This research explores how expressed emotional arousal in a consumer review affects reader perceptions of its helpfulness. Drawing from research on written communication and lay theories of emotion, the authors propose a pattern of diminishing returns, in which the marginal effect of arousal on perceived helpfulness is positive at low levels of arousal but diminishes at higher levels. Results of a field study using Apple’s App Store, a follow-up survey, and two laboratory experiments provide consistent evidence for the predicted pattern. In addition, the results suggest that the nonlinear effect is explained in part by perceptions of reviewer effort and that the effect is stronger for products that are utilitarian in nature. By revealing a nuanced relationship between emotional expression and perceived helpfulness, these findings offer valuable implications for effective word-of-mouth communication.

Keywords: word of mouth, arousal, emotion, helpfulness, consumer reviews

Online Supplement: http://dx.doi.org/10.1509/jmr.13.0379
from exaggerating their feelings, classifying “rants and raves” as unhelpful; in contrast, Amazon (2016) encourages reviewers to share all their “energy and enthusiasm (both favorable and critical)” about a product. This disparity reflects broader disagreement about the role of emotions in judgment and decision making. Conventional wisdom often treats emotion as an impediment to rationality, and ample evidence has shown that intense emotional experiences can interfere with deliberation and self-control (Johnson and Tversky 1983; Loewenstein 1996; Shiv and Fedorikhin 1999; Tice, Bratslavsky, and Baumeister 2001). However, emotions also serve an important informational and motivational role, such that their impact on cognitive tasks is variable and sometimes beneficial (Damasio 2005; Isen 2001; Lerner and Keltner 2000; Naqvi, Shiv, and Bechara 2006).

Although there are myriad ways in which emotional arousal in a review may affect reader perceptions, we focus on a pathway of particular interest—inferring a tendency “fill in the blanks” by making inferences about reviewers (Naylor, Lamberton, and Norton 2011). An especially relevant inference concerns the amount of effort exerted by reviewers in deliberating on their experience and constructing a thoughtful review (Yin, Bond, and Zhang 2014). Building on contemporary models of emotional communication (Harris and Paradice 2007; Planalp 1998), we argue that readers make inferences of reviewer effort on the basis of emotional arousal, which represents a fundamental dimension of emotional expression. Specifically, we theorize a nonlinear, diminishing-returns relationship between expressed emotional arousal and perceived review helpfulness, such that at low levels of expressed arousal, readers associate additional arousal with greater effort and a more helpful review, but at higher levels, the effect of additional arousal diminishes and may even become negative. Importantly, our prediction deals only with perceptions of reviewer effort and helpfulness (regardless of the accuracy of those perceptions).

To examine our prediction, we presented a field study, a survey, and two laboratory experiments. Across our studies, expressed arousal is either measured or manipulated, in the form of verbal cues, nonverbal cues, or their combination. In addition, we explore perceived reviewer effort as a potential explanatory mechanism, and we examine the nature of the product (utilitarian or hedonic) as a theoretically relevant moderator (Moore 2015). By exploring the role of expressed arousal in consumer reviews, our research leads to a growing body of knowledge on emotions in word of mouth (WOM). Recent work has focused on the motivation of message recipients to share a message, based on the extent to which they find its content arousing (Berger 2011; Berger and Milkman 2012). We complement this work by focusing on inferences that recipients make about the sender and the message, based on the extent to which message content expresses arousal. More broadly, our exploration contributes to an important emerging literature on the social function of emotions (Van Kleef, De Dreu, and Manstead 2010).

THEORETICAL DEVELOPMENT AND HYPOTHESES

The Expression of Emotional Arousal

Psychologists have proposed a vast range of models to classify fundamental dimensions of emotional experience (Brosch, Pourtois, and Sander 2010; Mano 1991; Osgood 1966; Watson and Tellegen 1985). Although these models vary considerably, they consistently identify dimensions of valence and arousal to be particularly important (Niedenthal 2008; Russell 1980). Valence describes the extent to which a person perceives an experience as pleasant or unpleasant and is used to distinguish “positive” and “negative” affective states. Arousal (also “activation” or “intensity”) describes the extent to which a person is energized by an experience. Scales measuring arousal utilize endpoints such as “calming” versus “exciting,” “soothing” versus “agitating,” and so on (Bachorowski and Braaten 1994; Heilman 1997). Although valence and arousal sometimes covary, the two dimensions are independent and distinguishable, both phenomenologically and physiologically (Niedenthal 2008; Russell and Barrett 1999; Smith and Ellsworth 1985).

It is widely accepted that emotional response manifests through an “emotional reaction triad” of experience, physiology, and expression (Ekman 1993; Lang 1995; Levenson 1994; Scherer 2000). Applied here, experienced arousal captures the intensity of one’s subjective feelings, while physiological arousal captures bodily changes and physical responses (e.g., muscular tension, increased heart rate and blood pressure; Cannon 1927; Schachter and Singer 1962). In a consumer review setting, the experienced and physiological arousal of a review writer are not directly observable but must be inferred from the expression of arousal in the review content itself. Therefore, we focus on expressed arousal, and we use the terms “arousal” and “expressed arousal” interchangeably.

Emotional expression during communication occurs through verbal and nonverbal cues (Scherer 2000). Verbal cues represent the specific words embedded in a message; for example, communicators may reveal emotions with descriptive words (e.g., “love,” “jealousy”) and convey emotional arousal with linguistic markers (e.g., “a bit,” “very,” “really”) (Harris and Paradice 2007). In contrast, nonverbal cues represent auxiliary signals that depend on the medium. In spoken communication settings, nonverbal cues may include facial expressions, vocal expressions, and body language. In written communication, a wide variety of nonverbal cues are available (Carey 1980; Planalp 1998): for example, writers may use emoticons (e.g., “_:” and other symbolic text to imitate facial or vocal expressions, or they may convey arousal through the use of grammatical markers. Capitalized words and exclamation marks that connote high volume (“shouting”) are commonly used to indicate high-arousal emotions (Allen 1988; Schandorf 2013).

Arousal Cues and Reader Inferences

The online review setting represents an “actor–observer” context, in which reviewers provide product-relevant content, and readers later make sense of both the content and the reviewer. Given widespread evidence that observers attend to and elaborate on emotional content (Van Kleef 2010), it is logical to expect that review-embedded emotions will be important in the sense-making process. Emotion-related words are often processed automatically, and perceivers tend to utilize emotional cues even when other sources of information are sufficient (Gendron et al. 2012; Gernsbacher, Hallada, and Robertson 1998). In computer-mediated communication (Harris and Paradice 2007; Vandergriff 2013), readers
incorporate both verbal and nonverbal cues to assess the emotional state of senders, such that the presence of more arousal cues leads to assessments of greater sender arousal. Applied in our context, review readers who encounter more arousal cues are likely to infer that the reviewer experienced greater arousal during the process of constructing the review and, in turn, generate related inferences about the reviewer and review.

Of the various inferences that may be relevant to our setting, we focus on inferences regarding reviewer effort. We define perceived reviewer effort as the extent to which a reader believes that a reviewer exerted thoughtful deliberation on the content and construction of his or her review (note that this definition does not include effort expended in purchasing or using the product itself). Our focus on effort derives from both practical and theoretical considerations. Consumer processing of decision-relevant information typically involves an evaluation of the information source, and this is especially true in the case of product recommendations (Gershoff, Broniarczyk, and West 2001). Evidence from review contexts has suggested that consumers form spontaneous inferences about reviewer characteristics, which are typically unknown or ambiguous (Naylor, Lamberton, and Norton 2011). Moreover, consumers appear to recognize that the effort underlying a review is relevant to its usefulness, as objective indicators of reviewer effort predict helpfulness perceptions even after other factors are controlled for (Mudambi and Schuff 2010). Prior research has indicated that consumers infer reviewer effort from the discrete emotions in a review (Yin, Bond, and Zhang 2014).

Given that arousal is a fundamental component of emotional expression, it is reasonable to expect that perceived arousal may itself influence inferences regarding effort.

Research on the mental representation of emotion has identified commonly held beliefs regarding the characteristics, antecedents, and consequences of emotional experience (e.g., Shaver et al. 1987). Among these is the belief that emotional arousal serves an important motivational function. People exhibiting low levels of arousal are commonly described by others as “relaxed,” “calm,” or “bored,” indicating limited task-directed attention, information processing, or elaboration (O’Hanlon 1981). In contrast, people exhibiting high levels of arousal are described by terms such as “excited” or “enthusiastic,” indicating readiness to expend energy and take action (Lang 1995; Seo, Barrett, and Bartunek 2004). This connection is corroborated by research on lay theories of specific emotions. For example, Frijda, Kuipers, and Ter Schure (1989) show that targets experiencing low-arousal emotions such as sadness or boredom are expected to generate little motivated action (e.g., “apathy,” “rest”), whereas targets experiencing high-arousal emotions such as anxiety or anger are expected to engage in effortful exertion (e.g., “attending,” “reactant”). Applying this logic, we predict that, in general, readers will associate greater expressed arousal in a review with greater effort expended in constructing the review.

Importantly, however, we also predict that the positive association between expressed arousal and perceived reviewer effort will weaken as arousal increases. At higher levels of arousal, the level of implied effort will become more implausible, and readers will be increasingly likely to form alternative inferences. For example, readers who encounter high levels of arousal may perceive that a reviewer was utilizing a more subjective versus objective writing style, was simply “gushing” or “venting” about his or her experience (Jensen et al. 2013; Park, Lee, and Han 2007), or was engaging in strategic puffery—“trying too hard” to convince prospective consumers one way or the other (Fristad and Wright 1994; Xu and Wyer 2010). In addition, readers may interpret extreme levels of arousal as a signal of irrationality. Although emotion is often an important input to reasoning tasks (Damasio 2005; Pham 2007), highly elevated arousal is associated with numerous cognitive shortcomings (Fedorikhin and Patrick 2010; Gorn, Pham, and Sin 2001; Pham 2007; Sanbonmatsu and Kardes 1988), and research on lay theories of emotion has revealed a common belief that intense emotion is an impediment to reasoning (Oatley and Johnson-Laird 2014). At very high levels of expressed arousal, therefore, the relationship between arousal and perceived effort will weaken or even reverse.

Our final argument is that perceptions of reviewer effort are positively associated with perceptions of review helpfulness. This argument is both intuitive and consistent with broad interdisciplinary evidence that perceivers assume effort and performance to be correlated (see Skinner, Chapman, and Baltes 1988; Weiner and Kukla 1970). As a form of discursive writing (Vygotsky 1964), reviewing is deliberative in nature and involves various subtasks that each require cognitive resources: reconstructing one’s experience with a product, translating those thoughts into coherent form, incorporating assumptions about the audience, and so on. Readers will naturally assume that reviewers who have expended greater effort have performed these subtasks more thoroughly (e.g., by reconstructing a more complete account of the experience), thereby producing more helpful reviews. Taken together, our arguments lead to the following:

H1: The level of expressed arousal in a review is associated with perceived review helpfulness in a nonlinear (diminishing returns) manner, such that the effect of additional arousal is positive at low levels of arousal but becomes smaller as arousal increases.

H2: The nonlinear effect of expressed arousal on perceived review helpfulness is mediated by perceived effort, such that (a) expressed arousal is associated with perceived reviewer effort in a nonlinear (diminishing returns) manner, and (b) perceived reviewer effort is positively associated with perceived review helpfulness.

Arousal-Based Inferences Across Product Categories

As typically defined, utilitarian products are primarily instrumental and consumed for functional purposes, whereas hedonic products are primarily experiential and consumed for pleasure (Dhar and Wertenbroch 2000; Sloot, Verhoef, and Franses 2005). Hedonic products tend to be more “affectively rich” than utilitarian products and to evoke stronger emotional reactions (Babin, Darden, and Griffin 1994; Botti and McGill 2011). Given these intrinsic differences, it stands to reason that consumers will expect WOM for hedonic products to contain greater emotion than WOM for utilitarian products. Empirical evidence has confirmed that these expectations are often born out, as emotional language is more prevalent for online reviews in hedonic categories (Kronrod and Danziger 2013).

The tendency for reviews of hedonic products to be more emotional has important implications for reader inference. Research on affective decision making in intrapersonal
contexts has shown that emotions that “stand out” are more likely to be utilized in subsequent decisions (Albarracin and Kumkale 2003; Greifeneder, Bless, and Pham 2011; Siemer and Reisenzein 1998). Applying similar logic to our setting, readers should perceive arousal cues to be more diagnostic for reviews of utilitarian products (where they are less common) than reviews of hedonic products (where they are more common). As a result, readers will be more likely to draw on those arousal cues to make inferences about the reviewer and review. Therefore, the impact of expressed arousal on perceived review helpfulness will be stronger for utilitarian products than hedonic products, resulting in a “steeper” curve:

$$H_2: \text{The nonlinear effect of expressed arousal on perceived review helpfulness is greater for utilitarian products than hedonic products.}$$

**OVERVIEW OF STUDIES**

We investigated our hypotheses with four studies using distinct methodologies. In Study 1, we gathered real-world evidence in the form of reviews at Apple’s App Store. We measured helpfulness directly by user votes and measured expressed arousal through textual analysis. In Study 2, we administered a survey in which respondents assessed 400 App Store reviews in terms of helpfulness, arousal, and reviewer effort. In the final two studies, we conducted laboratory experiments that permitted the direct manipulation of expressed arousal. Studies 1–3 also investigated the role of our proposed moderator, product type.

**STUDY 1**

Our first study utilized secondary data to test $H_1$ and $H_3$ in a real-world, online WOM setting. Data consisted of user reviews from Apple’s App Store, which had accumulated nearly two years of user reviews at the time of data collection (April 2010). Visitors evaluate an app by leaving a one- to five-star rating and writing a text review (see Figure 1). Below each review, the platform asks readers the question “Was this review helpful?” and provides “Yes” and “No” options. For reviews that have received at least one vote, both the number of “Yes” votes and the number of total votes are displayed. The App Store platform provides developers a valuable channel for informing and persuading prospective customers. Moreover, the popularity of the platform ensures a wide range of apps and reviews, making it well suited to our investigation.

**Data Collection**

At the time of data collection, the App Store contained 20 mutually exclusive categories (e.g., “entertainment,” “productivity”). We began by identifying for each category the apps ranked in the top 500 by popularity over the preceding three months. Of 62,266 apps identified in this manner, 40,417 had accumulated at least one review, and we retrieved all historical reviews for those apps. For each review, we collected the following information: rating, text review content, helpful votes (i.e., the number of readers who voted “Yes”), and total votes. We also collected app-level information including the following: average rating, count of all ratings, category, and whether the app was free. Prior to the analyses, we filtered reviews with non-English characters (94,815 total), reviews without any text (2,743 total), and reviews with a rating of zero (presumably due to system errors, 38 total). Of 1,623,497 reviews that remained, 418,415 had received at least one helpful vote, and these formed the pool for our analysis.

**Variables**

To measure the perceived helpfulness of a review, we used its total number of helpful votes as the dependent variable, with the total number of votes included as a covariate.\(^1\) Importantly,
Table 1
STUDY 1: DESCRIPTIVE STATISTICS AND VARIABLE CORRELATIONS

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Helpful votes</td>
<td>2.37</td>
<td>7.10</td>
<td>0</td>
<td>984</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Total votes</td>
<td>3.87</td>
<td>9.87</td>
<td>1</td>
<td>1,031</td>
<td>.93</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Rating</td>
<td>3.45</td>
<td>1.68</td>
<td>1</td>
<td>5</td>
<td>.05</td>
<td>-0.3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Length</td>
<td>42.02</td>
<td>49.04</td>
<td>1</td>
<td>1,134</td>
<td>.12</td>
<td>.09</td>
<td>.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Reading difficulty</td>
<td>7.09</td>
<td>4.07</td>
<td>.4</td>
<td>461.6</td>
<td>.04</td>
<td>.03</td>
<td>.04</td>
<td>.33</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Emotional valence</td>
<td>1.89</td>
<td>.14</td>
<td>1</td>
<td>3</td>
<td>.00</td>
<td>-0.2</td>
<td>.33</td>
<td>-13</td>
<td>-11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7. Emotional arousal</td>
<td>1.68</td>
<td>.12</td>
<td>1</td>
<td>2.89</td>
<td>-0.2</td>
<td>-0.2</td>
<td>.10</td>
<td>-16</td>
<td>-15</td>
<td>.38</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: N = 414,336. Table 2 describes the operationalization of each variable.

the default ordering of App Store reviews during the period of data collection was not by helpfulness but rather by date posted, with more recent reviews appearing first.

We measured emotional valence and arousal using text analysis software, the Revised Dictionary of Affect in Language (RDAL; Whissell 2009). The software has been widely used to quantify emotional content in psychology and linguistics, and substantial evidence has supported its reliability and validity (Whissell 2009; Whissell et al. 1986). The RDAL dictionary contains 8,742 words characteristic of natural language and has been shown to match 90% of words in most samples. The dictionary includes not only emotional words but also words that appear commonly in natural language. Each word is assigned a score for both valence (1 = “unpleasant,” and 3 = “pleasant”) and arousal (1 = “passive,” and 3 = “active”). Whissell developed the RDAL valence and arousal scores by asking volunteers to rate random samples of words along one or the other dimension. For example, the arousal score for the word “run” is substantially higher than that for “walk” (M = 2.8 vs. M = 2.1), but valence scores for the two words are similar (M = 2.2 vs. M = 2.0).

To analyze the content of a text sample, RDAL matches each word in the sample with words in its internal dictionary; whenever a match is found, the scores for valence and arousal are retrieved. After all words in the sample have been analyzed, final scores for valence and arousal are computed by averaging retrieved scores for all identified words. Among the reviews in our set, more than 99% were assigned values for valence and arousal; words lacking either value were dropped from the analysis. Table 1 presents descriptive statistics, and Table 2 summarizes the operationalization of all variables.

To measure product type (utilitarian or hedonic) for each review, we conducted a pretest with 43 undergraduate students. Participants were asked to evaluate each of the 20 App Store categories, in random order. Definitions instructed utilitarian apps as “useful, practical, functional, something that helps you achieve a goal or solve a specific problem” and experiential apps as “pleasing and fun, something that is enjoyable and appeals to your senses” (the term “experiential” was used to aid comprehension). Each category name was presented along with a brief description of the category and names of several representative apps (see Web Appendix A). Participants answered the following, adapted from Sloot, Verhoef, and Franses (2005): “Using the scales below, how would you describe <category name> apps?” The question included two separate, nine-point items: “not utilitarian/very utilitarian,” and “not experiential/very experiential.” After compiling the results, we created a “utilitarian value” for each category by subtracting its average experiential rating from its average utilitarian rating.

As control variables in the analyses, we included three items that have been used in previous literature on antecedents of review helpfulness (Korffatis, Rodriguez, and Sicilia 2008; Mudambi and Schuff 2010): review rating, length, and reading difficulty. Review rating indicates the star rating provided in a review; ratings ranged from one star to five stars (M = 3.45 stars). Review length indicates the number of words in a review (M = 42 words). We calculated reading difficulty using the Gunning Fog index (see Gunning 1969), which estimates the years of formal education needed to understand a piece of text on first reading (M = 7 years). In addition, we controlled for the following characteristics at the app level: average rating, count of ratings, and price. Average rating serves as an indicator of overall satisfaction, while the count of ratings indicates popularity. Price was captured with a dummy variable (0 = free, 1 = paid).

Data Analysis and Results

As a preliminary visual analysis, we created the chart in Figure 2, which depicts the relationship between review helpfulness and emotional arousal in the raw data. Consistent with predictions, the figure suggests a nonlinear trend of diminishing returns.

Given that the number of helpful votes was a count variable whose variance (50.47) exceeded its mean (2.37), we utilized negative binomial regression in the formal analysis. Table 3 presents the results. To examine H1, we conducted a series of regressions in which we first entered control variables along with utilitarian value, expressed arousal, and squared arousal. As shown in Model 1, the coefficient of the squared term of arousal was negative and significant (β = -.883, p < .01), indicating a nonlinear relationship in the form of diminishing returns. Consistent with H1, increases in expressed arousal were associated with greater review helpfulness at low levels of arousal, but this positive association became weaker as arousal increased, and it eventually became negative. Using the obtained coefficients, we estimated the inflection point to be 1.69 on the three-point arousal scale.

To investigate the role of product type, we created a new model (Model 2 in Table 3) that included the interaction of utilitarian value with both arousal and squared arousal.
Consistent with H3, the estimated coefficient for the nonlinear interaction term was significant and negative ($\beta = -0.027, p < .05$), indicating that the nonlinear effect of expressed arousal on perceptions of review helpfulness was greater for app categories that were relatively more utilitarian. Figure 3 depicts the interaction graphically by plotting the expected value of review helpfulness against expressed arousal separately for more utilitarian categories (+1 SD) and more hedonic categories (−1 SD). The “steeper” curve for utilitarian versus hedonic categories in the figure indicates support for our model.

Discussion

Using a sample of real-world reviews, Study 1 provided initial evidence that the relationship between the expressed arousal in a review and perceptions of its helpfulness follows a nonlinear pattern of diminishing returns, even after controlling for other variables. The impact of additional arousal was beneficial at lower levels of arousal, but this positive effect was reduced (and turned negative) at higher levels of arousal. Consistent with our inference-based account, the nonlinear effect of arousal was greater for utilitarian apps than hedonic apps.

However, the use of field data necessitated noteworthy limitations. First, expressed arousal was measured by software rather than human perceivers. The RDAL is inherently constrained by its dictionary-matching algorithm, which ignores figurative language, grammatical markers, and other nonverbal characteristics (e.g., emoticons). Second, our framework proposes inferences regarding reviewer effort as a mechanism underlying the nonlinear impact of expressed arousal, but the design precluded measurement of such inferences. We conducted Study 2 to address these limitations. More broadly, the findings were correlational and thus subject to potential confounds; we address this concern with experiments in Studies 3–4.

STUDY 2

Our second study consisted of a survey in which respondents were given a set of real-world reviews, drawn from the pool in Study 1, and asked to provide their perceptions of each review. The design enabled us to investigate our primary hypothesis using a subjective measure of arousal and to examine our proposed mediator, perceived reviewer effort.

Stimulus Materials and Procedure

We began with the full sample of reviews collected in Study 1. After dropping reviews less than one sentence long (to ensure that content was sufficient to generate an impression), we randomly selected 400 reviews to form the pool used in the study.

One hundred twenty-eight undergraduate students participated in exchange for course credit. Participants were each shown 20 randomly selected reviews, one at a time and on separate screens. Each screen displayed the name of an app, its category, and the content of one review for that app.
Table 3  
STUDY 1: NEGATIVE BINOMIAL MODEL COEFFICIENTS  

<table>
<thead>
<tr>
<th>DV: Number of Helpful Votes</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of total votes</td>
<td>.109*** (.001)</td>
<td>.109*** (.001)</td>
</tr>
<tr>
<td>Rating</td>
<td>.154*** (.001)</td>
<td>.154*** (.001)</td>
</tr>
<tr>
<td>Length</td>
<td>.002*** (.000)</td>
<td>.002*** (.000)</td>
</tr>
<tr>
<td>Reading difficulty</td>
<td>.008*** (.001)</td>
<td>.008*** (.001)</td>
</tr>
<tr>
<td>Average rating</td>
<td>−.092*** (.002)</td>
<td>−.092*** (.002)</td>
</tr>
<tr>
<td>Number of ratings</td>
<td>−.000*** (.000)</td>
<td>−.000*** (.000)</td>
</tr>
<tr>
<td>Paid</td>
<td>.099*** (.003)</td>
<td>.099*** (.003)</td>
</tr>
<tr>
<td>Utilitarian value</td>
<td>.006*** (.000)</td>
<td>−.054* (.032)</td>
</tr>
<tr>
<td>Emotional valence</td>
<td>.083*** (.012)</td>
<td>.086*** (.012)</td>
</tr>
<tr>
<td>Emotional arousal</td>
<td>2.979*** (.185)</td>
<td>3.243*** (.197)</td>
</tr>
<tr>
<td>Arousal × Utilitarian value</td>
<td>−.883*** (.053)</td>
<td>−.966*** (.057)</td>
</tr>
<tr>
<td>Arousal × Utilitarian value</td>
<td>−.883*** (.053)</td>
<td>−.966*** (.057)</td>
</tr>
<tr>
<td>Constant</td>
<td>−2.986*** (.164)</td>
<td>−3.201*** (.174)</td>
</tr>
<tr>
<td>N</td>
<td>414,336</td>
<td>414,336</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>−681,765.4</td>
<td>−681,753.0</td>
</tr>
</tbody>
</table>

*p < .10,  
**p < .05,  
***p < .01.

Notes: Standard errors in parentheses. DV = dependent variable.

Data Analysis and Results

Because each participant evaluated different random reviews from the pool, analyses were conducted at the review level (N = 400). On average, 6.4 participants evaluated each review. For each review, we calculated average ratings for helpfulness, valence, arousal, and effort. Cronbach’s alphas exceeded .90 for all constructs, indicating satisfactory reliability. Table 4 summarizes descriptive statistics and correlations of all major variables.

To examine H1, we performed ordinary least squares regression by first entering arousal and all control variables, then adding a term for squared arousal. Results appear in Table 5 (Model 1). The coefficient for the squared term of arousal was negative and significant (β = −.120, p < .01). Consistent with H1, increased arousal was associated with greater review helpfulness at low levels of arousal, but the marginal effect of arousal became smaller as arousal increased, and it eventually turned negative. The estimated inflection point was 5.89 on the nine-point arousal scale.

To investigate the role of perceived effort, we utilized the MEDCURVE macro (Hayes and Preacher 2010), which applies a nonparametric test that accounts for nonlinear relationships (Fritz and Mackinnon 2007). Figure 4 illustrates our nonlinear mediation model, which included a quadratic effect of arousal on perceived effort and a linear effect of perceived effort on review helpfulness, in addition to a quadratic direct effect of arousal on review helpfulness.

Analyses were conducted using 5,000 bootstrap resamples, and the control variables in Table 2 were included as covariates. Estimated model coefficients appear in Figure 4. In support of H2, results indicated the presence of nonlinear mediation. The quadratic effect of arousal on perceived effort was negative and significant (β = −.108, p < .01), and perceived effort was positively associated with review helpfulness (β = .934, p < .01). After controlling for the quadratic effect of arousal on perceived effort, the quadratic effect of arousal on helpfulness became nonsignificant (β = −.019, p = .35). To better understand the pattern of indirect effects, we calculated θ, the instantaneous indirect effect of expressed arousal on helpfulness through perceived effort, at low (−1 SD) and high (+1 SD) levels of arousal. Results revealed a significant positive indirect effect at low arousal (θ = .331, 95% confidence interval = [.208, .491]), suggesting that perceived effort could explain the positive impact of arousal at low levels. Results also revealed a significant negative indirect effect at high arousal (θ = −.195, 95% confidence interval = [−.381, −.005]), suggesting that perceived effort could explain the negative impact of arousal at high levels.

To investigate the moderating role of product type, we constructed an additional model that included the interaction of utilitarian value obtained from Study 1 with both arousal and squared arousal (see Table 5, Model 2). The estimated coefficient for the nonlinear interaction term was not significant (p > .4). Thus, H3 was not supported.

Discussion

Using a survey design that allowed for subjective measures of our key variables, Study 2 replicated the main result of our field study: the relationship between expressed arousal and perceptions of review helpfulness followed a nonlinear pattern...

Notes: The curves depict expected values for review helpfulness based on Model 3 of Table 3. The utilitarian curve was generated using a utilitarian value one standard deviation above the mean (utilitarian value = 4.15), and the hedonic curve was generated using a utilitarian value one standard deviation below the mean (utilitarian value = −5.92). Mean values were used for all other variables in the model.
of diminishing returns. Consistent with our theorizing and H2, additional findings provided direct evidence that the nonlinear relationship was explained by perceptions of reviewer effort.

In contrast, Study 2 did not reveal evidence for the moderating role of product type. We speculate that this (non)result may be attributable to the format of the review stimuli, which we kept simple to limit participant fatigue across 20 trials. Given that participants saw only the name of each app and its category, with no further description, they may not have been able (or motivated) to judge with confidence whether the app was utilitarian or hedonic in nature.

Neither of our first two studies provides definitive evidence of causation, because of the potential for unobserved confounds in the real-world reviews that they utilized. In particular, the level of arousal expressed in the reviews may have been associated with their information content, argument strength, and so on; if so, then the observed effects of arousal may simply reflect differences in objective quality. We addressed these concerns in our next two studies by conducting laboratory experiments in which expressed arousal was manipulated directly.

**STUDY 3**

Study 3 examined expressed arousal and product type as manipulated (rather than measured) variables in a laboratory setting. Participants took part in a hypothetical shopping task in which they observed reviews for six different apps. The study incorporated a mixed design, with expressed arousal manipulated within-subjects at three levels (low, moderate, and high) and product type manipulated between-subjects at two levels (utilitarian and hedonic).

**Stimulus Materials**

We began by identifying real-world apps that were relatively unknown, relevant to a broad audience, and likely to evoke reviews varying in expressed arousal. To represent utilitarian and hedonic products, respectively, we narrowed our focus to the App Store “reference” and “entertainment” categories. From the reference category, we selected a location-based app that presents users with Wikipedia articles relevant to their surroundings. From the entertainment category, we selected a fish pond simulation, in which different varieties of fish swim on the screen and respond to user input.

The treatment reviews appear in Table 6 and were developed in two stages. In the first stage, we created three baseline reviews for each of the two apps. The baseline reviews were designed to provide one overall evaluative statement and one descriptive statement, with minimal emotional content. To construct the baseline reviews, we retrieved actual reviews from the App Store and then removed emotional words and phrases. For example, a baseline review for the reference app was: “High quality app. It tells you how far away you are from the landmarks and even shows those landmarks on a map.”

In the second stage, we varied each baseline review to create three treatment versions exhibiting low, moderate, or high arousal, through a combination of verbal cues and grammatical markers (Allen 1988; Schandorf 2013). The process involved four steps. First, we varied the presence and number of exclamation marks, such that sentences in the low-, moderate-, and high-arousal conditions ended with a period, an exclamation mark, or three exclamation marks, respectively. Next, we varied specific words in the text to convey different degrees of arousal (e.g., “noticed” vs. “was surprised” vs. “was completely shocked”). Third, we appended an emotional sentence to the moderate- and high-arousal reviews (e.g., “Happy with it!” vs. “Thrilled with it!!!”). Finally, we capitalized all letters in one word of each high-arousal review. The end result for each category was three sets of three reviews, varying in expressed arousal.

**Procedure**

Eighty-one undergraduate students participated in the study in exchange for course credit. Participants were randomly
assigned to the utilitarian or hedonic category. In the cover story, they were asked to imagine that they were considering purchasing a specific variety of app (either “location-based” or “fish pond”), which was then briefly described (“learn about surprising new places and things around you”; “enjoy a beautifully rendered fish pond in your pocket”). Participants were informed that an initial search had returned six alternatives, with similar prices and similar average ratings (approximately four out of five stars). Participants were then told that to aid their decisions, they would be reading and evaluating reviews of each app.

The following screens presented six text reviews, one at a time. Participants were told that each review described a different app in their consideration set. Three “filler” reviews were presented in positions 1, 3, and 5 in the sequence. The three treatment reviews were presented in positions 2, 4, and 6, and consisted of one low-arousal, one moderate-arousal, and one high-arousal review. Treatment reviews were selected randomly from the three sets in Table 6, subject to the constraint that each review came from a different set, and were presented in random order. This approach ensured that for each of the three sets, each level of arousal would be presented a similar number of times across participants.

After reading each review, participants were asked to report their perceptions of review helpfulness and perceived effort. As a manipulation check, participants also rated the level of arousal expressed in the review. Each variable was measured with three items, which included the two items used in Study 2 and one additional item (see Web Appendix D).

Results

Initial examination revealed satisfactory internal reliability for all measures (Cronbach’s alphas > .9). Analysis of the manipulation check revealed that perceived arousal in the high-, medium-, and low-arousal conditions followed the intended pattern (M = 8.26 vs. M = 6.90 vs. M = 4.14; all ps < .001). Thus, the manipulation of arousal was deemed successful.

To examine the effect of arousal on perceived review helpfulness, we conducted a mixed analysis of covariance (ANCOVA), with arousal entered as a within-subject factor, product type entered as a between-subject factor, and review order entered as a categorical covariate with six levels. The quadratic effect of arousal was specified as a polynomial contrast. Results of the analysis revealed no evidence for an overall linear effect of arousal (p > .2). Consistent with H1 and our first two studies, however, the results revealed a significant quadratic effect of arousal (F(1, 74) = 6.133, p < .05). As Figure 5 illustrates, pairwise comparisons revealed a significant increase in perceived helpfulness from low to moderate arousal (M = 5.01 vs. M = 5.63; t(80) = 2.63, p < .05) and a nonsignificant decrease in perceived helpfulness from moderate to high arousal (M = 5.34, t(80) = −1.29, p > .2).

We next investigated the role of effort in mediating effects of arousal (H2). Because the study involved a repeated-measure variable with nonlinear effects, customary methods for assessing mediation were not applicable. Therefore, we adapted the three-step procedure proposed by Judd, Kenny, and McClelland (2001) for testing linear mediation in repeated-measure designs. Given that helpfulness did not significantly differ between moderate and high arousal, we restricted the analysis to low and moderate levels. First, a repeated-measures ANCOVA revealed a positive effect of arousal on perceived effort (M = 3.24 vs. M = 4.26; t(80) = 5.41, p < .001). Second, a regression revealed that greater perceived effort was associated with greater perceived helpfulness at both low and
moderate arousal levels (low: $\beta = .62$, $t = 5.63$; moderate: $\beta = .67$, $t = 6.76$; $p < .01$). Third, differences in perceived effort predicted differences in perceived helpfulness ($\beta = .70$, $t = 6.79$, $p < .01$). Thus, all three conditions were established, suggesting that the positive effect of arousal below moderate levels was mediated by perceived effort.

Finally, we investigated the potential moderating effect of product type. The interaction between the quadratic effect of arousal and product type did not reach conventional significance ($F(1, 74) = 2.680$, $p = .11$). For exploratory purposes, however, we conducted follow-up analyses for each product type separately. Results revealed a pattern of arousal and product type. The interaction between the quadratic effect of arousal and product type was not significant ($F(1, 34) = 7.249$, $p < .05$) but was not significant for the hedonic category ($F(1, 35) = .086$, $p > .7$). Within the utilitarian category, pairwise comparisons revealed a significant increase in helpfulness perceptions from low to moderate arousal ($M = 5.39$ vs. $M = 6.25$; $t(39) = 2.48$, $p < .05$) and a marginally significant decrease in helpfulness perceptions from moderate to high arousal ($M = 5.70$, $t(39) = -1.76$, $p = .09$). Within the hedonic category, no pairwise comparison was significant ($ps > .4$).

Discussion

Extending our investigation to an experimental setting, Study 3 provided additional evidence that perceptions of review helpfulness are affected by expressed arousal in a manner of diminishing returns. In addition, mediation analyses supported our argument that the effect is due in part to inferences regarding reviewer effort. Finally, although the moderating effect of product type was not significant, follow-up analyses suggested that the nonlinear effect of arousal was stronger for utilitarian products, consistent with Study 1.

Despite the value of its experimental approach for establishing causation, conclusions from Study 3 are constrained by the fact that verbal review content was not identical across levels of arousal. Specifically, the three versions contained different emotional words (e.g., “happy” vs. “thrilled”), and reviews in the low-arousal condition contained one fewer sentence than those in the moderate- and high-arousal conditions. It is plausible that such differences influenced perceptions of either effort or helpfulness. Our next experiment addressed these concerns by manipulating arousal in a different manner, which relied exclusively on nonverbal cues.

STUDY 4

The design and procedure of Study 4 were similar to that of Study 3. However, stimuli were constructed in a manner by which arousal cues could be manipulated while holding constant the verbal content of reviews.

Stimulus Materials

Stimuli consisted of reviews for the same location-based reference apps presented in Study 3. We constructed the reviews using an approach similar to the prior study, with the
following exceptions. First, we included only positive reviews. Second, we manipulated expressed arousal using grammatical markers alone, without varying any words in the reviews. We applied two forms of grammatical markers: exclamation marks and capitalization (Allen 1988; Schandorf 2013). Reviews in the low-, moderate-, and high-arousal conditions were embedded with zero, two, and eight exclamation marks respectively, and reviews in the high-arousal conditions were embedded with capital letters. Third, in addition to the treatment versions of each review, we created a “baseline” version as a control. Baseline reviews contained the descriptive sentences without the added emotional sentence, and without any exclamation marks or capitalization. Table 7 presents the four sets of reviews that resulted from this process.

3In Studies 1 and 2, the estimated coefficient for the interaction of review rating and squared arousal was positive and significant (p < .05), suggesting that the nonlinear impact of arousal was greater for negative reviews than positive reviews. Given that valence is not part of our theoretical framework, we decided to hold valence constant in Study 4.

Procedure

One hundred fifty-seven undergraduate students participated in the study for course credit. The procedure was similar to that of Study 3. After seeing the cover story (Web Appendix E), participants read and evaluated five reviews, including an initial filler review followed by four treatment reviews. Treatment reviews were selected randomly from the four sets in Table 7, subject to the constraint that each review came from a different set, and the reviews were presented in random order. Measures of perceived helpfulness, arousal, and reviewer effort were identical to those in Study 3.

Results

Examination of the manipulation check confirmed that perceived arousal was greater for high-arousal reviews than that for moderate-arousal reviews (M = 8.38 vs. M = 6.85; t(156) = 17.62, p < .001), which was in turn greater than that for low-arousal reviews (M = 4.78; t(156) = 12.42, p < .001). Therefore, the manipulation of expressed arousal was deemed successful. Perceived arousal for the baseline reviews (M = 4.08) was lower than that for all three treatment conditions (vs. low-arousal: t(156) = −4.91, p < .001).

To examine the effect of arousal on perceived helpfulness, we conducted an ANCOVA in which arousal was entered as a within-subject factor and review order was entered as a covariate. Results revealed that the linear effect of arousal was negative and significant (F(1, 133) = 35.856, p < .001), indicating that reviews containing more arousal were on average deemed less helpful. More importantly, results revealed a significant quadratic effect of arousal (F(1, 133) = 14.008, p < .001). As we illustrate in Figure 6, pairwise comparisons revealed no significant difference in perceived helpfulness from low to moderate arousal (M = 6.61 vs. M = 6.56; p > .7), but a significant decrease in perceived helpfulness from moderate to high arousal (M = 6.56 vs. M = 5.53; t(156) = 6.29, p < .001). Additional analyses indicated that reviews exhibiting low and moderate levels of arousal were perceived as more helpful than baseline reviews (M = 6.10; tlow vs. baseline(156) = 2.96, tmoderate vs. baseline(156) = 2.64, ps < .01), but reviews exhibiting high levels of arousal were perceived as less helpful than baseline reviews (t(156) = −2.87, p < .01).

As in Study 3, we next investigated the mediating role of perceived reviewer effort by adopting the Judd, Kenny, and McClelland (2001) procedure for testing linear mediation in within-subject designs. Because helpfulness did not significantly differ between low and moderate levels of arousal, we restricted the analysis to moderate and high levels. The first step revealed a negative effect of arousal on perceived effort (Mmoderate = 4.98 vs. Mhigh = 4.56; t(156) = −3.23, p < .01). The second step revealed that greater perceived effort was associated with greater helpfulness at both moderate and high levels of arousal (moderate: β = .49, t = 7.01; high: β = .67, t = 10.44; ps < .001). The third step revealed that differences in perceived effort predicted differences in helpfulness (β = .63, t = 7.90, p < .001). Thus, all three conditions were established, suggesting that the negative effect of arousal beyond moderate levels on perceived helpfulness was mediated by perceived effort.
<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Baseline</th>
<th>Low Arousal</th>
<th>Moderate Arousal</th>
<th>High Arousal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High quality app. You can enter a location of interest, and it is able to look up places other than your current location. This is convenient for trip planning.</td>
<td>High quality app. You can enter a location of interest, and it is able to look up places other than your current location. This is convenient for trip planning.</td>
<td>High quality app! You can enter a location of interest, and it is able to look up places other than your current location! This is CONVENIENT for trip planning! I am PLEASED!!!</td>
<td>HIGH QUALITY app!!! You can enter a location of interest, and it is able to look up places OTHER THAN your current location! This is CONVENIENT for trip planning! I am PLEASED!!!</td>
</tr>
<tr>
<td>2</td>
<td>The interface is easy to use. It provides a nice graphical view of the information. Plus, it can show landmarks on a map and tell you how far away they are.</td>
<td>The interface is easy to use. It provides a nice graphical view of the information. Plus, it can show landmarks on a map and tell you how far away they are.</td>
<td>The interface is easy to use! It provides a nice graphical view of the information. Plus, it can show landmarks on a map and tell you how far away they are. I'm very happy with this.</td>
<td>The interface is EASY to use!!! It provides a NICE graphical view of the information! Plus, it can show landmarks on a map AND tell you how far away they are! I'm VERY HAPPY with this!!!</td>
</tr>
<tr>
<td>3</td>
<td>Lots of options and customizable. The text size is readable and can be adjusted. And it allows white text on black background, which is much easier on the eyes.</td>
<td>I love it. Lots of options and customizable. The text size is readable and can be adjusted. And it allows white text on black background, which is much easier on the eyes.</td>
<td>I love it! Lots of options and customizable. The text size is readable and can be adjusted. And it allows white text on black background, which is much easier on the eyes!</td>
<td>I LOVE it!!! Lots of options and CUSTOMIZABLE! The text size is READABLE and can be adjusted! And it allows white text on black background, which is MUCH EASIER on the eyes!!!</td>
</tr>
<tr>
<td>4</td>
<td>Reading articles is like flipping through a real book. It makes the information easy to access. When you tap a topic, related topics spread out from it in a nice animation.</td>
<td>I really like this app. Reading articles is like flipping through a real book. It makes the information easy to access. When you tap a topic, related topics spread out from it in a nice animation.</td>
<td>I really like this app! Reading articles is like flipping through a real book. It makes the information easy to access. When you tap a topic, related topics spread out from it in a nice animation!</td>
<td>I REALLY LIKE this app!!! Reading articles is like flipping through a REAL book! It makes the information EASY to access! When you tap a topic, related topics spread out from it in a NICE animation!!!</td>
</tr>
</tbody>
</table>
Discussion

In support of our main prediction, the results of Study 4 again revealed a nonlinear effect of expressed arousal on perceived review helpfulness, even when the verbal content of reviews was held constant. Moreover, follow-up analyses confirmed that the detrimental effects of arousal at high levels could be partly explained by inferences regarding reviewer effort. In contrast to Studies 1–3, the results of Study 4 did not reveal evidence for a beneficial effect of arousal at lower levels. Though unanticipated, this finding may simply reflect differences in calibration across the four studies (see next section).

GENERAL DISCUSSION

Supplementing an emerging stream of research on the perceived value of WOM for decision making (Chen and Lurie 2013; He and Bond 2013; Moore 2015; Mudambi and Schuff 2010; Yin, Mitra, and Zhang 2016), our research explores the consequences of emotional arousal in consumer reviews on subsequent reader perceptions. Four studies offered methodological triangulation through archival data analysis, surveys, and laboratory experiments. Expressed arousal was both measured and manipulated, using verbal cues (Study 1), nonverbal cues (Study 4), and a combination (Studies 2 and 3). Table 8 summarizes the major findings. In support of our main prediction, the four studies provided consistent evidence that expressed arousal affects reader perceptions of review helpfulness in a nonlinear manner of diminishing returns. In addition, we obtained both direct and indirect support for an inference-based account, in which the nonlinear effect of arousal operates through reader inferences regarding the effort expended by the reviewer in constructing his or her review. Evidence for the moderating role of product type was not conclusive; however, the pattern of results across studies was consistent with our claim that the nonlinear effect of arousal is magnified for utilitarian products.

Theoretical Implications

Our findings contribute to growing scholarship on consumer WOM and decision making as well as the broader topic of affective processes in communication. Existing research on determinants of review “helpfulness” has tended to focus on readily observable variables, including ratings and reviewer characteristics (Chen and Lurie 2013; Forman, Ghose, and Wiesenfeld 2008; Mudambi and Schuff 2010; Yin, Mitra, and Zhang 2016). However, an emerging trend has begun to explore review content more directly (Cao, Duan, and Gan 2011; Moore 2015; Yin, Bond, and Zhang 2014). We supplement this trend by investigating the role of expressed emotion, which is abundant in review settings but has received little prior attention.

Almost all prominent emotion frameworks include fundamental dimensions of valence and arousal (Niedenthal 2008; Russell 1980). Within the vast marketing scholarship on affect and emotion, research examining the valence dimension is plentiful, but research on the arousal dimension is rare (cf. Fedorikhin and Patrick 2010; Gorn, Pham, and Sin 2001). Reflecting this pattern, consumer WOM research has tended to focus on valence (Chen and Lurie 2013; Ludwig et al. 2013), but a few notable exceptions have explored other aspects. Yin, Bond, and Zhang (2014) examine the influence of discrete emotions (such as anxiety and anger) in consumer reviews, demonstrating that same-valence emotions often have distinct effects on reader perceptions. The present findings suggest that beyond the specific emotions conveyed, review-embedded arousal cues themselves directly influence inferences regarding both reviewer and review. Berger and colleagues have incorporated arousal directly in a model of WOM transmission (Berger 2011; Berger and Milkman 2012), showing that message content which evokes greater arousal in perceivers is more likely to be shared by recipients. Our research focuses instead on the arousal of senders, as signaled by cues in their messages, and is consistent with an “actor–observer” framework in which

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Study 1</th>
<th>Study 2</th>
<th>Study 3</th>
<th>Study 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: The effect of arousal on review helpfulness follows a trend of diminishing returns</td>
<td>Supported (inverted U-shaped)</td>
<td>Supported (inverted U-shaped)</td>
<td>Supported (diminishing returns)</td>
<td>Supported (diminishing returns)</td>
</tr>
<tr>
<td>H2: Perceived effort mediates the nonlinear effect of arousal on review helpfulness</td>
<td>—</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
</tr>
<tr>
<td>H3: Product type moderates the nonlinear effect of arousal on review helpfulness</td>
<td>Supported</td>
<td>Not supported</td>
<td>Partially supported</td>
<td>—</td>
</tr>
</tbody>
</table>
recipients use evidence of arousal to inform message-relevant judgments.

Research on affect and information processing provides a complex picture regarding the impact of emotional arousal on cognitive performance and task outcomes (Naqvi, Shiv, and Bechara 2006; Tice, Bratslavsky, and Baumeister 2001), and well-known findings in the marketing literature have linked elevated arousal to both positive and negative consumer outcomes (e.g., Isen 2001; Shiv and Fedorikhin 1999). Moreover, consumers differ in the extent to which they view affective experience as detrimental or beneficial to rational judgment (Avnet, Pham, and Stephen 2012; Epstein 1994; Hsee et al. 2003; Isen 2001). Acknowledging this ambivalence, our framework allows for a nuanced role of emotional arousal in interpersonal communication, in which the expression of additional arousal can have a beneficial, neutral, or even detrimental effect on the perceived value of message content, depending on the level of arousal already expressed and the situational context. More generally, our findings support a growing consensus that reliance on affect is heavily dependent on situational factors, even for the same individual and judgment (Pham 2009).

At a broader level, our inference-based account reflects a growing trend in which scholars are beginning to explore the various interpersonal, communicative functions served by emotion (Hareli and Hess 2012). Emotional experiences not only motivate and guide our own behavior but also provide observers with valuable social information (Van Kleef, De Dreu, and Manstead 2010). The social functions of emotion may be especially relevant to the realm of online communication, in which messages have a long “shelf life” and recipients tend to have little information about the sender or context. Our findings shed light on the process by which lay theories are applied by consumers to process the information conveyed by review-embedded emotion.

Implications for WOM Collection and Management

Many real-world review platforms offer formal guidelines for prospective reviewers. These guidelines vary considerably with respect to the expression of emotion: some platforms advise review authors to restrain their feelings, while others recommend that authors express those feelings freely (e.g., the Yelp and Amazon examples presented previously). Our findings suggest that in many contexts, neither approach is optimal: future readers may perceive reviewers who appear overly “relaxed” or “fired up” as less helpful, even when they have provided objectively useful and relevant information. Instead, an implication of our research is that guidelines should be designed to encourage a moderate level of emotional expression (e.g., “express your feelings freely, without holding back or exaggerating”). Moreover, our findings suggest that appropriate instructions may vary systematically across settings. In particular, guidelines concerning emotional expression may be especially useful for retailers in utilitarian categories, in which the effects of arousal on perceived helpfulness appear to be strongest.

Marketers are increasingly turning to online forums as a tool for customer research and relationship management (Decker and Trusov 2010; Ludwig et al. 2013). Faced with an enormous volume of communications, firms often wish to identify those messages (and senders) that are especially influential. However, no definitive measure of message influence exists, and available proxies (e.g., comments, likes, helpfulness votes) require time to accumulate. Complementing research on characteristics of viral WOM (Berger and Milkman 2012; Chen and Lurie 2013), our work offers implications for marketers working to develop methods of estimating, a priori, the extent to which specific WOM will be deemed helpful. In addition to other relevant variables (e.g., source characteristics, information content), our findings suggest that message-embedded arousal should be considered directly in this estimation. Using readily available software tools (e.g., the RDAL applied in Study 1), measurement of expressed arousal can be automated and incorporated into existing communication-monitoring tools.

Limitations and Further Research

Although findings of all four studies supported the pattern of diminishing returns predicted by H1, the specific form of the observed relationship differed. Such variance is unsurprising given the different contexts and stimuli used across the studies and may simply result from differences in calibration (i.e., arousal that seems “extreme” in one context may seem “moderate” in other contexts). Indeed, the inflection point of the observed nonlinear curve varied substantially across studies. Following this logic, the lack of a positive effect of arousal at lower levels in Study 4 may reflect the fact that “low-arousal” reviews were close to the inflection point of the nonlinear curve. In support of this speculation, a comparison of manipulation checks suggests that perceived arousal in the low-arousal condition was notably higher in Study 4 than Study 3 (M = 4.78 vs. M = 4.14). More generally, perceptions of extreme arousal may be difficult to induce in experimental settings. Nonetheless, our findings suggest an opportunity to identify specific contexts under which different patterns of nonlinearity obtain.

Our theoretical arguments (and the results of Studies 3 and 4) suggest that the nonlinear effect of expressed arousal can be explained in part by inferences regarding reviewer effort. Though effort is known to play a critical role in helpfulness perceptions (Mudambi and Schuff 2010; Yin, Bond, and Zhang 2014), we acknowledge that a wide variety of other mechanisms may contribute to the nonlinear effect. For instance, reviews exhibiting a moderate level of arousal might be processed more fluently, resulting in more favorable global evaluations (e.g., Lee and Labroo 2004). Readers may perceive high-arousal reviewers to be more subjective rather than objective (e.g., “gushing” or “venting”; Jensen et al. 2013; Park, Lee, and Han 2007), or readers may assume that such reviewers are “trying too hard” to influence prospective consumers, undermining their own credibility (Friestad and Wright 1994; Xu and Wyer 2010). The interplay of these factors is worthy of future investigation.

Similarly, our arguments for the moderating role of product type (H2) rest on the assumption that emotional cues will be deemed more diagnostic when they are unexpected or atypical. This logic is consistent with the view that emotions that “stand

---

4 A common theme among many of these alternative inferences is that the reviewer may have been thinking irrationally. Studies 2–4 collected measures of perceived reviewer rationality, and we present supplementary analyses in Web Appendix F. Results suggest that at high levels of arousal, the negative effects of additional arousal are explained in part by negative inferences about reviewer rationality.
“arousal, identity disclosure, and assumed similarity in shap-
which message recipients tend to assume underlying sim-
entry, and so on. Similarly, forums differ markedly in the
arousal in a product-related Facebook post, personal blog
review, for example, readers may expect higher levels of
channels are likely to evoke different consumer expecta-
tional Unconscious, The
Cao, Qing, Wenjing Duan, and Qweii Gan (2011), “Exploring
Determinants of Voting for the ‘Helpfulness’ of Online User
Carey, John (1980), “Paralanguage in Computer Mediated Com-