Does Paying For Online Product Reviews Pay Off? the Effects of Monetary Incentives on Content Creators and Consumers

Andrew Stephen, University of Pittsburgh, USA
Yakov Bart, INSEAD, Singapore
Christilene Du Plessis, INSEAD, Singapore
Dilney Goncalves, IE Business School, Spain

We show that although incentivizing product reviews results in more helpful content, disclosing payment lowers review consumers’ product quality expectations. This effect occurs because disclosure induces doubt in product quality and persists when more objective information is available, irrespective of disclosure specificity, across product categories and even after product trial.

[to cite]:

[url]:
http://www.acrwebsite.org/volumes/1012713/volumes/v40/NA-40

[copyright notice]:
This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at http://www.copyright.com/.
New Directions in Word-of-Mouth
Chairs: Ezgi Akpinar, Erasmus University, The Netherlands
Jonah Berger, University of Pennsylvania, USA

Paper #1: When Controversy Begets Conversation
Zoe Chen, Georgia Tech, USA
Jonah Berger, University of Pennsylvania, USA

Paper #2: Rating with Confidence: Rating Certainty and Word-of-Mouth Behavior
Yu-Jen Chen, University of Maryland, USA
David Godes, University of Maryland, USA

Paper #3: Valuable Virality: The Effect of Advertising Appeals and Brand Integralness
Ezgi Akpinar, Erasmus University, The Netherlands
Jonah Berger, University of Pennsylvania, USA

Paper #4: Does Paying For Online Product Reviews Pay Off? The Effects of Monetary Incentives on Content Creators and Consumers
Andrew Stephen, University of Pittsburg, USA
Yakov Bart, INSEAD, France
Christilene Du Plessis, INSEAD, France
Dilney Goncalves, IE Business School, Spain

SESSION OVERVIEW
Talking with others is a fundamental consumer motive. People tell friends about new restaurants, forward online content to coworkers, and post online reviews about products they like (or hate). Further, these behaviors have a large impact on consumer behavior. Word-of-mouth influences the products people buy, movies they watch, and online communities they join (Chevalier and Mayzlin 2006; Trusov, Bucklin and Pauwels 2009; Huffaker et al. 2011). Consequently, there has been lots of recent interest in word-of-mouth and social media (e.g., Facebook and Twitter). Firms spend a good deal of effort trying to generate word-of-mouth (e.g., encouraging people to write reviews) and create content that consumers will share.

But while it is clear that word-of-mouth is both frequent, and important, less is known about how firm activities shape WOM generation and its consequences. Intuition suggests that more controversial ads will be talked about more, but is that really the case? Companies often ask consumers to rate their experiences, but could using a 5 vs. 100 point scale affect consumer’s propensity to share WOM? How can brands create valuable virality, or content that is not only shared but also has positive downstream consequences for the brand (e.g., boosts evaluation or purchase)? How does paying people to write reviews impact review quality and how does disclosure of payment influence consumer evaluations?

This session examines these, and related questions, as it integrates various research perspectives to examine how firm activities affect both the generation and outcomes of word-of-mouth. Chen and Berger show that while controversy makes topics more interesting to talk about, it also reduces comfort. Consequently, the effect of controversy on conversation depends on how aspects of the situation (e.g., identity disclosure) and the audience (friends vs. strangers) shape these complementary mechanisms. Chen and Godes show that rating an experience on a 5-point (vs. 100-point) scale leads to higher WOM intention. These results reflect a rating-certainty effect where the rater believes that her rating captures accurately her underlying utility. Akpinar and Berger provide evidence that while soft sell ad appeals (e.g., stories or cool content) are more likely to be shared than hard sell appeals (e.g. emphasizing product features), they are less likely to have a positive impact on purchase likelihood and brand evaluation. Consequently, to create valuable virality, brands must make soft appeal ads where the brand is an integral part of the narrative. Finally, Stephen et al. show how paying for reviews and disclosing payment impacts actual and perceived review quality. While paying consumers actually leads to higher quality reviews and does not negatively impact perceptions of review quality, disclosure can either increase or decrease perceptions of product quality because of the contrast between expectations and actual experience.

Consistent with ACR 2012’s theme, this session offers diversity in several ways. First, while prior work usually examines the drivers and effects of word-of-mouth separately, this session bridges the gap to simultaneously examine the causes and consequences of word-of-mouth. Second, rather than focusing solely on one methodological approach, this session brings together a variety of methods (e.g., experiments and empirical analysis of field data) to address the questions at hand. Third, each paper attacks these phenomena from a unique perspective, delivering diverse but complementary insights. The co-chairs (Ezgi Akpinar and Jonah Berger) will integrate the talks and open discussion about potential directions for future work in this exciting area.

Given how fundamental word-of-mouth and social transmission is to our daily lives, and recent interest in social media, we expect this session will be of substantial interest to a host of contingencies. Not only should it appeal to researchers working on word of mouth and attitude change, but also to those who study advertising, social influence, consumer-generated content, persuasion and the effects of rating scales. This should produce a fruitful interdisciplinary discussion that will encourage more research in the area.

When Controversy Begets Conversation

EXTENDED ABSTRACT
Conversation topics vary on how controversial they are, or whether they are marked by opposing views (Merriam-Webster 2003). Some advertisements, topics and brands are relatively non-controversial while others are more divisive. GoDaddy.com, for example, often makes risqué Super Bowl ads, while Chevy’s ads tend to be less contentious. Topics like the weather are less controversial than gay marriage.

But how does controversy impact conversation? The lay belief among marketers and consumers is that controversy increases conversation (Steel 2011). But is that actually the case?

Further, people engage in all types of conversations. They post anonymously on online discussion boards or have conversations on Facebook; they chat face-to-face with friends or converse with strangers. Might these different contextual factors impact the relationship between controversy and conversation, and if so, how?

We theorize that the impact of controversy on conversation will depend on two countervailing forces. On the one hand, controversy can boost conversation because it is interesting. Differences in opinion can liven things up, and people talk about interesting things to entertain others (Heath et al. 2001) or fulfill self-enhancement goals (e.g., signal that the speaker is interesting, Berger and Schwartz 2011). On the other hand, controversy can decrease conversation by making people uncomfortable. People want to be socially accepted, and interpersonal conflicts, such as those generated by con-
troversial topics, can jeopardize acceptance. How controversy shape conversation, then, will depend on the strength of the two underlying processes.

Contextual factors such as identity disclosure and conversation partner (e.g., friend vs. stranger) should moderate the relationship between controversy and conversation by acting upon these underlying processes. The effect of interestingness is likely to remain static across situations since interesting topics are likely to remain interesting across settings. However, if discomfort reflects concerns about social acceptance, then it should be weaker in contexts when social acceptance concerns are less salient (e.g., when there is no identity disclosure) or less threatened by discussion of controversial issues (e.g., when talking to friends).

Three studies examine (1) how controversy affects conversation, (2) the underlying drivers, and (3) how this relationship varies in different conversation contexts.

Field Study 1: Our first study uses articles from a real news website to look at how the amount of controversy an article evokes impacts the number of comments it receives. We collected over two hundred articles as well as the number of comments each received (~5,000 comments overall. Two independent coders then rated each article on controversy (1 = not at all controversial, 7 = very controversial). Regressing the number of comments articles received on both the linear and squared controversy ratings shows an inverted-U relationship between controversy and conversation. Low levels of controversy increase conversation. But past a moderate controversy, additional controversy fails to increase (and even decreases) conversation. This is particularly noteworthy given the moderate level at which the effects of controversy start to reverse.

Study 2: In our second study, we experimentally manipulate controversy to observe directly examine its causal impact on conversation. We also test the underlying processes of interest and comfort and also explore how identity disclosure moderates the relationship between controversy and conversation by affecting comfort.

Participants were randomly assigned to condition in a 2 Disclosure (No disclosure vs. disclosure) x 3 Controversy (low vs. moderate vs. high) design. They listed a broad topic in current events (e.g., “taxes”) and then listed three subtopics, one that is high, middle, and low in controversy (e.g., “corporate bailout”, “taxing the rich”, and “sales tax”). Participants were asked to imagine having an online conversation with a stranger in which everyone was either using aliases (No Disclosure condition) or real names (Disclosure condition). Then, they were randomly given one of the subtopics they listed and asked to rate how likely they would be to talk about the subtopic in that conversation. They also rated how interesting they found the subtopic to be and how (un)comfortable they would feel talking about the subtopic. Results show that disclosure moderate the relationship between controversy and conversation. In the no disclosure condition, controversy has an inverted-U relationship with conversation where moderate level of controversy increases conversation ($M_{low} = 4.24$ vs. $M_{moderate} = 5.61, p = .02$). However, additional controversy decreases conversation ($M_{moderate} = 5.61$ vs. $M_{high} = 4.04, p < .01$). In the disclosure condition, controversy doesn’t increase conversation (low same as middle same as high, $p > .23$) and appears to decrease conversation ($M_{high} = 3.58$ vs. $M_{low} = 4.87, p < .05$). Mediation analyses show that when people do not disclose identity, conversation is driven primarily by interestingness. When identity is disclosed, however, conversation is driven by both interestingness and comfort.

Study 3: In our third study, we further test our framework and examine how conversation partner (friend vs. stranger) affects people’s propensity to talk about controversies.

We again asked participants to list a general topic and three subtopics that varied in level of controversy. Then we manipulated the audience by asking them to imagine having a face-to-face conversation with either a friend or stranger. Then, they rated each of the three subtopics they listed on likelihood of talking (main DV), interestingness, and discomfort.

Results reveal that conversation partner moderates the relationship between controversy and conversation. When talking to friends, controversy increases conversation. A move from low to moderate levels of controversy increases conversation ($M_{low} = 3.65$ vs. $M_{moderate} = 4.81, p = .02$). Beyond that point, however, additional controversy fails to further increase conversation ($M_{moderate} = 4.81$ vs. $M_{high} = 5.00, p = .68$). When talking to strangers, conversation is flat across the three controversy conditions ($M_{low} = 3.78$ vs. $M_{moderate} = 3.52$, and $M_{high} = 3.48$, pairwise comparisons all insignificant). Mediation analyses reveal that when talking to friends, conversation is driven primarily by interestingness; however, when talking to strangers, the positive effect of interesting is canceled out by discomfort.

Taken together, result show that (1) controversy doesn’t necessarily increase conversation, (2) interesting and comfort drive the effects, and that (3) situational factors moderate the outcome by affecting the underlying processes.

Rating with Confidence: Rating Certainty and Word-of-Mouth Behavior

EXTENDED ABSTRACT

Online product ratings represent a common outlet for consumers to express their post-consumption (dis)satisfaction about product performance. Consumers, as product reviewers, evaluate products based on their consumption experience via ratings and reviews on retailer websites such as Amazon.com and Yelp.com or manufacturer websites such as Apple.com and dominos.com. It has been consistently shown that product ratings, as a part of WOM, have a significant impact on product sales (Chevalier and Mayzlin 2006; Chintagunta et al 2011; Zhu and Zhang 2010). While most research has focused on the impact of product reviews on those reading them, little attention has been paid to their impact on those providing them. We address this question here by investigating a product reviewer’s subsequent WOM behavior after writing a review or rating a product.

In a pilot study, we found that different rating scales led to different levels of WOM. Respondents asked to rate an experience on a 5-pt scale report higher WOM intention compared with those rating on a 100-pt scale. To our knowledge, no existing theory explains this effect. We hypothesize that the result is driven by a general psychological construct which we label “rating certainty.” According to this theory, a product reviewer’s “true” utility must be mapped to the platform’s chosen rating scale. The fidelity of the mapping process in communicating the rater’s true utility varies across contexts. This fidelity – the extent to which one’s chosen rating captures accurately her underlying utility – is the essence of rating certainty. Finally, when a consumer has higher rating certainty, we argue that she is more likely to transmit WOM.

While our theory is based on rating certainty, it is also entirely plausible that attitude certainty may play a similar role. That is, when people are more certain about their attitude, the likelihood of performing attitude-associated behavior, such as engaging in WOM, will increase (Bassili 1996; Tormala and Petty 2002). However, we stress that rating certainty differs from attitude certainty in that the former captures one’s certainty toward the rating while the latter captures certainty toward the attitude target. To demonstrate that these two constructs are distinct, we present the results of a discriminant
validity analysis. Moreover, we control for attitude certainty in all analyses.

Next, we examined the impact of rating scale on WOM and the mediating role of rating certainty. One hundred and twenty two adults from Amazon Mechanical Turk participated in a video rating study. The study design was a 2 (rating scale: 2pt or 5pt) x 2 (certainty measure order: rating certainty first (RC) or attitude certainty first (AC)). Participants first watched a 2-minute video and then rated the video on either a 2-point or 5-point rating scale anchored from awful to excellent. In the RC condition, we measured rating certainty via four 9-point items which had been chosen based on a previous scale development study. These included: “How sure are you that the rating score you assigned is precise?” “How definite is your rating score of the video?”, “The rating score I assigned was clear”, and “The rating score I assigned was precise.” We combined these four items to generate a rating certainty index (Cronbach’s $\alpha = .95$). Next, we measured participants’ attitude toward the video and their certainty with respect to the attitude measure (Barden and Petty 2008). In the AC condition, we first measured participants’ attitude certainty and then their rating certainty. We counterbalanced the question order of RC and AC to control for any order effect. Since the results show no order effect, we combined these two conditions in our analysis. Finally, we assessed WOM likelihood on a 7-point scale.

We found the predicted rating-scale effect: participants were more likely to transmit WOM in the 5-pt condition than in the 2 pt condition ($M_{2pt}=2.2$, $M_{5pt}=2.8$, $p < .04$). Moreover, rating certainty was higher in the 5 pt condition than in the 2 pt condition ($M_{2pt}=6.4$, $M_{5pt}=7.6$, $p < .04$). Notably, we find no effect of rating scale on attitude certainty: participants in the two scale conditions are equally certain about their attitude toward the video ($M_{2pt}=8.0$, $M_{5pt}=7.9$, $p = .69$). This suggests that rating scale is not a source of attitude certainty and, thus, that attitude certainty does not explain the effect of scale on WOM. Finally, we perform a mediation test which supports the hypothesis that the effect of rating scale on WOM can be explained by rating certainty. The five-point scale induces higher levels of rating certainty leading to higher levels of WOM. In sum, we develop and validate a measure of rating certainty and demonstrate its discriminant validity with respect to attitude certainty. Finally, we demonstrate that rating certainty can be affected by the chosen rating scale and that, in turn, it may be an antecedent to WOM intention.

**Valuable Virality: The Effect of Advertising Appeals and Brand Integralness**

**EXTENDED ABSTRACT**

What influences whether ads are not only highly shared (i.e., viral) but also increase product evaluation and sales of the brands that create them (i.e., valuable)?

Recent research has shown that word-of-mouth can boost diffusion and sales (e.g., Godes and Mayzlin 2004, 2009; Goldenberg et al. 2009). Consequently, many organizations and brands have invested lots of resources in trying to craft content that consumers will talk about and share. But there are two difficulties with this approach. On the one hand, most people don’t like to share things that look like ads, so companies try to create outrageous or funny content (e.g., Subservient Chicken or Old Spice). On the other hand, while it’s great to make viral content, the content will only help the brand if it changes downstream consumer behavior, such as increasing brand evaluation or purchase. How can companies craft content that is not only viral, but also valuable?

We suggest that valuable virality depends on two key factors: ad appeal (hard vs. soft sell) and brand integralness (whether brand is integral to the advertising plot or not). While hard sell appeals focus on product features using a sales orientation, soft sell appeals convey a story through indirect mechanisms (i.e., showing beautiful scenes or evoking affective reactions such as humor and surprise, Okazaki, Mueller and Taylor 2010). As noted above, people don’t like to share things that look like direct persuasion attempts and soft sell appeals might be shared more because they provide more interesting content.

That said, soft sell appeals are not very useful if they don’t boost product evaluation or choice. Further, consumers realize advertisers are trying to persuade them and may dislike ads that they know are trying to trick them (Fristad and Wright 1994). Consequently, we suggest that to generate valuable virality, the ad must not only be interesting content but the brand must be integral to the narrative - woven in so deeply that it’s directly connected to the underlying story.

We test this possibility in two studies. In Study 1, we exposed participants to either (1) a soft sell-brand integral ad; (b) a soft sell-brand not integral ad or (c) a hard sell ad (brand integral by default). Then they reported their evaluation of the ad and likelihood of sharing it, as well as their evaluation of the brand, and likelihood of purchase.

Results show how appeal type and brand integralness influence different dependent variables differently (Figure 1). As predicted, soft sell appeals boosted sharing and ad evaluation. Compared to the hard sell appeal, participants said they would be more likely to share either of the soft sell appeals and evaluated them more favorably.

The effects differed, however, for brand evaluation and purchase. As predicted, while hard sell appeals boosted brand evaluation and purchase above soft sell appeals where the brand was not integral, soft sell appeals where the brand was integral had the most positive effects.

In Study 2, we rule out alternative explanations and provide deeper insight into the mechanism behind the effects. While the results of Study 1 are supportive, one might wonder whether it was something about the ads themselves, rather than the integralness of the brand, that drove the effects. To rule out this possibility, we created two versions of each ad. The ad content itself was the same, but we varied what type of product the ad was for to manipulate integralness. So across conditions, the same ad was either integral or not depending on whether the product it was for fit with the ad content. We also tested why non integral soft sell appeals reduce brand evaluation and choice by measuring how participants feel about persuasive attempts.

A 2 (product category) x 3 (ad type) ANOVA was conducted on the various dependent measures. As expected, and extending the results of Study 1, the manipulations had different impacts on different
dependent variables. While soft sells boosted sharing and ad evaluation compared to hard sells, only integral soft sell appeal boosted brand evaluation and purchase. Further, mediation analyses show that these effects are driven by the extent to which consumers find the ad persuasive attempts as acceptable and not very manipulative. When the brand is integral to the ad, persuasive attempts seemed more acceptable and less manipulative and therefore those brands receive more favorable evaluations and purchase intentions.

Taken together, these results show how content (e.g., ads) can be crafted so that it is both highly shared and beneficial to the brand or organization that made it. While soft sell appeals are more likely to be shared, the brand must be integral to the ad to receive the benefits of brand evaluation and purchase. This is driven by consumers’ perceptions of advertiser’s persuasive attempts. Overall the findings suggest how to generate virality that is also valuable to the brand.

**Does Paying For Online Product Reviews Pay Off? The Effects of Monetary Incentives on Content Creators and Consumers**

EXTENDED ABSTRACT

This research considers whether offering incentives to encourage consumers to write online product reviews is beneficial for firms. We propose that paying customers will result in more helpful reviews because offering content creators an explicit incentive may encourage them to take a more professional approach to the review-writing task. Paradoxically if content consumers know that the content creator was paid they may form less favorable expectations of product quality based on the review. Specifically, when incentive provision is disclosed, review consumers doubt the quality of the reviewed product. The doubt induced by payment disclosure is hypothesized to distort product judgments resulting in a shift of evaluations toward a neutral or indifferent attitude when the review is valenced. Paying content creators may therefore ironically decrease review-based product evaluations even though the review itself is more helpful.

We test our predictions in four experiments. In all experiments, participants were members of a large online panel in the U.S. In the first part of Study 1, participants were content creators who played an online game and then wrote a review about it. Participants were randomly assigned to either receive no incentive or a $1 incentive for writing their review. We find that paying an incentive had no impact on perceived effort, or difficulty related to writing the review.

In part 2 of Study 1, subjects were content consumers. Each was randomly assigned one of the reviews generated in part 1 and asked questions about the helpfulness, positivity, and objectivity of the review. Using random effects regression, we find a significant positive effect of incentive provision on helpfulness ($b = .18, t = 2.16, p = .03$), but not positivity ($p = .78$) or objectivity ($p = .73$). We also show that increasing review helpfulness (through incentive provision) may lead to content consumers being more interested in product trial.

Using a similar procedure, Study 2 examined how content consumers’ product evaluations are affected by incentive disclosure. Participants were randomly assigned to one of four conditions in a 2 (no disclosure vs. disclosure) x 2 (no incentive vs. $1 incentive) between-subjects design. In the disclosure (no-disclosure) conditions, participants were (were not) given information about whether the content creator was paid to write the review. The incentive manipulation was the same as that used in Study 1.

We tested how incentives and disclosing them affected content consumers’ product evaluations (1 to 5 stars) and their willingness to pay for the game as an app download ($0 to $0.99). We find that product evaluations are lower when content creators are incentivized and content consumers are aware of this. In addition, willingness to pay dropped from 81 cents to 25 cents when the reviewer was incentivized and incentive provision was disclosed.

In Study 3 we replicate the Study 2 findings in a different product category (music) and rule out the possibility that consumer-generated reviews are only influential in the absence of more reliable information.

Participants were randomly assigned to one of six conditions in a 3 (silent vs. not paid vs. paid) x 2 (no additional information vs. additional information) between-subjects design. The first factor manipulated the specificity of the incentive disclosure (silent: no information regarding incentives provision; not paid: “the person who wrote this review was not paid”; paid: “the person who wrote this review was paid”). This new manipulation allowed us to check whether our results hold when disclosure is ambiguous.

The second factor manipulated the inclusion of information from a more realistic source in the participant’s information set (no additional information: basic description of the band whose music video was the subject of the review; additional information: basic description of the band and a three page *Rolling Stone* magazine article that positively described the band’s history and was created for this study).

We replicate our prior findings by showing that disclosure has a significant main effect on product evaluation ($F(2, 212) = 3.94, p = .02$). The main effect of additional information and the two-way interaction were not significant ($p > .6$). The mean product evaluation was lower when the participants thought the content creator had been paid ($M_{paid} = 3.38$) than when nothing was disclosed ($M_{silent} = 3.69$) or when receiving no payment was disclosed ($M_{not-paid} = 3.62$).

The evaluation-lowering effect therefore holds despite changes to how incentive disclosure is conveyed and whether or not additional product information is available to content consumers.

Study 4 shows that the effect of incentive disclosure on content consumers’ product evaluations is mediated by an increase in doubt in product quality and that this process only occurs when reviews are positive. Additionally, this study examines whether changes in review-based product evaluations caused by incentive disclosure carry over to affect post-experience evaluations.

The product used in this study was the game from the first two studies. Participants were randomly assigned to one of six conditions in a 3 (paid $0 vs. paid $0.75 vs. paid $1.25) x 2 (positive review vs. negative review) between-subjects design. The incentive-disclosure manipulation was similar to that used in the previous studies except that we used two non-zero conditions ($0.75$ and $1.25$) to show that our previous results were not specifically due to disclosure of a $1 incentive. The valence manipulation involved randomly presenting participants with either a positive or a negative review.

Using the procedure for testing moderated mediation detailed in Preacher, Rucker, and Hayes (2007), we find that the effect of incentive disclosure on pre-experience product evaluation was mediated by doubt when the review was positive. Furthermore, we find that the effect of incentive disclosure on product evaluations through doubt carries over to affect product evaluations when additional experiential information is available.

This research shows that information related to whether a content creator is incentivized can make a difference not only to how the generated content is interpreted by a consumer but also in their assessment of the focal product in both the short-term (at the stage of forming expectations of product quality) and long-term (lowering product evaluations even after having first-hand product experience).