhatia et al. 2019). We began with the predefined visual category from LIWC as our conceptual anchor and used distributional semantics to extend this set into a richer, high-dimensional representation of the underlying construct. This helped identify semantically related language based on patterns of word use in large text corpora. Each review was then evaluated based on how similar its language was to this extended representation, allowing us to capture

Table 1. Yelp Data Set: Descriptive Statistics

| *N* = 6,713,325 reviews in Yelp | Mean | SD | Minimum | Maximim |
|---|---|---|---|---|
| Average number of words per review | 97.2 | 92.3 | 1 | 1,134 |
| Share of visual words per review | 60% | 1.18 | 0% | 50% |
| Proportion of reviews with a photo | 20% | | | |
| Average number of photos per review (for reviews with at least one photo) | 2.46 | 2.18 | 1 | 30 |
| Proportion of 5-star reviews | 48% | | | |
| Proportion of reviews by elite users | 11% | | | |
| Average number of reviews written by the reviewer | 128.7 | 313.7 | 1 | 15,516 |

Table 2. Tripadvisor Data Set: Descriptive Statistics

| $N$ = 245,569 reviews in TripAdvisor | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|
| Average number of words per review | 76.0 | 74.7 | 1 | 2,489 |
| Share of visual words per review | 59% | 1.33 | 0% | 50% |
| Proportion of reviews with a photo | 23% | | | |
| Average number of photos per review (for reviews with at least one photo) | 2.85 | 1.17 | 1 | 4 |
| Proportion of 5-star reviews | 48% | | | |
| Average number of reviews written by the reviewer | 157.9 | 315.2 | 1 | 11,436 |
| Proportion of reviews written on a mobile device | 33% | | | |
| Proportion of reviews written by international visitors[a] | 13% | | | |

[a] Excluded from the analyses as the reviews are written in a language other than English.

visual language even when reviews did not contain exact terms in the original visual dictionary.

Formally, we calculated the degree of visual language in each review by using the Word2Vec embeddings model (Mikolov et al. 2013). We first tokenized and vectorized each review $i$, representing it as a 300-dimensional vector $w_i$. We then obtained word vectors $w_k$ for all LIWC visual words and averaged them to obtain a single reference vector $v$ representing the visual construct. Finally, for each review $i$, we computed the cosine distance between its vector $w_i$ and the visual reference vector $v$. Cosine similarity ranges from $-1$ to $+1$, such that higher scores indicate greater semantic alignment with the visual construct and thus, reflecting greater visual language use.

### Results

**Effect of Photos on Visual Language (Word-Frequency Method).** We tested whether the presence of a photo in a review is associated with an increase in visual language in the review text, indicating redundancy. For each data set, we examined the relationship between photo presence (1 = *photo present*, 0 = *photo absent*) and the degree of visual language in the text, using both LIWC and the word embedding approaches.

To account for unobserved heterogeneity in visual appeal across restaurants, we estimated a model with restaurant-level fixed effects. We clustered standard errors at the restaurant level to address potential within-business correlation in the residuals. In a separate model, we included review date fixed effects to account for temporal trends in photo usage, as camera-enabled phones became more common over time. Both models controlled for review length and star rating, and

used OLS estimation using the *lfe* package from R (Gaure 2013).

Using fixed-effects models at the restaurant level, we found evidence of redundancy in both data sets: the presence of a photo predicted greater use of visual language in review text; Yelp: $b = 0.09$, $t(6,690,649) = 63.72$, $p < .001$; TripAdvisor: $b = 0.08$, $t(193,531) = 7.29$, $p < .001$ (tables 3 and 4). The effect held including time fixed effects or excluding all fixed effects in both data sets.

For robustness, we also examined whether the number of photos predicted the use of visual language. Results showed that photo count also predicted greater use of visual language in review text; Yelp $b = 0.02$, $t(6,690,649) = 51.45$, $p < .001$; TripAdvisor $b = 0.02$, $t(193,531) = 5.84$, $p < .001$ (tables 5 and 6). The effect was robust including time fixed effects or excluding both (restaurant and time) fixed effects.

**Effect of Reviewer Status on Visual Language (Word-Frequency Method).** We also examined whether being an elite reviewer predicted visual language use. Elite status explicitly conferred on Yelp was associated with greater visual language use; $b = 0.04$, $t(6,690,649) = 63.72$, $p < .001$. In the TripAdvisor data set that does not confer status, we used the total number of reviews written by the reviewer (i.e., past review count) as a proxy for elite status. We observed a similar effect; $b = 0.006$, $t(193,531) = 2.46$, $p = .014$. As a robustness check, we reestimated the Yelp model including past review count as an alternative measure to elite status: results remained consistent (see app. B (available online)).

We next examined whether elite reviewers use photos and visual language differently from nonelite reviewers. Although elites used more visual language overall, when a

Table 3. Photo Presence Increases Visual Language in Yelp Reviews Using Word-Frequency Method

| | Dependent variable: LIWC visual words | | | | | |
|---|---|---|---|---|---|---|
| | Main | | | Interact | | |
| | Restaurant FE (1) | Time FE (2) | No FE (3) | Restaurant FE (4) | Time FE (5) | No FE (6) |
| Photo Present | .09*** (.001) | .11*** (.002) | .10*** (.002) | .10*** (.002) | .11*** (.002) | .11*** (.002) |
| User is Elite | .004** (.002) | .01*** (.002) | .01*** (.002) | .02*** (.002) | .03*** (.002) | .03*** (.002) |
| Log Review length | .09*** (.002) | .09*** (.002) | .10*** (.002) | .09*** (.002) | .09*** (.002) | .10*** (.002) |
| Star Rating | −.01*** (.001) | −.01*** (.001) | −.01*** (.001) | −.01*** (.001) | −.01*** (.001) | −.01*** (.001) |
| Photo present X user is elite | | | | −.04*** (.003) | −.05*** (.003) | −.05*** (.003) |
| Restaurant FE | Yes | No | No | Yes | No | No |
| Review date FE | No | Yes | No | No | Yes | No |
| Clustering: restaurant | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | .05 | .01 | .01 | .05 | .01 | .01 |

Note.— There were 6,713,325 observations.
** $p < .01$.
*** $p < .001$.

review included a photo (vs. no photo), the increase in their use of visual language was smaller than that of (nonelite) reviewers. This pattern was reflected in a negative and significant interaction between photo presence and elite status in Yelp: $b = -0.012, t(6,690,649) = 20.45, p < .001$. In the TripAdvisor data set, a negative relationship was also observed between photo presence and past review count; however, it failed to reach significance: $b = -0.01, t(193,531) = -1.34, p = .18$ (see tables 3 and 4).

**Word-Embedding Method.** Mirroring the results from the word-frequency method, the word embedding analyses also revealed a significant positive association between photo presence and visual language in the review text; Yelp: $b = 0.002, t(6,690,609) = 54.91, p < .001$ and TripAdvisor: $b = 0.002, t(193,513) = 5.42, p < .001$ (tables 7 and 8). Results were similar using photo counts as the independent variable (see apps. C and D). Further mirroring findings from the word-frequency method, the word embedding method revealed that being an elite user increased visual language use (Yelp: $b = 0.002, t(6,690,609) = 37.81, p < .001$, TripAdvisor: $b = 0.001, t(193,513) = 14.42, p < .001$). As before, the increase in elites' use of visual language

was smaller than that of nonelite reviewers; Yelp: $b = -0.003, t(6,690,609) = -41.77, p < .001$ and TripAdvisor: $b = -0.0003, t(193,513) = -1.54, p = .12$.

*Discussion*
Study 1 demonstrates, across two real-world data sets, that the presence of a photo in a review increases reviewers' visual language use in the accompanying text, supporting hypothesis 1. Although the photo already conveys visual elements of one's experience, reviewers still reference visuals in words, suggesting that they communicate redundantly across modalities.

Furthermore, elite users, who likely hold chronic impression goals, use more visual language, in line with our prediction (hypothesis 3). Moreover, the negative interaction between photo presence and elite status suggests that the presence of a photo more strongly impacts nonelite reviewers, prompting them to use more visual language. Elites, who use visual language more frequently overall, are less affected by situational factors, such as visual cues.

However, our computational approaches have limitations: the word-frequency method relies on a limited set of visual words, and word embeddings cannot capture multiword expressions like "check out" or "look at" as unified

Table 4. Photo Presence Increases Visual Language in Tripadvisor Reviews Using Word-Frequency Method

| | Dependent variable: LIWC visual words | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | Main | | | Interact | | |
| | Restaurant FE (1) | Time FE (2) | No FE (3) | Restaurant FE (4) | Time FE (5) | No FE (6) |
| Photo present | .08*** (.01) | .12*** (.01) | .12*** (.01) | .13*** (.04) | .20*** (.04) | .21*** (.04) |
| Log user Reviews | .01* (.002) | .002 (.003) | .002 (.003) | .01*** (.002) | .01 (.003) | .01* (.003) |
| Log Review length | .08*** (.01) | .09*** (.01) | .09*** (.01) | .08*** (.01) | .09*** (.01) | .09*** (.01) |
| Star Rating | .01** (.004) | .004 (.004) | .003 (.004) | .01** (.004) | .004 (.004) | .002 (.004) |
| Device is Mobile | −.02* (.01) | −.02 (.01) | −.02 (.01) | −.02* (.01) | −.02 (.01) | −.02 (.01) |
| Photo present X log user reviews | | | | −.01 (.01) | −.02* (.01) | −.02* (.01) |
| Restaurant FE | Yes | No | No | Yes | No | No |
| Review date FE | No | Yes | No | No | Yes | No |
| Clustering: Restaurant | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | .11 | .04 | .004 | .11 | .04 | .005 |

NOTE.— There were 200,527 observations.

\* $p < .05$.

\*\* $p < .01$.

\*\*\* $p < .001$.

semantic units. To validate our computational measures and ensure that what we identify as visual language corresponds to how humans perceive reviews, we next test whether human judges corroborate these computationally derived these linguistic patterns.

## STUDY 2: VALIDATING VISUAL LANGUAGE USE IN YELP REVIEWS BY HUMAN JUDGES

Across two different approaches, study 2 uses human judges to validate and extend study 1's computational findings. Study 2A tests whether the linguistic signal identified computationally aligns with readers' perceptions of the extent to which each review focuses on visual aspects. Beyond validation, this study develops a set of measures of visual focus applicable to smaller data sets where dictionary and embedding approaches are less reliable (Newman et al. 2003). Using this measure, visual focus can be assessed based solely on the review text, allowing us to compare photo and no-photo reviews and test hypothesis 1.

However, higher visual focus does not automatically imply greater redundancy. A review could describe visual aspects in

detail while the photo depicts entirely different visual elements (e.g., describing the ambiance while the photo shows a dessert). To assess the effect of being an elite on redundant communication directly (hypothesis 3), study 2B focuses only on reviews with photos and measures redundancy between photo and text (hypothesis 3).

### Study 2A: Assessment of Visual Focus by Human Judges

**Participants and Design.** Because it was not feasible to rate all study 1 reviews, we randomly selected two samples of 440 reviews each for human validation. Each sample was balanced: half of the reviews included photos and half did not, half were created by elite users and half were not. All reviews were positive (i.e., 4- and 5-star), with an average length of 89.5 words (SD = 20.34).

We recruited 200 US-based participants (46.5% female, $M_{age}$ = 42.3, SD = 13.2) to evaluate sample 1 and 201 US-based participants (50.2% female, $M_{age}$ = 43.7, SD = 12.4) to evaluate sample 2 via Cloud Connect. Each participant rated six randomly selected reviews yielding 9,624 total

Table 5. Photo Count Increases Visual Language in Yelp Reviews Using Word-Frequency Method

| | Dependent variable: LIWC visual words | | | | | |
|---|---|---|---|---|---|---|
| | Main | | | Interact | | |
| | Restaurant FE (1) | Time FE (2) | No FE (3) | Restaurant FE (4) | Time FE (5) | No FE (6) |
| Photo count | .02*** | .02*** | .02*** | .02*** | .03*** | .03*** |
| | (.0004) | (.001) | (.001) | (.0005) | (.001) | (.001) |
| User is elite | .01*** | .02*** | .01*** | .02*** | .03*** | .03*** |
| | (.002) | (.002) | (.002) | (.002) | (.002) | (.002) |
| Log review | .09*** | .09*** | .10*** | .09*** | .09*** | .10*** |
| Length | (.002) | (.002) | (.002) | (.002) | (.002) | (.002) |
| Star rating | −.01*** | −.01*** | −.01*** | −.01*** | −.01*** | −.01*** |
| | (.001) | (.001) | (.001) | (.001) | (.001) | (.001) |
| Photo count X user is elite | | | | −.01*** (.001) | −.02*** (.001) | −.02*** (.001) |
| Restaurant FE | Yes | No | No | Yes | No | No |
| Review date FE | No | Yes | No | No | Yes | No |
| Clustering: restaurant | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | .05 | .01 | .01 | .05 | .01 | .01 |

NOTE.— There were 6,713,325 observations.
*** $p < .001$.

evaluations. Due to random assignment of reviews to respondents, some reviews were not rated ($n = 55$), resulting in a total of 825 unique reviews evaluated across both samples.

**Procedure.** To ensure participants' judgments were based solely on text and to enable comparison between reviews, we removed all photos that accompanied each review. Participants rated each review text on four items (1 = *not at all*, 9 = *a great deal*) assessing the degree each review conveyed visual aspects of the experience (i.e., visual focus): "*How much does the review describe the appearance of the food, restaurant, or outdoors?*"; "*How much does the review mention what the writer saw at the restaurant?*"; "*How much does the review describe how things looked at the restaurant?*"; "*How much visual detail does the review include?*" The four items showed high internal consistency ($\alpha = .95$) and were averaged into a composite score with higher values indicating greater visual focus. In addition, we calculated two separate visual language scores for each review using (1) LIWC (word-frequency method) and (2) the word-embedding method as described in study 1.

**Results.** The LIWC-based visual language score was modestly but significantly associated with human perceptions of visual focus, $r(2406) = 0.08$, $p < .001$. The embedding-based visual language score showed a stronger association, $r(2406) = 0.26$, $p < .001$, providing convergent evidence that both computational measures aligned with human judgements.

Further, we tested the effects of photo presence and elite status on perceptions of visual focus using a linear mixed-effects model with random intercepts for participants. Consistent with study 1, reviews originally accompanied by a photo ($M = 4.05$, SD = 2.26) were rated as more visually focused than those without a photo ($M = 3.92$, SD = 2.30); $b = 0.25$, SE = 0.05, $t(9348) = 4.72$, $p < .001$. Also, reviews written by elite users ($M = 4.13$, SD = 2.46) were rated as more visually focused than those written by nonelite reviewers ($M = 3.83$, SD = 2.38), $b = 0.33$, SE = 0.05, $t(9337) = 6.21$, $p < .001$.

Finally, we found a significant and negative interaction between photo presence and elite status, $b = −0.16$, SE = 0.07, $t(9332) = −2.22$, $p = .026$. Whereas reviews written by elite reviewers were more visually focused overall, for these reviews, photo presence did not significantly increase the visual focus ($b = 0.08$, SE = 0.05, $t(4719) = 1.44$, $p = .15$). In contrast, for reviews written by nonelite reviewers, photo presence significantly increased visual focus ($b = 0.24$, SE = 0.06, $t(4541) = 4.34$, $p < .001$).

Table 6. Photo Count Increases Visual Language in Tripadvisor Reviews Using Word-Frequency Method

| | Dependent variable: LIWC visual words | | | | | |
|---|---|---|---|---|---|---|
| | Main | | | Interact | | |
| | Restaurant FE (1) | Time FE (2) | No FE (3) | Restaurant FE (4) | Time FE (5) | No FE (6) |
| Photo Count | .02*** (.004) | .04*** (.01) | .04*** (.01) | .04** (.01) | .07*** (.02) | .08*** (.02) |
| Log user Reviews | .01* (.002) | .001 (.003) | .002 (.003) | .01*** (.002) | .01 (.003) | .01* (.003) |
| Log review Length | .08*** (.01) | .09*** (.01) | .09*** (.01) | .08*** (.01) | .09*** (.01) | .09*** (.01) |
| Star rating | .01** (.004) | .005 (.004) | .003 (.004) | .01** (.004) | .004 (.004) | .002 (.004) |
| Device is mobile | −.02* (.01) | −.02 (.01) | −.02 (.01) | −.02* (.01) | −.02 (.01) | −.02 (.01) |
| Photo | | | | −.004 | −.01** | −.01** |
| Count X log user reviews | | | | (.003) | (.003) | (.003) |
| Restaurant FE | Yes | No | No | Yes | No | No |
| Review date FE | No | Yes | No | No | Yes | No |
| Clustering: restaurant | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | .11 | .04 | .004 | .11 | .04 | .005 |

NOTE.— There were 200,527 observations.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

**Discussion.** The positive correlations between human judgments and both LIWC- and embedding-based measures provide convergent support for the validity of these linguistic indicators. Further, consistent with hypothesis 1 and study 1 results using computational methods, human judges rated reviews that originally had been accompanied by a photo as more focused on visual aspects than those without (hypothesis 1), even though no photos were shown. Similarly, reviews written by elite reviewers were perceived as more visually focused than those written by nonelite reviewers (hypothesis 3). Also, photo presence increased visual focus less for elite reviews than nonelite reviews, consistent with study 1 results.

### Study 2B: Assessment of Redundancy by Human Judges

Because our theory and hypotheses focus on redundancy between review text and review photo, study 2B examines only reviews with photos and asks human judges to assess redundancy to directly test hypothesis 3.

**Participants and Design.** We used the 220 reviews with a photo from sample 2 in study 2A. Half of these reviews were written by elite users and half by nonelite users, with an average length of 89.5 words (SD = 20.34). All reviews were positive (4- and 5-star ratings).

We recruited 150 US-based participants (50.0% female, $M_{age}$ = 42.36, SD = 11.45) via Cloud Connect. Each participant rated 14 or 15 randomly selected reviews on two redundancy items, yielding approximately 4302 total evaluations of photo-text redundancy.

**Procedure.** Study 2B randomly presented participants with both the review text and photo. Participants rated each review on two items (1 = *not at all*, 9 = *a great deal*): "*How similar is the information in the review text to the information conveyed in the photo?*" and "*To what extent did the review text repeat the information conveyed in the photo?*" The two items showed high internal consistency ($r$ = 0.79) and were averaged into a composite redundancy score, with higher values indicating greater redundancy.

Table 7. Photo Presence Increased Visual Language in Yelp Reviews Using Word-Embedding Method

| | Dependent variable: Visual words using embeddings | | | | | |
|---|---|---|---|---|---|---|
| | Main | | | Interact | | |
| | Restaurant FE (1) | Time FE (2) | No FE (3) | Restaurant FE (4) | Time FE (5) | No FE (6) |
| Photo Present | .002*** (.0000) | .004*** (.0001) | .003*** (.0001) | .003*** (.0000) | .004*** (.0001) | .004*** (.0001) |
| User is Elite | .002*** (.0000) | .003*** (.0001) | .003*** (.0001) | .003*** (.0001) | .005*** (.0001) | .004*** (.0001) |
| Log Review length | .03*** (.0000) | .03*** (.0001) | .03*** (.0001) | .03*** (.0000) | .03*** (.0001) | .03*** (.0001) |
| Star Rating | .002*** (.0000) | .002*** (.0000) | .002*** (.0000) | .002*** (.0000) | .002*** (.0000) | .002*** (.0000) |
| Photo present X user is elite | | | | −.003*** (.0001) | −.003*** (.0001) | −.004*** (.0001) |
| Restaurant FE | Yes | No | No | Yes | No | No |
| Review date FE | No | Yes | No | No | Yes | No |
| Clustering: restaurant | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | .49 | .43 | .43 | .49 | .43 | .43 |

NOTE.— There were 6,713,285 observations.

*** $p < .001$.

**Results.** We tested the effect of elite status on photo-text redundancy using a linear mixed-effects model with random intercepts for participants. Consistent with hypothesis 3, reviews written by elite users ($M = 5.77$, SD = 2.23) were rated as significantly more redundant than those written by nonelite reviewers ($M = 5.30$, SD = 2.19), $b = 0.44$, SE = 0.09, $t(2054) = 5.01$, $p < .001$.

**Discussion.** Study 2B provides direct evidence that elite reviewers communicate more redundantly when using both photos and text. This pattern is consistent with the finding in study 2A that elite reviewers describe visual aspects more, providing convergent evidence across different measures. Together, this pattern is consistent with hypothesis 3 and supports our theoretical framework that elite reviewers with more chronic impression goals communicate more redundantly, likely reinforcing key information across multiple modalities.

## STUDY 3: TESTING THE EFFECT OF PLATFORM DESIGN AND GOALS ON VISUAL FOCUS

Study 3 examined whether platform design (i.e., photo selection order) and communicators' goals (i.e., impression motive) causally influence visual language use in consumer reviews. The study used a two-stage design: in stage 1, par-

ticipants reviewed a recent dining experience; in stage 2, independent raters evaluated the visual focus in those reviews.

### Participants and Design

In stage 1, we recruited 710 US-based participants (51.1% female, $M_{age} = 39.4$, SD = 12.1) via Cloud Research Connect. While hypotheses 2 and 3 only make main effect predictions, for efficiency we manipulated both factors (order and impression goal, respectively) in the same study. Participants were randomly assigned to create reviews in one of four experimental conditions in a 2 (order: photo-first vs. text-first) × 2 (goal: impression vs. control) between-subjects design.

In stage 2, we recruited a separate sample of 401 US-based participants (50.4% female, $M_{age} = 39.4$, SD = 12.6), also via Cloud Research Connect.

### Procedure

**Stage 1- Review Writing.** Participants were instructed to recall a recent (in the past three months), highly positive (5-star) dining experience at a local, nonchain restaurant they had visited in their town. We focused on positive experiences because they represent the majority of reviews on Yelp and other platforms (East, Hammond, and Wright

Table 8. Analyzing Visual Language in Tripadvisor Reviews Using Word-Embedding Method

| | Dependent variable: Visual words using embeddings | | | | | |
|---|---|---|---|---|---|---|
| | Main | | | Interact | | |
| | Restaurant FE (1) | Time FE (2) | No FE (3) | Restaurant FE (4) | Time FE (5) | No FE (6) |
| Photo | .002*** | .003*** | .003*** | .003** | .004*** | .005*** |
| Present | (.0003) | (.0004) | (.0004) | (.001) | (.001) | (.001) |
| Log | .001*** | .001*** | .001*** | .001*** | .001*** | .001*** |
| User reviews | (.0001) | (.0001) | (.0001) | (.0001) | (.0001) | (.0001) |
| Log | .04*** | .04*** | .04*** | .04*** | .04*** | .04*** |
| Review length | (.0003) | (.0002) | (.0003) | (.0003) | (.0002) | (.0003) |
| Star | .002*** | .002*** | .002*** | .002*** | .002*** | .002*** |
| Rating | (.0001) | (.0002) | (.0002) | (.0001) | (.0002) | (.0002) |
| Device | .001*** | .001*** | .001*** | .001*** | .001*** | .001*** |
| Is mobile | (.0002) | (.0003) | (.0003) | (.0002) | (.0003) | (.0003) |
| Photo | | | | −.0003 | −.0003 | −.0004 |
| Present X log user reviews | | | | (.0002) | (.0002) | (.0002) |
| Restaurant FE | Yes | No | No | Yes | No | No |
| Review date FE | No | Yes | No | No | Yes | No |
| Clustering: restaurant | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | .50 | .43 | .41 | .50 | .43 | .41 |

NOTE.— There were 200,509 observations.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

2007). After writing down the name of the restaurant, participants imagined writing a Yelp review about this experience. They were shown a mock-up of a Yelp review interface, including a star rating scale, a text box, and a photo upload button (see app. E). This visual interface ensured that all participants knew they would write a review text and also include a photo regardless of the order condition they were in.

All participants first rated the restaurant using the 5-star scale. Those in the impression goal condition then read: "*Your review will be read by many other Yelp users who are deciding whether to visit this restaurant. Write a review that will impress readers, meaning that they will think highly of you for your restaurant choice.*" Participants in the control condition just read "*Your review will be read by many other Yelp users who are deciding whether to visit this restaurant.*" Then, in each condition, half of the participants began by writing their review text, while the other half began by selecting a photo for their review. To select the photo, participants searched for their chosen restaurant on Google, selected

a photo they wanted to include with their review, and uploaded it.

After completing the review task, participants responded to a two-item goal manipulation check: "How important was it to make others think highly of you?" and "How important was it to impress others with your restaurant choice?" on a 7-point scale (1 = *Not at all important*, 7 = *Extremely important*). The two items showed high internal consistency ($\alpha$ = .90) and were averaged into a composite score of impression goal rating. Participants also reported several covariates, including dining format, spending at the restaurant, frequency of online reviewing, total number of reviews written, gender, and age.

**Stage 2: Review Evaluation.** Independent human judges assessed the reviews created in Stage 1 and rated the degree of visual details they conveyed. Similar to study 2A, we showed participants only the review text (word count =84.03, SD = 15.81). Each participant rated six randomly selected reviews,

yielding 9,620 total evaluations.[1] Due to random assignment of reviews to respondents, some reviews were not rated ($n = 25$), resulting in a total of 685 unique reviews evaluated in the study. Participants rated the reviews using the same four items from study 2A ($\alpha = .96$). We averaged these items into a composite score with higher values indicating greater visual focus.

### Results

**Manipulation Check (Stage 1).** To assess whether the goal manipulation was successful, we regressed participants' impression goal ratings on goal condition (control = −1 vs. impression = 1), order (text-first = −1 vs. photo-first = 1), and their interaction. The analysis revealed a significant main effect of goal condition: participants in the impression condition reported greater impression motivation ($M = 3.51$, SD = 1.91) than those in the control condition ($M = 2.72$, SD = 1.65), $b = 0.73$, SE = 0.20, $t(706) = 3.69$, $p < .001$. Neither the main effect of order ($p = .78$) nor the interaction between order and goal conditions ($p = .64$) was significant.

**Perceived Visual Focus (Stage 2).** We regressed visual focus ratings from stage 2 on goal condition (control = −1 vs. impression = 1), order (text-first = −1 vs. photo-first = 1), and their interaction (see fig. 1). This analysis revealed a significant main effect of photo selection order, $b = 0.07$, SE = 0.02, $t(9376) = 3.43$, $p < .001$ and an interaction between order and goal, $b = -0.09$, SE = 0.02, $t(9374) = 4.51$, $p < .001$.

To unpack the interaction, we first examined the effect of order within each goal condition. Consistent with hypothesis 2, in the control condition, reviews written after selecting a photo ($M = 4.28$, SD = 2.42) were rated as more visually focused than those written before ($M = 3.87$, SD = 2.37), $b = 0.31$, SE = 0.06, $z = 5.57$, $p < .001$. This effect was attenuated in the impression goal condition,

_____

1. Due to attrition, the 710 participants creating reviews were unequally distributed among the four conditions ($X^2 = 5.96$, $p = .015$, $n$ (Impress photo first) = 171, $n$ (Impress text first) = 199, $n$ (Control photo first) = 159, $n$ (Control text first) = 171). Tests assessing whether condition affected any of the control variables collected, only found a small main effect on gender, such that the impression condition included fewer females. Notably, conditions did not differ in terms of indicators of effort (i.e., word count of review, time spent on study), as such we believe attrition is unlikely to have meaningful consequences for the stimuli created in stage 1 and subsequently evaluated for stage 2.
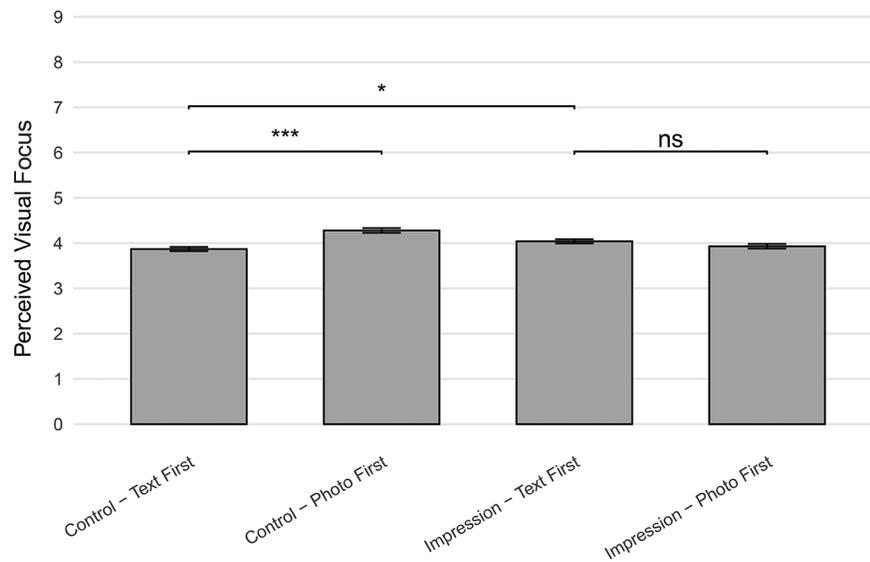
where selection order did not significantly impact perceived visual focus ($p = .45$).

Second, to isolate the effect of goals without the influence of photo selection order, we compared the effect of goals within the text-first condition. Supporting hypothesis 3, reviews written under impression goal ($M = 4.04$, SD = 2.37) were rated as more visually focused than those written under control goal ($M = 3.87$, SD = 2.37), $b = 0.12$, SE = 0.05, $z = 2.21$, $p = .027$.

### Discussion

Study 3 demonstrates that impression goals causally increase visual focus in reviews. Supporting hypothesis 3, reviews written under situational impression goals were perceived as more visually focused than control reviews when the text was written first.

Importantly, impression goals also moderated the effect of platform design. Consistent with hypothesis 2, in the control goal condition, reviews were perceived as more visually focused when the photo was selected before writing text (vs. after). However, when impression goals were active, this selection-order effect disappeared, suggesting that communicators with impression goals maintain high visual focus regardless of platform prompts.

These findings align with studies 1 and 2A, in which elite reviewers, who likely hold chronic impression goals, extensively described visual aspects whether or not they included a photo. Together, these findings suggest that impression goals, whether chronic or situational, lead to consistent visual language use that is less dependent on platform design features.

## GENERAL DISCUSSION

Researchers have long explored the psychological drivers of social transmission. With the widespread adoption of camera-enabled phones, photos have become increasingly central to how consumers share their experiences. This article examines how people create WOM when both visual and verbal modalities are at their disposal. Across two large real-world data sets, and controlled laboratory experiments, we consistently find evidence that people communicate redundantly, describing visual aspects in their text even when photos are present. Our findings suggest that communicators align more with Grice's maxim of relation (relevance and emphasis on important details) than with Grice's maxim of quantity (brevity and avoiding redundancy). This tendency is heightened when review platforms prompt users to begin their communication by selecting a photo and when communicators focus on how their review is received by others.

Figure 1. Perceived visual focus use as a function of selection order and goals.

## Theoretical Contributions

This work makes several theoretical contributions. Most importantly, our findings extend research on how consumers generate WOM. While prior work has focused primarily on verbal WOM (Berger and Milkman 2012; Packard and Berger 2017), we examine how people create visual-verbal WOM. Our findings demonstrate that people communicate redundantly, using both modalities to convey similar information rather than substituting one for the other. We also identify platform-induced and communicator-driven factors that shape redundant communication.

Further, we contribute to the literature on the role of platform design in shaping WOM creation (e.g., Aerts et al. 2017; Liu et al. 2025). We build on this literature by showing that the order in which communicators produce each modality matters. Beginning by selecting a photo before writing text heightens redundancy.

Our work also contributes to research on how communicative goals shape online expression. Prior work shows that communicators' active goals, such as impression goals, systematically influence what and how people share (Barasch and Berger 2014; Schlosser 2020). Adding to this literature, we show that visual-verbal communication becomes more redundant when impression goals are active: reviewers are more likely to reinforce key visual aspects in their verbal descriptions when they focus on how their review will be received.

Our findings further speak to a broader question in linguistics about how communicators balance communicative efficiency with impact on the receiver (Zipf 1949). Whereas some work suggests that languages evolve for efficiency (Hahn, Jurafsky, and Futrell 2020), other research finds that communicators sometimes prioritize clarity and impact on the receiver (Degen et al. 2020). Our findings align with the latter research stream, revealing that communicators choose redundancy to enhance audience impact in visual-verbal communication. Unlike emojis, which often replace certain words (e.g., a happy face can replace the word "happy"; Luangrath, Peck, and Barger 2017), photos are not treated as substitutes for verbal content. Instead, communicators accompany photos with language that highlights visual features.

We also contribute to the advertising and journalism literature that examines the interplay between photos and words. Prior work recommends professionals should minimize redundancy by ensuring that text and photos provide distinct rather than overlapping information (Houston, Childers, and Heckler 1987; McIntyre, Lough, and Manzanares 2018). In contrast, we find that in consumer-generated content, communicators often do the opposite: they create redundancy across modalities, using text to reinforce visual information conveyed by photos.

## Substantive Contributions

Substantively, understanding what drives redundancy has important implications. Recent research finds that reviews with greater photo-text similarity, i.e., greater redundancy, are perceived as more helpful (Ceylan et al. 2024). Redundancy may enhance helpfulness by reinforcing key information across modalities, reducing cognitive effort and

increasing message clarity and persuasiveness (Mayer 2019; Paivio 1991).

Our findings offer platforms actionable strategies to facilitate creation of helpful reviews. Platforms can design interfaces that encourage redundancy. For instance, platforms can provide real-time feedback during review creation by keeping the uploaded photo visible while reviewers write their text and offering gentle prompts like "Consider describing what's shown in your photo" if visual focus is low. Alternatively, they can activate impression goals (e.g., telling reviewers "Thousands of travelers will see your review") that would increase redundancy and ultimately helpfulness.

Platforms can also use redundancy metrics to inform which reviews are displayed most prominently. Since visual focus in text is computationally detectable, platforms can weight redundancy in their "most helpful" algorithms or default sort orders, alongside existing signals like review length and upvotes. This would enable platforms to quickly determine most helpful reviews during submission, even before helpfulness votes accumulate and help consumers to quickly identify helpful reviews without having to scroll through numerous posts.

Beyond platforms, redundancy can help companies interpret consumer feedback more accurately. When consumers show and describe the same feature, such as a meal's presentation, the verbal description clarifies what the reviewer focuses on in the photo and grounds the visual information. Reviews with higher redundancy thus offer firms more diagnostic information about what specifically drives satisfaction or dissatisfaction. Just as Instagram's visual culture has already prompted restaurants to design more photogenic dishes and interiors (Lee 2017), visual-verbal redundancy in reviews can signal which aspects of the consumer experience are most noticed, valued or criticized. By analyzing redundant reviews, companies can better identify and enhance the features that matter most to consumers and improve those that consistently result in disappointment.

### *Limitations and Directions for Future Research*
This article offers initial understanding how people create visual-verbal WOM. However, many interesting questions remain. We focused on restaurant reviews because consumers frequently review their dining experiences using photos and food consumption is inherently multisensory, with visual elements playing a central role (Krishna 2012; Spence et al. 2010; Zellner et al. 2010). This highly visual product category may amplify tendencies toward visual-verbal redundancy. Future research should examine whether our

findings generalize to categories in which visual elements play a more peripheral role (e.g., software products). Additionally, while we focused on photo-text redundancy, other forms of multimodal communication (e.g., videos) may involve not just visual but also auditory redundancy, offering another promising avenue for research.

Additionally, our investigation focused on online WOM in broadcasting contexts (one-to- many communication). Online WOM can differ from offline WOM (Eisingerich et al. 2015). Further, people communicate differently when addressing a broad audience versus one person or a few selected individuals (Barasch and Berger 2014). Future research should examine whether and how audience size and context shape multimodal communication patterns, such as photo-text redundancy in private messages versus public reviews.

Further, we examined communication with audiences who were not part of the original experience. However, communicators often share experiences with people who were present (e.g., a joint vacation). Prior research in object identification (Rubio-Fernández 2016) suggests that communicators often mention information redundantly, even when sharing the same context with receivers, such as when both looked at the same objects. Similarly, when conveying experiences to receivers who shared the original context, communicators may still engage in redundancy to reinforce important details, or they may reduce redundant communication assuming that the shared context makes it unnecessary. Future research may explore whether shared experiences amplify or reduce redundancy.

Future research could also explore how photo content influences redundancy in visual- verbal WOM. For instance, Barasch, Zauberman, and Diehl (2018) find that when people share photos with others who were not present during the experience, they tend to select prototypical images, such as Christmas trees, stockings, and gifts for a holiday celebration. This raises an important question: Does photo prototypicality reduce or enhance redundancy in accompanying text? If an image already conveys a familiar scene, communicators may feel less need to reinforce it with words. Alternatively, they may still communicate redundantly to highlight key aspects for the receiver. Exploring this dynamic would further clarify how communicators balance efficiency and informativeness in visual-verbal WOM.

We identify impression goals as one driver of redundant communication, but other motivations may also play a role. For example, those aiming to persuade may view redundancy as more convincing, whereas self-conscious individuals may avoid redundancy to prevent appearing boastful and braggy.

Future research may investigate how different communication goals shape visual-verbal WOM.

While our studies used reviews from established platforms where authenticity is generally assumed, an important question is whether redundancy patterns differ between genuine and fraudulent reviews. Future research could examine whether authentic reviewers who photograph their own experiences show different visual-verbal correspondence patterns compared to fraudulent reviewers using stock images or appropriated photos. Such work could inform computational approaches to review authenticity detection.

From writing to snapping, technology continuously offers new ways to communicate across multiple channels and modalities, reshaping social interactions. As communication evolves, understanding visual-verbal redundancy becomes increasingly important. Though many interesting questions remain, we hope our work provides a foundation for examining how people strategically combine visual and verbal elements in user-generated content.

## REFERENCES

Aerts, Goele, Tim Smits, and Peeter W. J. Verlegh (2017), "The Platform Shapes the Message: How Website Design Affects Abstraction and Valence of Online Consumer Reviews," *Decision Support Systems*, 104, 104–12.

Barasch, Alixandra, and Jonah Berger (2014), "Broadcasting and Narrowcasting: How Audience Size Affects What People Share," *Journal of Marketing Research*, 51 (3), 286–99.

Barasch, Alixandra, Gal Zauberman, and Kristin Diehl (2018), "How the Intention to Share Can Undermine Enjoyment: Photo-Taking Goals and Evaluation of Experiences," *Journal of Consumer Research*, 44 (6), 1220–37.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1–25.

Berger, Jonah, and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192–205.

Bhatia, Sudeep, Russell Richie, and Wanling Zou, (2019), "Distributed Semantic Representations for Modeling Human Judgment," *Current Opinion in Behavioral Sciences*, 29, 31–36.

Burtch, Gordon, Qinglai He, Yili Hong, and Dokyun Lee (2022), "How Do Peer Awards Motivate Creative Content? Experimental Evidence from Reddit," *Management Science*, 68 (5), 3488–506.

Ceylan, Gizem, Kristin Diehl, and Davide Proserpio (2024), "Words Meet Photos: When and Why Photos Increase Review Helpfulness," *Journal of Marketing Research*, 61 (1), 5–26.

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, Herbert (1992), *Arenas of Language Use*, University of Chicago Press.

Davenport, Thomas H., and John C. Beck (2001), "The Attention Economy," *Ubiquity*, May (6), 1, https://doi.org/10.1145/376625.376626.

Degen, Judith, Robert D. Hawkins, Caroline Graf, Elisa Kreiss, and Noah D. Goodman (2020), "When Redundancy Is Useful: A Bayesian Approach to 'Overinformative' Referring Expressions.," *Psychological Review*, 127 (4), 591–621.

Diehl, Kristin, Alixandra Barasch, Minjeong Ko, and Gal Zauberman (2025), "Captured Memories: The Impact of First-Person Versus Third-Person Viewpoint Photographs on Remembering Personal Experiences," *Journal of Applied Research in Memory and Cognition*, http://dx.doi.org/10.2139/ssrn.5236265.

East, Robert, Kathy Hammond, and Malcolm Wright (2007), "The Relative Incidence of Positive and Negative Word of Mouth: A Multi-Category Study," *International Journal of Research in Marketing*, 24 (2), 175–84.

Eisingerich, Andreas B., HaeEun Helen Chun, Yeyi Liu, He Michael Jia, and Simon J. Bell (2015), "Why Recommend a Brand Face-to-Face but Not on Facebook? How Word-of-Mouth on Online Social Sites Differs from Traditional Word-of-Mouth," *Journal of Consumer Psychology*, 25 (1), 120– 28.

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.

Garten, Justin, Joe Hoover, Kate M. Johnson, Reihane Boghrati, Carol Iskiwitch, and Morteza Dehghani (2018), "Dictionaries and Distributions: Combining Expert Knowledge and Large Scale Textual Data Content Analysis: Distributed Dictionary Representation," *Behavior Research Methods*, 50 (1), 344–61.

Gaure, Simen (2013), "OLS with Multiple High Dimensional Category Variables," *Computational Statistics and Data Analysis*, 66, 8–18.

Goodman, Noah D., and Michael C. Frank (2016), "Pragmatic Language Interpretation as Probabilistic Inference," *Trends in Cognitive Sciences*, 20 (11), 818–29.

Grice, H. (1975), "Logic and Conversation," in *Syntax and Semantics*, vol. 3, ed. K. Moulton, Brill, 41–58, https://brill.com/view/book/edcoll/9789004368811/BP000003.xml.

Hahn, Michael, Dan Jurafsky, and Richard Futrell (2020), "Universals of Word Order Reflect Optimization of Grammars for Efficient Communication," *PNAS*, 117 (5), 2347–53.

Houston, Michael J., Terry L. Childers, and Susan E. Heckler (1987), "Picture-Word Consistency and the Elaborative Processing of Advertisements," *Journal of Marketing Research*, 24 (4), 359–69.

Humphreys, Ashlee, and Rebecca Jen Hui Wang (2018), "Automated Text Analysis for Consumer Research," *Journal of Consumer Research*, 44 (6), 1274–306.

Krishna, Aradhna (2012), "An Integrative Review of Sensory Marketing: Engaging the Senses to Affect Perception, Judgment and Behavior," *Journal of Consumer Psychology*, 22 (3), 332–51.

Kronrod, Ann, Ivan Gordeliy, and Jeffrey K. Lee (2023), "Been There, Done That: How Episodic and Semantic Memory Affects the Language of Authentic and Fictitious Reviews," *Journal of Consumer Research*, 50 (2), 405–25.

Leary, Mark R., and Robin M. Kowalski (1990), "Impression Management: A Literature Review and Two-Component Model," *Psychological Bulletin*, 107 (1), 34–47.

Lee, Sarah (2017), "Picture Perfect? How Instagram Changed the Food We Eat," December 29, https://www.bbc.com/news/uk-england-london-42012732.

Liu, Miaomiao, Xiaohua Zeng, Cheng Zhang, and Yong Liu (2025), "What Happens When Platforms Disclose the Purchase History Associated with Product Reviews?" *Decision Support Systems*, 188 (January), 114367.

Luangrath, Andrea Webb, Joann Peck, and Victor A. Barger (2017), "Textual Paralanguage and Its Implications for Marketing Communications," *Journal of Consumer Psychology*, 27 (1), 98–107.

Mathwick, Charla, and Jill Mosteller (2017), "Online Reviewer Engagement: A Typology Based on Reviewer Motivations," *Journal of Service Research*, 20 (2), 204–18.

Mayer, Richard E. (2019), "Thirty Years of Research on Online Learning," *Applied Cognitive Psychology*, 33 (2), 152–59.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014), "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, 104 (8), 2421–55.

McIntyre, Karen, Kyser Lough, and Keyris Manzanares (2018), "Solutions in the Shadows: The Effects of Photo and Text Congruency in Solutions Journalism News Stories," *Journalism and Mass Communication Quarterly*, 95 (4), 971–89.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013), "Distributed Representations of Words and Phrases and Their Compositionality," *Advances in Neural Information Processing Systems*, 3111–19.

Moore, Sarah G. (2012), "Some Things Are Better Left Unsaid: How Word of Mouth Influences the Storyteller," *Journal of Consumer Research*, 38 (6), 1140–54.

Moore, Sarah G., and Katherine C. Lafreniere (2020), "How Online Word-of-Mouth Impacts Receivers," *Consumer Psychology Review*, 3 (1), 34–59.

Newman, Matthew L., James W. Pennebaker, Diane S. Berry, and Jane M. Richards (2003), "Lying Words: Predicting Deception from Linguistic Styles," *Personality and Social Psychology Bulletin*, 29 (5), 665–75.

Packard, Grant, and Jonah Berger (2017), "How Language Shapes Word of Mouth's Impact," *Journal of Marketing Research*, 54 (4), 572–88.

Paivio, Allan (1991), "Canadian Journal of Psychology Outstanding Contributions Series Dual Coding Theory: Retrospect and Current Status," *Canadian Journal of Psychology*, 45 (3), 255–87.

Partan, Sarah, and P. Marler (1999), "Communication Goes Multimodal," *Science*, 283 (5406), 1272–73.

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.

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

Rubio-Fernández, Paula (2016), "How Redundant Are Redundant Color Adjectives? An Efficiency-Based Analysis of Color Overspecification," *Frontiers in Psychology*, 7, 153.

Schlosser, Ann E. (2020), "Self-Disclosure Versus Self-Presentation on Social Media," *Current Opinion in Psychology*, 31, 1–6.

Schwarz, Norbert (1994), "Judgment in a Social Context: Biases, Shortcomings, and the Logic of Conversation," *Advances in Experimental Social Psychology*, 26 (C), 123–62.

Spence, Charles, Carmel A. Levitan, Maya U. Shankar, and Massimiliano Zampini (2010), "Does Food Color Influence Taste and Flavor Perception in Humans?" *Chemosensory Perception*, 3 (1), 68– 84.

Sperber, D., and D. Wilson (1986), *Relevance: Communication and Cognition*, 2nd ed., Harvard University Press.

Tausczik, Yla R., and James W. Pennebaker (2010), "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," *Journal of Language and Social Psychology*, 29 (1), 24– 54.

Valsesia, Francesca and Kristin Diehl (2022), "Let Me Show You What I Did Versus What I Have: Sharing Experiential Versus Material Purchases Alters Authenticity and Liking of Social Media Users," *Journal of Consumer Research*, 49 (3), 430–49.

Wilson, Deirdre (1993), "Relevance and Understanding," *Pragmalinguistica*, 1, 335–66.

Yeomans, Michael (2019), "Some Hedonic Consequences of Perspective-Taking in Recommending," *Journal of Consumer Psychology*, 29 (1), 22–38.

Zellner, Debra A., Mia Lankford, Lisa Ambrose, and Paul Locher (2010), "Art on the Plate: Effect of Balance and Color on Attractiveness of, Willingness to Try and Liking for Food," *Food Quality and Preference*, 21 (5), 575–78.

Zhang, Mingyue, Xuan Wei, and Daniel Dajun Zeng (2020), "A Matter of Reevaluation: Incentivizing Users to Contribute Reviews in Online Platforms," *Decision Support Systems*, 128, 113158.

Zipf, George Kingsley (1949), *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*, Addison-Wesley.