Online Reviews, Helpfulness Ratings, and Consumer Attitudes: An Extension of Congruity Theory to Multiple Sources in Web 2.0

Joseph B. Walther
Professor in Communication, Michigan State University, USA

Yuhua (Jake) Liang
Department of Communication, Michigan State University, USA

Tina Ganster
Department of Social Psychology: Media and Communication, University of Duisburg-Essen, Germany

Donghee Yvette Wohn
Telecommunication, Information Studies & Media, Michigan State University, USA

Josh Emington
Department of Communication, Michigan State University, MI

In online shopping, users may contribute reviews that are subject to overt evaluations from other users’ helpfulness ratings. Still other users may provide comments indicating agreement or disagreement with the original reviewer. This research modified congruity theory (Osgood & Tannenbaum, 1955) to predict effects on attitudes of readers who observe contributions from three potential sources of influence: review valence, other users’ aggregated helpfulness rating of the review, and another user’s verbal agreement or disagreement with the review. Results supported the hypothesized interaction effect among the 3 factors affecting attitudes toward a product, toward reviewers, and commenters. The findings suggest that congruity theory provides a framework for understanding effects from a juxtaposition of sources and messages within a participatory Web system.

Key words: Congruity theory, Online Reviews, Helpfulness Ratings, User-generated content

doi:10.1111/j.1083-6101.2012.01595.x

Participatory websites provide complex communication systems that confront readers with information from a variety of sources in a variety of forms. Web 2.0 sites that host online shopping are an increasingly important part of overall consumerism; the percentage of adult online shoppers in the US tripled from 22% in 2000 to 66% in 2008 (Horrigan, 2008). Accompanying the rise of online shopping is consumers’ reliance on user-generated reviews and ratings that appear on these sites (Willemsen, Neijens, Bronner, & de Ridder, 2011).
Research has begun to investigate the influence of user-generated reviews on other consumers’ purchase decisions. Consumers apparently regard user-generated reviews as more trustworthy than traditional advertising information (Huang, Chou, & Lan, 2007). Influential properties of these reviews include the positive (vs. negative) user-generated descriptions of products (Chevalier & Mayslin, 2006; Dellarocas, Zhang, & Awad, 2005) and the sheer volume of product reviews (Duan, Gu, & Whinston, 2005; Liu, 2006). Research also shows that source characteristics associated with the author of a product review shape consumers’ perceptions of products (Forman, Ghose, & Wiesenfeld, 2006). A variety of message features also affect responses (Dellarocas et al., 2005), such as valence, argument density, argument diversity, and the writer’s expertise claims (Willemsen et al., 2011).

Despite the growing knowledge about the effects of various components in participatory review systems, an important characteristic of these sites has eluded investigation. Most research concerning user-generated product evaluations examined one facet of these systems to the exclusion of others. However, like many other participatory websites, review sites offer a juxtaposition of multiple sources of influence (see Walther et al., 2010a), such as the users’ reviews, aggregated user representations reflecting others’ ratings of the reviews, and individual comments about reviews, provided by different individuals and groups. Decoding the interplay of these various elements when they appear simultaneously and how they affect readers’ responses comprises one of the major challenges in understanding Web 2.0 as a communication system. The present research brings theory to bear that offers specific predictions for the influence of the combination of these types of sources and messages on consumers. An original experiment tested derived predictions using mock-up online product pages.

Multiple Sources of Influence and Web 2.0 Systems

Consumers are often exposed to multiple user-generated sources and messages when shopping for products online. A typical site may offer at least three different sources of influence. First, user-generated reviews present experiences and opinions written by other consumers. The positivity or negativity of a review’s content is the valence of the recommendation (see Lee, Rodgers, & Kim, 2009). Second, other users evaluate these online reviews and generate helpfulness ratings (Otterbacher, 2009; Willemsen et al., 2011), which the site compiles and displays as aggregate ratings for yet other readers to see. These displays are often presented as a proportion of users who found others’ reviews helpful (e.g., “95 out 100 people found this review helpful”). Third, online review sites solicit comments from other readers, whose verbal statements may connote agreement or disagreement. Most comments reflect personal experiences or evaluations that support or rebut the original review. Examples of agreeing comments may include “yes, that’s what happened!” or “I also think this product is great!” Disagreeing comments may be “that’s not what happened!” or “I disagree that this product worked well!”

Predicting the relative effects of these multiple sources of social influence presents conceptual challenges. Walther et al. (2010a) suggest approaching the challenge by classifying sources in terms of their relationship to the receiver, and on that basis, applying traditional theories to help explain their likely effects. Congruity theory offers one such possibility.

Congruity Theory

Numerous theories explain the influence of an individual source’s recommendation (or rejection) of a specific product or object on a receiver’s attitude. Some of these theories deal with sequential messages, but most of these theories focus on sequential messages from a single source. Few theories explain how receivers deal with two sequential messages from two independent sources (cf., McGuire, 1964).
The present study offers a sequential modification of congruity theory (Osgood & Tannenbaum, 1955) to understand how participatory websites that feature several juxtaposed recommendations may affect subsequent readers’ attitudes.

Congruity theory specifies that when an individual source of opinion praises or disparages some object, a receiver’s attitude toward that object changes depending in part on how well the receiver regards the particular source. The theory does not focus on the quality of arguments, but whether a source makes an “associative assertion” about an object (praises, recommends, advocates, or aligns his or herself with an object) or a “dissociative assertion” about the object (disparaging, denouncing, condemning, or otherwise expressing a negative evaluation of it). The theory offers specific predictions that incorporate how well the receiver regards the source and the object before the source makes an assertion, all of which affect receivers’ attitudes toward both the object and the source in the end.

**Associative and Dissociative Statements**

When a source and an object initially differ from one another in terms of a receiver’s liking for each of these entities, and a source expresses an *association* between the source and object, the discrepancy creates degrees of pressure toward reconciliation, or congruity, among the source and the object. The receiver changes attitudes about these entities so that the evaluation of each becomes more similar. If a liked source makes an associative assertion about a disliked object, receivers’ attitudes become more favorable toward the object and less favorable toward the source. The attitude changes most for that entity (source or object) that is initially less polarized. If either entity occupies a neutral initial attitude, the terminal attitude toward that entity should match the initial attitude toward the more polarized of the two. When a *dissociative* assertion is made, receivers come to evaluate the two entities more differently from one another than before the derogation takes place. Osgood and Tannenbaum offered specific formulas for the computation of these operations and much research shows consistency with these predictions to the extent that data were amenable to analysis (see for review Insko, 1967).1

**A Two-Stage Model of Congruity Theory**

Although congruity theory was originally conceived to describe the effect of a single message from a single source, the logic of the theory offers some openness for its adaptation to multiple messages (Tannenbaum & Norris, 1965). The present application employs the theory through iteration, as a two-message/two-source model.

A first iteration predicts the effects of one source (and the source’s initial valuation) making an associative (or dissociative) assertion regarding an object (with its own respective valuation), activating congruity operations and leading to a first-stage change in attitude toward the object. In other words, if one reviewer on a web site offers praise (or disparagement) about a product in the form of a positively (or negatively) valenced review, this affects readers’ attitude about that product. The attitude toward the product that results from these operations may be considered a transitional attitude, because it becomes an input to a second iteration of the process. In a second iteration, the attitude about the product that was derived from the first stage—the transitional attitude—serves as the initial object valuation, whereupon a second source either promotes or denigrates the object through agreement or disagreement with the first source, activating the appropriate congruity operations a second time. This effect renders the receiver’s terminal attitude toward the object (as well as toward the second source).

These predictions depend on the level of regard that a reader has toward the source of an online review or comment. We assume that readers generally feel positively towards online reviewers and
commenters. This assumption is based on principles from social identification theory applied to online interaction (see for review Walther & Carr, 2010) and past work on individuals’ responsiveness toward recommendations from generalized Internet users (Sundar & Nass, 2001). It is premised on the notion that online shoppers feel affinity toward those who have also shopped online, including those who left comments, especially when they are undifferentiated with respect to their individual identities. Walther, DeAndrea, Kim, and Anthony (2010b) verified this premise in a participatory web setting by demonstrating that other users’ comments significantly affected viewers’ attitudes toward videos they watched on YouTube.

Based on this assumption of minimally positive initial regard toward online reviewers and commenters, the following congruity predictions can be derived. When a reader regards a source positively, and a reader’s attitude toward a product is initially neutral, if the source recommends the product, the receiver’s attitude toward the source does not change as a result of the recommendation, but the source’s recommendation changes the reader’s attitude toward the product from neutral to positive. If, on the other hand, the liked source derogates the product instead of praising it, the reader’s attitude toward the source still does not change, but the reader’s attitude toward the product becomes negative.

At this point, the second source’s assertion takes place, but because the reader no longer evaluates the product neutrally (due to the first source’s statement), the influence of the second source may result in a different attitude toward both the object and toward the second source. If the first (liked) source praised a product (which receivers initially viewed neutrally), and as a result the reader now likes the product, and the second, liked source also recommends the product, the product remains liked and the second source remains liked. If the first source praised the product but the second source denigrates it, readers’ attitude toward the product decline to neutral, and reader’s attitude toward the second source also declines to neutral.

If the first (liked) source disparaged the product (which receivers initially viewed neutrally), and as a result the readers dislike the product, and if the second (liked) source recommends the product, then readers’ attitudes toward the product improve to neutral, and readers’ attitudes toward the second source decline to neutral (as was the case in the previous example). Finally, if the first source disparages the product and the second source also disparages the product, readers’ attitudes toward the product remains negative and their attitude toward the second source remains positive.

Although these formulations are sufficient to generate hypotheses, the relationships are also susceptible to interaction effects with yet another factor, and therefore, hypotheses are specified after further explication.

Helpfulness Ratings

The preceding predictions include the effects on attitude change due only to review valence and comment agreement (or disagreement). However, in participatory websites, consumers may view aggregate user representations (AURs) that evaluate reviews. In the following section, we review the basis for predicting the effect of one such AUR—helpfulness ratings—on the level of regard for reviewers, with implications for congruity hypotheses.

A number of online retailers allow consumers to signal their opinions about others’ product evaluations. The question, “Was this review helpful for you?” may be answered by the choice between helpful and unhelpful. A site may display the number or ratio of other users who found the review helpful or unhelpful. In research that examined the effects of helpfulness ratings on readers, Metzger, Flanagin, and Medders (2010) suggested that the helpfulness ratings of a review signal public endorsement of a particular message, which may be especially likely to affect readers who have no experience with a
product. Otterbacher (2009) found that reviews that are rated more helpful are associated with top selling products compared to those written about less popular products, but it is unclear whether useful reviews drove popularity or popularity drove useful reviews.

Social identification principles, mentioned above, also suggest that the opinions of previous users should affect readers’ evaluations of reviews. Especially when one’s group members are visually anonymous, social identification enhances overattributions of similarity and a tendency to conform to the group’s norms and preferences (Postmes et al., 2001). An AUR by its nature provides no indication of the individual identity of any of the contributors to that coefficient. Therefore, when readers see the aggregated opinions of other people like themselves, those ratings should cause them to evaluate matters similarly as their predecessors.

If the impact of users’ helpfulness ratings extend from evaluations of the reviews to include readers’ impressions of the reviewers (the sources of the reviews), this extension reconnects these effects to congruity theory. The influence of helpfulness ratings may affect readers’ initial evaluations of the primary reviewer (not the comments or commenters that follow up on the initial review). The helpfulness rating should therefore modify valuation of the reviewer in the first stage of the two-stage congruity process described above. These additional principles, and the previous arguments adapting congruity theory as a two-stage process of source effects on attitudes toward a product, generate the following hypotheses:

H1: Readers’ attitudes toward products result from an interaction between review valence, helpfulness ratings, and comment agreement, the impacts of which are also affected by readers’ level of regard for each of the sources.

H2: Readers’ attitudes toward reviewers result from an interaction between review valence, helpfulness ratings, and comment agreement, the impacts of which are also affected by readers’ level of regard for each of the sources.

H3: Readers’ evaluations of commenters result from an interaction between review valence, helpfulness ratings and comment agreement, the impacts of which are also affected by readers’ level of regard for each of the sources.

Specifically, the two-stage congruity process occurs as follows: Positive helpfulness ratings may improve attitudes toward already well-regarded sources; if readers initially regard reviewers somewhat positively by virtue of social identification, readers may like reviewers even more who are rated as helpful. Then congruity principles predict even more change in the transitional attitude toward the object (product) to which that reviewer refers because the readers’ attitude towards the reviewer is even more polarized. The valence of the transitional attitude still depends on the reviewer’s associative or dissociative statement about the product, and the transitional attitude becomes the initial valuation of the product in the second iteration (which otherwise operates the same as before). The predictions derived from congruity theory appear in Table 1, expressed in terms of positive or negative numerical weights calculated from the theoretical formulas.

**Method**

Participants (N = 410) were undergraduate students at a large, Midwestern U.S. university who participated in exchange for credit as part of a departmental subject pool. Demographic analyses
Table 1 Hypothesized Congruity Scores (Contrast Weights) for Reviewer, Commenter, and Product Attitudes

| Unhelpful Reviews of Neutrally Perceived Product | Predicted Terminal | Predicted Terminal | Predicted Terminal |
| Liked Reviewer (+1) | Transitional Product Attitude | Liked Commenter (+1) | Product Attitude | Reviewer Attitude | Commenter Attitude |
|---|---|---|---|---|
| Assertion | Associative | +1 | Associative | +1 | +1 | +1 |
| Associative | +1 | Dissociative | 0 | +1 | 0 |
| Dissociative | −1 | Associative | 0 | +1 | 0 |
| Dissociative | −1 | Dissociative | −1 | +1 | +1 |

| Helpful Reviews of Neutrally Perceived Product | Predicted Terminal | Predicted Terminal | Predicted Terminal |
| Helpful Liked Reviewer (+2) | Transitional Product Attitude | Liked Commenter (+1) | Product Attitude | Reviewer Attitude | Commenter Attitude |
|---|---|---|---|---|
| Assertion | Associative | +1.67 | Associative | +2 | +1.67 |
| Associative | +1 | Dissociative | +2 | +1 | −1 |
| Dissociative | −1 | Associative | −1 | +2 | −1 |
| Dissociative | −1 | Dissociative | −1.67 | +2 | +1.67 |

\[ t(407) = 19.35^* \]

\[ r_{\text{contrast}} = .69 \]

\[ .98 \]

\[ .87 \]

Note. *p < .001.

revealed moderate diversity in gender (42% female), ethnicity, and age (\(M = 19.94\), range = 18–23). Computer-based randomization assigned participants to one of eight stimulus conditions resulting from the \(2 \times 2 \times 2\) experimental design related to the valenced content of a product review (positive/negative), how helpful (helpful/unhelpful) other users appeared to have rated that review, and a responding comment to the review (agree/disagree).

Participants initially completed a questionnaire about their involvement with the product (vegetable oil). Then, they viewed a web page displaying an experimental stimulus reflecting one of eight conditions. When participants felt that they had examined the web page sufficiently, they completed the induction check and dependent measures.

**Experimental Inductions**

**Involvement**

The product should arouse little involvement in order for participants’ initial attitude about the product to be near neutral. Relying on previous research indicating its relatively low level of consumer
involvement (Kuenzel & Musters, 2007), the researchers chose vegetable oil as the focal product. Participants in the present study reported a mean involvement score of 3.53 on a scale ranging from 1 (low) to 7 (high). (Some participants added comments at the end of the study, including “I do not care about vegetable oil. I will buy whatever is on sale,” and “The topic, vegetable oil, was a bit difficult to have feelings toward either way.”)

Message valence
The research involved a number of steps to create stimuli that differed in valence, while maintaining authenticity. First, researchers selected a number of existing statements from reviews of vegetable oil on Amazon.com, and through consensus, categorized each of them as positive or negative based on their face characteristics. In order to develop 30 positive and 30 negative statements for further testing, additional statements were created by reversing the polarity connotations of existing statements. To identify the most positive and negative statements empirically, 91 offset participants (from the same subject pool as those who completed the main experiment) viewed the statements and rated the valence of each on a 7-point semantic differential, very negative to very positive. Researchers retained 15 of the most positive statements (with mean scores of 5.61 to 5.91) and 15 of the most negative (1.72 to 2.30). Using subsets of each 15-statement set, five sets of positive and five sets of negative statements were constructed containing three statements apiece. Each 3-statement combination constituted a stimulus in the experiment to which participants were randomly assigned. After data collection, the perceived valence of the sets did not differ statistically within either the positive condition, \( F(4, 203) = .98, p = .42 \), or the negative condition, \( F(4, 206) = 1.66, p = .16 \), so further analyses collapsed the sets within conditions. A sample of the messages appears in Appendix A.

Helpfulness
In order to indicate the apparent helpfulness of the reviews, the stimuli featured one of two different ratios of previous users who rated the review as helpful. To select the baseline number of ratings for these ratios, the researchers sampled reviews of the most popular products on Amazon.com and averaged the total number of ratings per review, which was 170. The next task was to select the experimental values for the proportions of those ratings which suggested helpful vs. unhelpful, in order to maximize experimental variance and minimize the risk of artificiality. The ratio “167 out of 170” signified the helpful condition, and “3 out of 170” formed the unhelpful stimulus. These ratios appeared on the stimulus pages in the sentence: “167 of 170 people found the following review helpful” or “3 of 170 people . . .”.

Comment
A verbal comment appeared on each stimulus page that expressed agreement or disagreement with the original review. The form of the comment was the same as is commonly found on Amazon.com, where commenters often state explicitly that they agree or disagree with the review located on the page. Statements also appeared refuting reviewers’ claims, adding no additional points. Comment examples appear in Appendix A.

Stimuli
Graphics software enabled the creation of eight mock Amazon.com web pages, reflecting the experimental variations in review, comment, and helpfulness factors (see Figure 1). The general design of these pages closely resembled Amazon.com’s product pages by using the same font size and style, logo, layout, and several other attributes. However, the stimuli did not include other elements on
Figure 1  Example of a stimulus page with a positive review, helpful rating, and agreeing comment.

typical Amazon.com pages such as product features, advertisements, and recommendations of other products. The page included a picture of a bottle of vegetable oil that is not actually sold in the United States, in order to avoid evoking prior brand involvement. Next to the picture of the oil, the simulated page included the name of the oil, its price, one user-generated review comprised of the three similarly valenced statements, the helpfulness sentence containing the ratio of people who rated the review as helpful, and a comment from another user that directly agreed or disagreed with the review.

Measures

Involvement
The Personal Involvement Inventory for Advertising (Zaichkowsky, 1994) measured participants’ involvement with the product of vegetable oil. It features ten 7-interval semantic differential items, e.g., important/unimportant, boring/interesting, and involving/uninvolving. Reversed items were recoded and scores were averaged so that greater scores reflected more involvement, Cronbach’s $\alpha = .90$.

Attitude toward product
Holbrook and Batra’s (1987) consumer attitude measure assessed participants’ terminal attitude toward the product. The scale included four 7-interval items (good/bad, favorable/unfavorable, like/dislike, and negative/positive), $\alpha = .96$. After recodes, greater scores reflected more favorable attitudes.
Source evaluation
A modified version of the source credibility scale (McCroskey & Teven, 1999) assessed participants’ final attitude about the sources’ trustworthiness using four 7-interval items (honest/dishonest, trustworthy/untrustworthy, moral/immoral, ethical/unethical). Participants completed the measure related to the reviewer and the commenter separately, both administrations achieving $\alpha = .90$.

Results

Induction Checks
Induction checks assessed whether the experimental manipulations aroused the expected differences in perceptions of reviews’ valence, helpfulness, and commenters’ agreement. After reviewing the experimental stimuli, participants completed a number of 5-interval semantic differential measures. For review valence, participants indicated if the review was positive/negative, good/bad, and supportive/unsupportive of the product ($\alpha = .99$). Participants rated the intended positive valence stimuli as better ($M = 4.68, SD = 0.77$), than the negative valence stimuli ($M = 1.44, SD = 0.99$), $t (413) = 37.12, p < .001$. The helpfulness ratings’ effect was verified by a 5-interval scale on which participants rated the review from unhelpful to helpful. The stimulus in which 167/170 users ostensibly rated the review as helpful garnered significantly greater perceived helpfulness ($M = 4.17, SD = 1.25$) than the condition in which only 3/170 users appeared to rate the review as helpful ($M = 2.88, SD = 1.69$), $t (411) = 8.87, p < .001$. Participants evaluated whether the comment agreed or disagreed with the review using a three-item, 5-interval semantic differential measure indicating agree/disagree, consistent/inconsistent, and approve/disapprove. The two versions of the comment differed significantly, with the intended agreement version leading to more perceived agreement ($M = 4.46, SD = .90$) than the disagreement version ($M = 1.89, SD = 1.27$), $t (413) = 23.87, p < .001$.

Hypothesis Tests
Contrast analysis provides a method to test whether specific, predicted patterns among means are reflected by the data (see Keppel & Wickens, 2004). In the present case, the formulas presented by congruity theory (and its extension as a two-stage model) present mathematical values that should correspond to cells containing each combination of stimulus factors. For instance, using congruity theory calculations, as depicted in the top row of Table 1, participants who read that a reviewer (who is not helpful) positively evaluates the product, with a commenter agreeing, yield a final attitude toward the product that results in a positive score (+1). For participants who see the reviewer positively evaluate the product, with the commenter disagreeing, we expect their final attitude about the product to be neutral (0). We also expect a neutral attitude for participants who see the reviewer negatively evaluate the product with the commenter disagreeing (0), whereas for those who see the reviewer evaluate the product negatively with the commenter agreeing, their final attitude toward the product should be negative (−1). The pattern among these numbers, derived from the congruity calculations, should reflect the pattern among the mean scores of participants’ attitudes, the nonchance probability of which is what contrast analysis tests. On that basis, we assigned contrast weights to expected means to reflect the specific predicted directions among review valence, helpfulness, and comment agreement with respect to their effects on readers’ attitude toward product, reviewer, and commenter (see Table 1).

For H1, the data were consistent with the predicted effects, $t (407) = 19.35, p < .001$, $r_{\text{contrast}} = .69$. These results indicate that the pattern of observed means for readers’ attitudes about the product reflected the congruity-based specifications due to the mutual effects of (a) the liking of the reviewer, which was
enhanced when helpfulness ratings were more positive, (b) associative or dissociative assertions (praise or derogation of the vegetable oil) in the initial review, and (c) associative or dissociative assertions (praise or derogation) by the commenter. See Table 2 for means and standard deviations.

Hypothesis 2 predicted that the terminal evaluation of the reviewer also followed the congruity theory specifications, including the prediction that liking for the reviewer is more positive when helpfulness ratings were greater compared to when helpfulness ratings were lower. The data were consistent with these predictions, $t(407) = 11.28, p < .001, r_{\text{contrast}} = .97$.

Evaluations of the commenter also showed a significant effect in the manner predicted by H3, $t(407) = 6.78, p < .001, r_{\text{contrast}} = .87$. These results indicate that the final evaluation of commenters was influenced by a constellation of factors, including the first-stage degree of liking toward the reviewer and the reviewer’s assertion about the product, which in turn affected the transitional attitude about the product, whereupon moderately liked commenters associated or dissociated themselves with the product, which affected readers’ attitudes toward the commenter (H3) as well as it did their attitude toward the product (H1).

Table 2  Means and Standard Deviations for Interactions of Review, Helpfulness, and Comment on Attitudes Toward Product, Reviewer, and Commenter

<table>
<thead>
<tr>
<th>Review Valence</th>
<th>Helpfulness Rating</th>
<th>Comment Agreement</th>
<th>n</th>
<th>Product</th>
<th>Reviewer</th>
<th>Commenter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Positive</td>
<td>Helpful</td>
<td>Agree</td>
<td>50</td>
<td>4.83</td>
<td>1.39</td>
<td>4.76</td>
</tr>
<tr>
<td>Positive</td>
<td>Unhelpful</td>
<td>Agree</td>
<td>53</td>
<td>4.87</td>
<td>1.03</td>
<td>4.57</td>
</tr>
<tr>
<td>Positive</td>
<td>Helpful</td>
<td>Disagree</td>
<td>51</td>
<td>4.75</td>
<td>0.94</td>
<td>4.51</td>
</tr>
<tr>
<td>Positive</td>
<td>Unhelpful</td>
<td>Disagree</td>
<td>50</td>
<td>3.89</td>
<td>1.63</td>
<td>4.02</td>
</tr>
<tr>
<td>Negative</td>
<td>Helpful</td>
<td>Agree</td>
<td>51</td>
<td>2.41</td>
<td>1.40</td>
<td>4.84</td>
</tr>
<tr>
<td>Negative</td>
<td>Unhelpful</td>
<td>Agree</td>
<td>53</td>
<td>2.36</td>
<td>1.33</td>
<td>4.75</td>
</tr>
<tr>
<td>Negative</td>
<td>Helpful</td>
<td>Disagree</td>
<td>54</td>
<td>3.57</td>
<td>1.25</td>
<td>4.47</td>
</tr>
<tr>
<td>Negative</td>
<td>Unhelpful</td>
<td>Disagree</td>
<td>53</td>
<td>3.56</td>
<td>1.36</td>
<td>4.44</td>
</tr>
</tbody>
</table>

Follow-up Tests

In order to assess the robustness of the effects produced by the contrasts reflecting congruity-theoretic predictions, further analyses compared these findings to those of alternative models of simple effects. Six orthogonal series of contrast weights reflected main effects of review valence, helpfulness rating, comment agreement, and all possible two-way interactions among these variables, for their expected influence on terminal attitude toward the product (see Table 3). Analysis of these alternative models indicates significant effects due to review valence, $t(407) = 12.59, p < .01, r_{\text{contrast}} = .53$, and commenter agreement, $t(407) = 6.69, p < .01, r_{\text{contrast}} = .31$. A valence by comment interaction effect also emerged, $t(407) = −2.25, p = .03, r_{\text{contrast}} = .11$. Although some of these factors received statistical support, the directions derived from congruity theory reported above produced a greater effect size ($r_{\text{contrast}} = .69$) than any of these alternatives, the closest of which was significantly worse, $Z = −3.66, p < .001$, indicating that the hypothesized double-congruity model accounted for participants’ responses better than unhypothesized main and interaction effects of the same factors.
Table 3  Alternative Model Tests on Attitude Toward Product

<table>
<thead>
<tr>
<th>Review Valence</th>
<th>Helpfulness</th>
<th>Comment Agreement</th>
<th>Valence Effect</th>
<th>Helpfulness Effect</th>
<th>Comment Effect</th>
<th>Valence x Helpfulness</th>
<th>Helpfulness x Comment</th>
<th>Valence x Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Low</td>
<td>Disagree</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Negative</td>
<td>Low</td>
<td>Agree</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Negative</td>
<td>High</td>
<td>Disagree</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Negative</td>
<td>High</td>
<td>Agree</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td>Low</td>
<td>Disagree</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td>Low</td>
<td>Agree</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td>High</td>
<td>Disagree</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td>High</td>
<td>Agree</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\( t(407) = 12.59, 1.47, 6.69, 1.59, -1.85, -2.25 \)

\( p < .01, .14, <.01, .11, .07, .05 \)

\( r_{\text{contrast}} = .53, .07, .31, .08, .09, .11 \)

Beyond the extension of congruity as a two-stage model, the most tenuous extension of the present work may be the inclusion of helpfulness ratings as a factor presumed to affect initial attitude toward the reviewer. Moreover, the impact of this application is a basis for all other effects that are tested. To assess the utility of this approach compared to a more traditional congruity approach in which initial attitudes toward sources are not affected by external social factors, the analyses were recomputed without the effect of helpfulness ratings. This alternative assumes that the reviewer was regarded mildly positively in each condition, which is therefore not testable, but its impact can be seen in participants’ observed attitudes toward the product and toward the commenter. The contrast test that held constant mildly positive regard for the reviewer produced a significant effect of assertion valence and commenter agreement on product attitude, \( t(407) = 13.61, p < .001, r_{\text{contrast}} = .53. \) However, the effect size for this result was significantly poorer than that for the model that included the variable influence of helpfulness ratings \( (r_{\text{contrast}} = .69), Z = -3.26, p = .001. \) The effect of assertion valence and commenter agreement on attitude toward the commenter, without including the influence of helpfulness ratings on regard toward reviewers, did not achieve significance, \( t(407) = 0.27, p = .79. \)

Discussion

This research explored the juxtaposition among different types of user-generated messages on consumers’ perceptions of a product depicted online. Specifically, user-generated content included the reviews that an individual provided about a product, users’ aggregated ratings of the helpfulness of the review, as well comments that expressed agreement or disagreement with the original product evaluation. A specific, predicted interaction among all three of these factors accounts best for their influence on readers.

This investigation proposed a novel modification of congruity theory to predict possible outcomes on consumers’ attitudes about a product, the reviewer, and the commenter. Two sources, one after the other, expressed either a positive or a negative opinion about a product. The two-step congruity model assumed that the first source’s (the reviewer’s) evaluation of the product leads in a specific way to a transitional attitude toward the product which is either positive or negative, depending on (a) whether the reviewer makes an associative or dissociative assertion and (b) how well the reviewer is regarded by a reader, which is determined in part by other users’ usefulness ratings of the review.
This transitional attitude is subject to change when readers consider the second source’s evaluative comment. The findings support congruity theory’s predictions about readers’ attitudes toward both the sources and the product. Results also support assumptions about helpfulness ratings and their effect on initial attitude toward the reviewer. Assertions by reviewers who appeared to be more helpful generated a stronger effect on readers’ attitude about the reviewer and a correspondingly stronger transitional attitude about the product. Our adaptation of congruity theory suggests that the effect of the assertion introduced by the second source is not as strong as the first one when a helpfulness rating affects the first source’s credibility, which ultimately affects terminal attitude scores accordingly.

Congruity theory depends upon specifying the initial levels of regard participants attribute to a source (i.e., reviewers and commenters). Based on previous studies and social identification principles, we assumed that readers regard reviewers and commenters in a mildly positive manner, all other things being equal. However, no pretest or manipulation check substantiated this assumption. Nevertheless, because the initial value of participants’ regard for sources is represented as a numeric value in congruity-theoretic calculations, the assumption can be tested by comparing the variance accounted for by analysis using the hypothesized value (+1) vs. a model in which the initial regard toward the reviewer or commenter is some value other than +1. A reanalysis of the data using contrast weights reflecting an initially neutral regard toward reviewers (0) produced an effect $r_{\text{contrast}} = .43$, slightly less than that obtained by the hypothesis test that assumed mildly positive regard ($r_{\text{contrast}} = .56$). Replications using values reflecting even less initial regard for the reviewer fit the data more poorly. The assumption that readers view reviewers and commenters positively seems to be a realistic one.

The interpretation of the results should be taken in light of the study’s limitations. First, participants received brief exposure to the stimuli, unlike an actual e-commerce site where they would be able to follow various links to other information and reviews, and cycle back to the product page indefinitely. The static stimuli in this study limit the generalizability of the findings. A second limitation is the initial neutral evaluation of and low involvement with the product in this study. Readers’ involvement with a product certainly affects how they evaluate products and how reviews affect them (Park, Lee, & Han, 2007). Future research may examine products with which readers have different prior experiences and evaluations. Third, this study concerned the context of a single review and comment. As users are usually confronted with multiple reviews or comments, the question arises whether the sequential model of congruity holds for more than two iterations. With more reviews, the use of other strategies to reduce cognitive effort might increase (e.g., primacy/recency effects), especially under low-involvement conditions.

In this framework, helpfulness ratings provided readers important information about reviewers, which colored all readers’ subsequent evaluations. Readers may use helpfulness ratings in other ways, such as heuristics about what to read in the first place, not just how to evaluate reviews post hoc. In practice, readers’ opportunities to exploit helpfulness ratings as ways to select reviews or evaluate reviewers are limited in some cases. For better or worse, some online systems such as Amazon.com automatically alter the presentation of reviews based on the relative rank of their aggregated helpfulness scores. Whether this approach helps readers or limits their decisions remains to be learned.

In conclusion, the predictions and results in the present research, although somewhat complex, offer an advancement in our understanding of online product reviews by considering the effects of user-generated reviews in combination with other sources of user-generated product and source evaluations. Previous research results examining only the effects of positive or negative reviews on consumers’ attitudes (e.g., Chevalier & Mayslin, 2006; Dellarocas et al., 2005) were replicated in this study, in the main effect of review valence on product attitudes. However, this effect was overridden by a specific directional interaction between review valence, helpfulness ratings, and the agreement/disagreement
of a secondary comment, reinforcing the importance of examining multiple sources of influence in juxtaposition to one another in Web 2.0 communication systems.

Notes

1 According to Osgood and Tannenbaum (1955) as summarized in Insko (1967), the congruity pressure for two objects linked with an associative assertion is calculated:

\[ P_{OJ1} = d_{OJ2} - d_{OJ1} \]  
\[ P_{OJ2} = d_{OJ1} - d_{OJ2} \]

For a dissociative assertion, the pressure is stated as:

\[ P_{OJ1} = -d_{OJ2} - d_{OJ1} \]  
\[ P_{OJ2} = -d_{OJ1} - d_{OJ2} \]

where \( P_{OJ} \) symbolizes total pressure toward congruity for the first object of judgment, \( d_{OJ} \); and \( P_{OJ2} \) symbolizes the total pressure toward congruity for the second object of judgment, \( d_{OJ2} \). The attitude toward the more polarized object of judgment does not change as much as the less polarized object. The change formulas are stated as:

\[ AC_{OJ1} = \frac{|d_{OJ2}|}{|d_{OJ1}| + |d_{OJ2}|} P_{OJ1} \]  
\[ AC_{OJ2} = \frac{|d_{OJ1}|}{|d_{OJ1}| + |d_{OJ2}|} P_{OJ2} \]

In these equations, \( AC_{OJ1} \) represents the change in attitude toward the first object of judgment (e.g., a source, i.e., a reviewer) and \( AC_{OJ2} \) represents change in attitude toward the second object (e.g., a product). Positively valenced reviews serve as associative assertions whereas negatively valenced reviews serve as dissociative statements. Specific predictions for the transitional attitude toward the product can be deduced from the congruity model regarding a reader’s attitude toward product and reviewer (see Table 1).

The first iteration shapes the reader’s valuation of the product, and this value then becomes \( d_{OJ2} \) in the second iteration. When a commenter provides a statement about the product in form of agreement or disagreement with the initial reviewer (which comprises associative or dissociative assertions about the product, resp.), the congruity formulas are applied a second time using these assertions, liking toward the second source, and the evaluation of the product as it was previously affected in Stage 1 (see Table 1).

2 Additional analyses add confidence in the hypothetically derived contrast analysis. When experiments involve a number of conditions there are several possible patterns of differences among means. Significant contrast tests indicate that “we found what the theory led us to look for,” according to Keppel and Wickens (2004, p. 81), but “we also need to know how well the predicted pattern matched the outcome—whether unexpected or unanticipated differences may also be present.” Keppel and Wickens therefore recommend analysis of the remaining variance among
scores that is not accounted for by theoretically specified contrasts. A nonsignificant test of residual variance suggests that the planned contrasts account for the majority of the variance in the data. We therefore analyzed the residual variance in participants’ attitudes about the product (after the effect of review valence, mild regard for the commenter, comment valence, and the role of helpfulness ratings on perceptions of the reviewer). The result, $F(7, 407) = .613, p = .75$, indicated that variance unaccounted for by the hypothesized contrasts was not significant. A reduced model without the effect of helpfulness on perceptions of the reviewer also proved to be adequate, residual $F(7, 407) = 1.07, p = .38$. However, with respect to participants’ final attitude toward the commenter, the reduced model (without helpfulness ratings’ effect on perceptions of the reviewer) did not render the unexplained residual variance nonsignificant, $F(7, 407) = 46.26, p < .001$, whereas the full model with all hypothesized factors did, $F < 1$.

References


About the Authors

Joseph B. Walther, Ph.D. (http://www.msu.edu/~jwalther/) is Professor in Communication, in Telecommunication, Information Studies & Media, and in the Center for Advanced Studies in International Development at Michigan State University. His research focuses on interpersonal, group, and educational applications of computer-mediated communication and social media. Address: 404 Wilson Road, Rm. 472, East Lansing, MI 48824 USA.

Yuhua (Jake) Liang, M.A. (liangyuh@msu.edu) is a doctoral student in the Department of Communication at Michigan State University. His research focuses on persuasion in social media and Web 2.0 environments. His previous work appears in the Journal of Communication and Media Psychology. Address: 404 Wilson Road, Rm 554, East Lansing, MI 48824 USA.

Tina Ganster (M. Sc.) is a research assistant and graduate student at the department of Social Psychology: Media and Communication at the University of Duisburg-Essen, Germany (http://www.uni-duesseldorf.de/sozialpsychologie/ganster.shtml). Her research focuses on social influence processes and the
impact of nonverbal cues within computer-mediated communication. Address: Building LE 218, Forsthausweg 2, 47057 Duisburg, Germany.

Donghee Yvette Wohn (yvettewohn.com) is a PhD student in Telecommunication, Information Studies & Media at Michigan State University. Her research focuses on the social and psychological effects of non-conscious behavior in online social environments. Address: 404 Wilson Rd. Rm. 409, East Lansing, MI 48824 USA.

Josh Emington is an Adjunct Professor in Communication for Adrian College and a Master’s candidate in the Department of Communication at Michigan State University. His research interests include interpersonal, group, and organizational communication with special attention to computer-mediated interactions. Address: 2299 Knob Hill Drive, Apt 4, Okemos, MI 48864

Appendix A

Example Review Valence Induction and Agreement or Disagreement Messages.

<table>
<thead>
<tr>
<th>Review Valence</th>
<th>Review Message</th>
<th>Agreement Message</th>
<th>Disagreement Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>I bought it, and was overwhelmingly satisfied with the results! It’s an excellent product. I can personally guarantee that this is a great oil.</td>
<td>I definitely agree with the reviewer! I was overwhelmingly satisfied with the results! I personally guarantee that it’s a great oil.</td>
<td>I disagree with the reviewer. I bought it and was NOT satisfied with the results! I do not think it’s an excellent product.</td>
</tr>
<tr>
<td>Negative</td>
<td>I was really disappointed in this oil. This is one of the lowest-quality oils I have tried. This oil completely fails.</td>
<td>I definitely agree with the reviewer! I was also very disappointed after using the oil. The quality of the oil is extremely low.</td>
<td>I disagree with the reviewer. I was not disappointed at all. I think it was one of the highest-quality oils I have ever tried.</td>
</tr>
</tbody>
</table>