

BONUS, DISCLOSURE, AND CHOICE: WHAT MOTIVATES THE CREATION OF HIGH-QUALITY PAID REVIEWS?

Completed Research Paper

Jing Wang

Stern School of Business, NYU
44 W. 4th Street, New York, NY 10012
jwang5@stern.nyu.edu

Anindya Ghose

Stern School of Business, NYU
44 W. 4th Street, New York, NY 10012
aghose@stern.nyu.edu

Panagiotis G. Ipeirotis

Stern School of Business, NYU
44 W. 4th Street, New York, NY 10012
panos@stern.nyu.edu

Abstract

The emergence of online crowdsourcing sites has opened up new channels for third parties and companies to solicit paid reviews from people. In this paper, we investigate 1) how the introduction of monetary payments affects review quality, and 2) the impact of bonus rewards, sponsorship disclosure, and choice freedom on the quality of paid reviews. We conduct a 2×2×2 between-subjects experiment on Amazon Mechanical Turk. Our results indicate that there are no significant quality differences between paid and unpaid reviews. The quality of paid reviews improves by both the presence of additional performance-contingent rewards and the requirement to add disclosure text about material connections, and deteriorates by the restrictions imposed on the product set to be reviewed. These results have implications for websites and companies who are seeking legitimate reviews for their online products from paid workers.

Keywords: User-generated content, Completion-contingent rewards, Performance-contingent rewards, Sponsorship disclosure, Choice freedom, Review quality, Between-subjects experiment, Econometric analyses

Introduction

Online reviews are becoming an increasingly prevalent way for consumers to publish and share their personal experiences with products and services. Potential buyers rely on such information to infer the underlying quality of products and make their purchase decisions. Product manufacturers are able to learn consumer preferences and adapt their design or marketing strategies accordingly.

Product reviews in many websites are contributed by volunteers. Although the total number of product reviews grows rapidly, it does not necessarily mean that each single product receives a large number of reviews. Jindal and Liu (2008) showed that, in the data they extracted from Amazon.com, there are a large number of products which get very few reviews (e.g. 50% of products have only 1 review). Duan et al. (2008) found that the volume of online reviews has a positive impact on box office revenues. Ghose and Ipeirotis (2011) also demonstrated that a larger number of reviews are associated with higher product sales. So, the lack of reviews for a product can be interpreted by the buyers as a bad signal for the quality of the product. At the same time, if buyers refrain from buying a product due to lack of the reviews, this also leads to lack of buyers that can write these reviews. To address the “cold-start” problem, some websites and companies now begin to provide monetary incentives to solicit reviews from workers in a variety of online crowdsourcing sites (e.g. eLance, Guru, Amazon Mechanical Turk, oDesk, and others).

When discussing paid reviews, the first words that come into our mind are “fake,” “biased,” “useless” and so on. This is indeed true when reviewers are posting either positive or negative reviews which differ from their real opinions. However, we believe that it would be inappropriate to consider all paid reviews untruthful. While on the Internet we are used to seeing reviews generated on a volunteer basis, paid reviews have been around for a long time. Many famous magazines (e.g. Consumer Reports, Wine Spectator) provide reviews all written by paid employees and critics. Even the well-known Saga books, which pioneered the concept of crowdsourcing reviews in an online world, had paid reviewers for aggregating the individual volunteer reviews into the short, snippet-based reviews for the restaurants. Also, there is a growing interest in blog advertising, which refers to the paid sponsorship of bloggers to mention, review, promote, or sell products in blog writing (Zhu and Tan 2007). Furthermore, it is well established that sponsored (paid) links that appear next to organic results are of high quality, which are also the powering force behind the development of many businesses on the web.

For this current study, we focus on the solicitation of paid reviews for which there are no polarity directions imposed on the expressed opinions. The solicitation of this type of paid reviews has the potential to increase significantly the number of reviews and foster the information sharing between consumers and manufacturers, especially for those rarely reviewed products. However, there is the legitimate concern that the quality of the reviews might deteriorate with the introduction of monetary incentives. On October 5, 2009 the US Federal Trade Commission (FTC) issued new guidelines that require advertisers to “disclose material connections (cash, free gifts, coupons, etc.)” for product or service endorsements. It is possible that this requirement to disclose material connections would affect the inherent motivation of reviewers and also the quality of reviews. Another feature for paid reviews is that the products to be reviewed are usually prescribed by the sponsors. Therefore, reviewers don't have the absolute freedom to choose products that they'd like to review, as what they do for the voluntary reviews. The effect of the restriction for product choice on the quality of the reviews is also an interesting question to explore.

The first question to explore is whether the provision of explicit monetary payments (completion-contingent rewards) would affect the quality of reviews. The second question, which is the main topic of this paper, is to investigate, in the presence of completion-contingent rewards, how the change in bonus payments (performance-contingent rewards), sponsorship disclosure, and choice freedom will influence the motivation of the reviewers as well as the review quality. We design a between-subject experiment to test how people behave under different treatments.

Our main contributions can be summarized as follows. First, subjects in our experiment behave similarly as they do when voluntarily participating in real life. Second, we find that, under the same condition, there are no significant quality differences between reviews solicited on a voluntary basis and those written for money. Our empirical results demonstrate that when completion-contingent rewards are provided: 1) extra performance-contingent rewards give subjects more incentive to generate high-quality reviews; 2) restricting the products that one can review would potentially undermine the review quality;

3) the requirement to disclose material connections can induce reviews that are considered more helpful when disclosure text is excluded from the review; however, if disclosure text is present, we find different reactions depending on the evaluators used: workers on Amazon Mechanical Turk (AMT) are somewhat indifferent to disclosure text; however, users on Amazon.com in general hold a negative view. This disparity illustrates that the acceptance of paid reviews as a legitimate medium for transferring information from reviewers to consumers is not just a matter of review quality but needs to also change the existing social norms that largely assume reviews are contributed in an unpaid and voluntary basis.

Theories and Hypotheses

Monetary Rewards and Review Quality

In real life, there are many kinds of monetary rewards that have been used by employers to attract, motivate and retain employees and achieve organizational goals. In this paper, we are interested in two types of rewards: *completion-contingent rewards*, which are rewards offered for completion of a task; and *performance-contingent rewards*, which require performing the task well, matching a standard of excellence, or surpassing a specified criterion. In crowdsourcing settings, reviewers usually get paid on a fixed rate per review; and they are rewarded with extra bonus only if the reviews are perceived as very helpful. Therefore, we ask two questions here: 1) How does the provision of completion-contingent rewards affect the quality of reviews? 2) When completion-contingent rewards are provided, do extra performance-contingent rewards lead to change in review quality?

Individuals are motivated to take certain actions or exhibit certain behaviors for a variety of reasons. *Intrinsic motivation* refers to motivation that is driven by an interest or enjoyment in the task itself, which comes from inside an individual, whereas *extrinsic motivation* refers to the performance of an activity in order to attain an outcome (e.g. money) which comes from outside an individual. When the target task is to write reviews, a reviewer is intrinsically motivated if she or he enjoys the feeling of sharing personal experiences and feelings with others, and externally motivated if she or he receives monetary compensation for that. The thousands of volunteers who have generated a vast amount of content over the Internet are mostly driven by their inherent interest, since there are no monetary rewards involved. The monetary compensation for writing reviews might be able to motivate individuals from another dimension-extrinsic motivation.

The effect of monetary rewards on performance has been debated for many years. According to Cognitive Evaluation Theory (CET; Deci 1975; Deci and Ryan 1980; Deci and Ryan 1985), monetary rewards have both a controlling aspect and an informational aspect. If rewards are perceived as *controlling*, they will crowd-out intrinsic motivation by inducing a shift in perceived locus of causality from internal (self-determined) to external (other-determined). Deci (1971) investigated the effects of external rewards on intrinsic motivation using two laboratory experiments and one field experiment. He found that intrinsic motivation tends to decrease in the presence of monetary rewards. Bock et al. (2005) showed that anticipated external rewards have a negative effect on individuals' knowledge-sharing attitudes. On the contrary, if monetary rewards are perceived as *informational*, they will crowd-in intrinsic motivation by serving the function of conveying relevant information about one's intrinsic competence in the rewarded activity. Cameron et al. (2005) found that achievement-based rewards during learning or testing increased participants' intrinsic motivation.

The standard economic principal-agent theory (e.g. Alchian and Demsetz 1972; Fama Jensen 1983) argues that individuals are motivated to work harder when monetary rewards rise. As a consequence, the common strategies employed by managers to motivate their employees are money, benefits, and other different forms of compensation (e.g. completion and performance-based rewards). They believe that the linkage between performance and rewards can encourage employees to exert more efforts which will, in turn, increase the overall performance. Burgess (2005) found that employees will spend more hours sharing knowledge if they perceive greater organizational rewards for doing so.

We now turn back to our first question about the effect of completion-contingent rewards on review quality. Completion-contingent rewards have a strong controlling aspect since people have to complete the reviews, but a very weak informational aspect since they contain little information about the person's competency in writing reviews; therefore, they are predicted to crowd-out intrinsic motivation. But these

rewards typically have a positive disciplining effect on extrinsic motivation (Frey 1994). The final outcome is determined by the relative salience of these two effects. As Alexy and Leitner (2007) have shown that total motivation of developers will increase when they are offered a completion-contingent reward for working on an open source software project, we hypothesize the following:

Hypothesis 1: *The provision of completion-contingent rewards will result in an increase in the quality of reviews.*

The second question we have is whether extra bonus payments (performance-contingent rewards) can improve the quality of paid reviews. Performance-contingent rewards communicate control in the sense that people have to meet a standard to maximize rewards; however, they can also signify the competence of people since the rewards are contingent on performance. Since the perceived locus of causality has probably shifted to external by the completion-contingent rewards, people will not feel much more controlled with the addition of performance-contingent rewards. Hence, we expect that, for performance-contingent rewards provided here, the informational aspect is more salient than the controlling aspect, which will result in a crowding-in effect. Moreover, performance-contingent rewards also have a positive disciplining effect. Therefore, we propose the following hypothesis:

Hypothesis 2: *When completion-contingent rewards are given, the provision of performance-contingent rewards will result in an increase in the quality of reviews.*

Sponsorship Disclosure and Review Quality

On October 5, 2009 the US FTC published the *Guides Concerning the Use of Endorsements and Testimonials in Advertising*, which took effect on December 1, 2009. One principle is that “*Advertisers are subject to liability for false or unsubstantiated statements made through endorsements, or for failing to disclose material connections between themselves and their endorsers*”. While most of other work is focused on the fairness and enforceability of such regulations, we want to examine how this requirement would affect the behaviors of reviewers.

Online user-generated content (UGC) is reshaping the way people share their personal knowledge or experience. Tons of information is created everyday over the Internet by thousands of volunteers. There are some papers which attempt to understand why people are willing to contribute their knowledge and effort for free. Self-image is commonly identified as a motive for the production of UGC. Daugherty et al. (2008) found that the *ego-defensive function*, which represents motivations designed to protect the ego from negative features of the self, contributes significantly to the attitude formation toward the creation of UGC. The creation of UGC helps consumers minimize their self-doubts, and possibly reduce guilty feelings about not contributing. Yang and Lai (2011) showed that internal self-concept-based motivation significantly influences the knowledge-sharing intention of Wikipedia contributors. Park et al. (2011) demonstrated that ego-involvement is highly associated with both the attitude toward uploading video content online and the intention to upload.

Nowadays, since most online reviews are believed to be contributed voluntarily (which is of course not true given the abundance of ads for paid reviews), paid reviews are generally not acceptable or at least unattractive. The disclosure of material connections (cash, free gifts, coupons, etc.) makes the fact of being paid obvious and salient, so it would unquestionably affect the credibility of the reviews, and undermine the image of the reviewers.

However, a well-written and honest review which truthfully reflects the opinion of the reviewer can reduce the guiltiness he has for being paid. It is reasonable to hypothesize that the reviewers who are motivated by self-image would attempt to protect their self-esteem by investing more time and efforts on the review writing process.

Hypothesis 3: *When completion-contingent rewards are given, the requirement to disclose material connections will result in an increase in the quality of reviews.*

Choice Freedom and Review Quality

Self-determination theory (SDT), developed by Deci and Ryan (Deci and Ryan 1985; Ryan and Deci 2000), asserts that the basic psychological needs for autonomy, competence, and relatedness are primary

factors that encourage intrinsic motivation. The need for *autonomy* is defined as an inherent desire to act with a sense of choice and volition, that is, to feel psychologically free from control, pressure and obligations. The need for *competence* is fulfilled by the feeling that one can reliably produce desired outcomes and achieve particular goals. The need for *relatedness* represents the need to feel securely connected with, and experience caring, recognition, and respect from others (Baumeister and Leary 1995).

SDT states that the provision of choice is an effective way to support the perception of autonomy. Therefore, the freedom to choose is expected to result in enhanced intrinsic motivation, higher levels of engagement, and better task performance. Kail (1975) conducted an experiment which demonstrates that subjects who freely choose a task persist longer at that task than those who are forced to do it. Monty and Rosenberger (1973) found that giving subjects an opportunity to exercise choice prior to the learning of a paired-associate task facilitates learning of that task. Patall et al. (2010) showed that subjects who receive a choice of homework report higher intrinsic motivation, achieve better completion rates, compared with subjects who do not have a choice.

When people are paid to write reviews, they are likely to lose the freedom of choice and feel less autonomous as their endorsers might restrict the set of products that they can choose from. This would, as previous literature implies, reduce the intrinsic motivation and overall performance of reviewers. Therefore, we posit that:

Hypothesis 4: *When completion-contingent rewards are given, the restriction for product choice will result in a decrease in the quality of reviews.*

Experimental Design

Since it is widely believed that carefully designed experiments can give researchers more control of the situation and make causal claims more convincing, we adopted an experimental methodology to test all these research hypotheses. We used a 2 (no bonus vs. bonus) × 2 (no disclosure vs. disclosure) × 2 (open vs. closed product selection) between-subjects experimental design and randomly assigned subjects to the eight treatments in our experiment.

Subjects

All the participants were recruited from AMT, which increasingly becomes a powerful tool for conducting online behavioral experiments (Mason and Suri 2011). We asked recruited workers to post their product reviews in Amazon.com—the world's largest online retailer which has reviews for all kinds of consumer durable products like books, CDs, electronics, home appliances, etc. We, as the requester, posted a HIT (Human Intelligence Tasks) on AMT titled “Answer a Survey on Online Review Writing and Post a Product Review on Amazon.com”. Only the workers who passed the basic qualification requirement ($\geq 95\%$ approval rate) were allowed to work on the task. Each worker was paid \$1 for successfully completing the task. We performed two checks in order to better understand the background of the participants.

Background Check 1: Survey on Reviewer Habits

Every participant who accepted the HIT was required to answer a short survey about online review writing, the details of which are displayed in Table 1. The purpose is to see how many participants have written reviews in the past, either voluntarily or after being paid. We would also be able to know if the reviewers are experienced reviewers or not.

Background Check 2: Review History

One key criticism for experiments in particular is that people might behave differently under experimental conditions than they would in real world. In our context, we are concerned about the inconsistency of reviewers' behaviors in our experiment and real-life scenarios. In other words, do reviewers who write high-quality reviews in the past tend to perform better in our experiment?

One significant advantage of using Mechanical Turk for our experiment is that the IDs of many Mechanical Turk workers are the same as their reviewer IDs on Amazon.com. Therefore, for each participant in our experiment, we can identify the reviews they have written and posted on Amazon.com

in the past. So, we extracted all the past reviews written by participants in our experiment from Amazon.com. This allows us to test if the quality of past reviews written by a reviewer can, in some extent, reflects the quality of the review written by the same reviewer in our experiment.

Table 1. Survey on Reviewer Habits
1. Have you written a review on Amazon.com before?
2. Have you written a review on any other website (not Amazon.com)? If yes, please provide the name of the website.
3. If you answered (Yes) to either 1 or 2, please indicate the total number of reviews you've written across all websites in the past.
4. Have you been paid to write a review on Amazon.com before?
5. Have you been paid to write a review on any other website before?
6. If you answered (Yes) to either 4 or 5, please indicate the percentage of the reviews for which you have been or expect to be paid.

Posting a Product Review

We have a total of eight conditions: 2 (no bonus vs. bonus) \times 2 (no disclosure vs. disclosure) \times 2 (open vs. closed product selection).

- **No bonus vs. bonus:** in the no bonus condition, there is no extra reward; in the bonus condition, workers are rewarded \$0.25 for each additional helpful vote collected on Amazon.com.
- **No disclosure vs. disclosure:** in the no disclosure condition, no disclaimer is needed; in the disclosure condition, workers are required to add a disclaimer after the review text: *“The reviewer was compensated for posting this review. However, the opinion stated in the review is that of the reviewer and the reviewer alone. Further, the reviewer independently selected this product to review and has no affiliation with the product maker/distributor, Amazon or the review requester.”*
- **Open vs. closed product set:** in the open condition, workers have the freedom to review any product available on Amazon.com that they have not previously reviewed on the site; in the closed condition, workers are restricted to review one of the five products given by us, which are randomly chosen from those products for which we have collected reviews in the open condition.

In AMT, each account is associated with a unique worker ID. To ensure random assignment of subjects to different conditions, each worker ID was allowed to participate in the study at most once, and was randomly assigned to one treatment. To maximally reduce the possible noise introduced by product heterogeneity between open and closed condition, we decide to use the products collected from the open condition to construct the product pool for the closed condition. Since we cannot get the product set until the open condition is completed, we split our experiment into two stages: open product stage and closed product stage. Within each stage, when a worker accepted our HIT, we randomly assigned her or him to one of the four conditions: 2 (no bonus vs. bonus) \times 2 (no disclosure vs. disclosure).

In our experiment, we do not require workers to have previous experience using the chosen products. But we collected information about the product usage by asking the following yes or no question: “Do you have experience using the product you selected to review?” This allows us to assess the impact of product usage on review quality later.

Evaluating a Product Review

Amazon.com has a helpful vote mechanism which evaluates review helpfulness based on the proportion of helpful votes. However, this mechanism is not always reliable, since it takes time to accumulate a reasonable number of ratings (Ghose and Ipeirotis 2007). Zhang and Varadarajan (2006) suggest that at least 10 votes per review are required in order to ensure the robustness of the ratings.

Since the reviews in our experiment are relatively recent, they often show up in the later pages in Amazon's ranking systems and end up with very few votes. Therefore, for each review that we collected in the 8 conditions, we asked 10 different workers in AMT the same yes or no question: "Was this review helpful to you?" We believe this approach can give us relatively reliable estimates for the review helpfulness.

In order to get an unbiased evaluation for the quality of reviews in the *Disclosure* condition, when we asked workers to evaluate these reviews, disclosure text was eliminated from the review text. We also asked workers to evaluate the past reviews of the participating workers that we collected in Background Check 2.

To make sure that the quality measure we get is reliable, we asked workers to evaluate some "gold" reviews for which we know the true quality: these are reviews posted on Amazon.com by other users, that have received a large number of votes on Amazon.com with a high fraction being either helpful or unhelpful votes. We manually collected 70 unhelpful reviews and 70 helpful reviews from Amazon.com as the "gold" reviews. In each HIT, we gave the worker 10 reviews to evaluate: eight reviews from our collected reviews, one review from the helpful review pool, and one review from the unhelpful review pool. Each worker was paid \$0.25 if she or he voted for all the 10 reviews and provided valid explanations for her or his decisions.

Preliminary Data

Descriptive Statistics

We collected a total of 552 reviews across the 8 conditions, as displayed in Table 2.

	No Bonus		Bonus	
	No Disclosure	Disclosure	No Disclosure	Disclosure
Open Product	81	86	101	73
Closed Product	83	39	54	35

Reliability Check

The helpfulness metric we used was based on the votes provided by AMT workers. Before proceeding to any further analysis, we need to make sure that this metric is consistent with the actual measure in Amazon.com. We checked the average scores for both helpful "gold" reviews and unhelpful "gold" reviews: the average fraction of helpful votes for helpful reviews is 0.847, which is a lot higher than 0.473—the average fraction of helpful votes for unhelpful reviews. This result gives us enough confidence in our evaluation metric for review quality.

Results for Background Check 1

Out of these 552 reviewers, 67% indicated that they had written online reviews in the past, and 13% had written reviews for monetary payment (4.7% had been paid to write reviews on Amazon.com). They also reported that they had been paid to write reviews on other online product sites such as iherb.com, ebay.com, Toys R Us, newegg.com, walmart.com, buy.com, etc.

Results for Background Check 2

For 193 out of 552 reviewers, we were able to collect their past review history. We run a simple linear regression based on

$$\text{Helpfulness} = \alpha + \beta \cdot \text{Helpfulness_history}$$

where *Helpfulness* is the quality score (measured by the fraction of helpful votes given by AMT workers) for each one of 193 reviews in our experiment, and *Helpfulness_history* is the average quality score (also evaluated by AMT workers) over all the past reviews of the same reviewer. The coefficient on *Helpfulness_history* is 0.259 (positive and significant at 0.3% level), which demonstrates that reviewers who wrote better reviews in the past are more likely to come up with high-quality reviews in our experiment, too.

Variables

Table 3 describes all the variables used and their explanations in our work. The dependent variable is *helpfulness*, measured by the fraction of people who found the review helpful. There has been some work (Mudambi and Schuff 2010; Ghose and Ipeirotis 2011) showing that review length is a good predictor for the helpfulness of reviews, therefore, we also report the length of reviews here, which is measured by the number of characters in the review.

Table 4 shows the descriptive statistics for all the variables.

Variable	Explanation
Helpfulness	the fraction of helpful votes (on a scale of 0-1)
Review Length	the total number of characters in the review
Product Usage	0 for no usage experience, 1 for some usage experience
Bonus	0 for no bonus, 1 for bonus
Disclosure	0 for no disclosure, 1 for disclosure
Closed	0 for open product selection, 1 for closed product selection

Variable	Obs.	Mean	Std. Dev.	Min	Max
Helpfulness	552	0.718	0.228	0.1	1
Review Length	552	490.27	356.19	73	3430
Product Usage	552	0.875	0.331	0	1
Bonus	552	0.476	0.500	0	1
Disclosure	552	0.422	0.494	0	1
Closed	552	0.382	0.486	0	1

Hypothesis Testing

In this section, we discuss the models used to test the four hypotheses that we brought up before, as well as the estimation results of these models.

Comparing Paid and Unpaid Reviews

As we stated before, one advantage of our experiment is the ability to track the reviews previously written by the reviewers (see “Background Check 2”). We also have an indication (based on the separate “Background Check 1” survey) on whether these reviews were written for payment or not. To see how completion-contingent rewards affect the characteristics of reviews, we created two different sets for paid and unpaid reviews, respectively. To eliminate the effects introduced by bonus, disclosure and closed selection, we only considered the condition for which the workers are free to choose whatever products they would like to review with no extra bonus and no requirement to add the “disclaimer”, which gave us a

total of 81 reviewers. Finally, we were able to collect past reviews for 28 reviewers (the rest of workers either have not written reviews on Amazon.com before or use different accounts for Amazon.com and AMT). We excluded the 2 workers who reported that they had written paid reviews on Amazon.com before. At the end, we got 82 unpaid reviews, previously posted on Amazon.com, and 26 paid reviews generated as part of our experiment. For each of the unpaid reviews, we asked 10 different workers in AMT to evaluate the helpfulness, as what we did for paid reviews.

Model Specification

In order to test our Hypothesis 1, we use a linear specification for the helpfulness estimation. There are a set of possible variables which might affect the quality of reviews besides the completion-contingent rewards. First, as is shown before, the inherent skills and expertise of reviewers differ from one to another. There are some workers who consistently write helpful reviews, regardless of the conditions imposed. Also, the product category matters. Mudambi and Schuff (2010) showed that product type (e.g. experience goods or search goods) affects the perceived helpfulness of the review.

To control for the effects of reviewer expertise and product type, we use a two-way fixed effects model, which is specified as follows:

$$Helpfulness_{kpr} = \alpha + \beta \cdot Paid_{kpr} + \mu_k + \eta_p + \varepsilon_{kpr}$$

The unit of observation in this analysis is a review. The dependent variable, $Helpfulness_{kpr}$, is for a review r in product category k written by reviewer p . $Paid_{kpr}$ is a variable indicating if the review r is from a paid reviewer. μ_k is a product category level *fixed effect* that accounts for unobserved heterogeneity across product categories. η_p is a reviewer level *fixed effect* that accounts for unobserved heterogeneity across reviewers. ε_{kpr} is the error term.

Empirical Results

Table 5. OLS Regressions with Product-Category-Level and Reviewer-level Fixed Effects	
Variable	Coef. (Std. Err)
Paid	0.080 (0.058)
Number of Observations	108
R-square (with fixed effects)	0.627
The dependent variable is Helpfulness.	

As shown in Table 5, the sign on the coefficient of *Paid* is positive, but not significant (The p-value of coefficient on *Paid* is 0.173). Therefore, we cannot reject the null hypothesis that there are no differences between the paid and unpaid reviews, and Hypothesis 1 cannot be accepted.

We provide the following explanation for this result. Staw et al. (1980) showed that when there was a norm for no payment, the introduction of monetary rewards significantly decreased overall task satisfaction. In the context of online reviews, people generally don't expect any monetary payments (i.e. there is a norm for no payment). When completion-contingent rewards are provided, the negative crowding-out effect increases to a level which is comparable to the positive disciplining effect, therefore, the overall motivation does not increase.

Testing the Effects of Bonus, Disclosure and Closed on Paid Reviews

The effects of bonus, disclosure and closed are tested on all the paid reviews we collected. Since we ensure random assignment for each worker, there is no need to control for the reviewer effect any more.

Model Specification

In order to test our Hypotheses 2 to 4, we use a linear specification for the helpfulness estimation.

$$\text{Helpfulness}_{kr} = \alpha + \beta_1 \cdot \text{Bonus}_{kr} + \beta_2 \cdot \text{Disclosure}_{kr} + \beta_3 \cdot \text{Closed}_{kr} + \mu_k + \varepsilon_{kr}$$

The unit of observation is a review and μ_k is a product category level *fixed effect* that accounts for unobserved heterogeneity across product categories. The fixed effects are equivalent to including a dummy for each product category. We have a total of 38 product categories in our sample. The number of observations per group ranges from 1 to 78.

Empirical Results

The results for the fixed effects model are shown in Table 6 column 2. *Closed* has the highest marginal effect, followed by *Bonus*, and then *Disclosure*.

The positive and statistically significant sign on the coefficient of *Bonus* implies that the provision of *performance-contingent rewards* increases the quality of paid reviews, which confirms Hypothesis 2. The presence of bonus increases the review helpfulness by 0.046.

We find that the coefficient of *Disclosure* is also positive and significant, which is consistent with our Hypothesis 3. The requirement to add disclosure text after review increases the review helpfulness by 0.038. As we are going to discuss later, this increase holds *no matter if disclosure text accompanies the review or not*, indicating that reviewers write better reviews when they are told that the payment will be disclosed. This confirms the prediction made by functional theory: consumers invest more effort on review writing to protect their self-image since writing reviews for monetary rewards is generally considered unfavorable in the community. This is good if our objective is to create a truthful and honest environment for online UGC. Although the credibility of the reviews might be undermined by the presence of disclosure text, the increase in review quality might compensate this loss to some extent. This finding gives companies and websites more incentive to obey the guidelines set by FTC which is, in turn, beneficial to the trust building of the online reviewing communities.

Moreover, the coefficient of *Closed* is very negative and statistically significant. The restriction for product set decreases the review helpfulness by 0.060, which gives strong support to Hypothesis 4. This implies that when workers are restricted to choose one product from a closed product set, their sense of freedom and volition are likely to be reduced, and so is their intrinsic motivation. As we will show next, this conclusion holds even if we exclude those reviews for which there is no product usage experience.

Subsample Analysis: With Product Usage

One criticism for our results is that the negative effect of restricted product choice on review quality may be due to the lack of product usage. Reviewers may choose to review products that they have never used before, especially when they don't have any previous usage experience with all the given products. So the bad performance of reviewers in the *Closed* condition is not because that reviewers are faced with a restricted choice of products, but because that they have no direct experience with the products they are reviewing.

To eliminate the biases induced by the heterogeneity in product usage, we rerun our regression, using only data (a reduced set of 483 reviews) for which the reviewers have some previous experience with the products they select to review.

The new results are shown in Table 6 column 3. We do not see any change in the coefficient signs of the three explanatory variables. However, the marginal effect of *Closed* condition decreases from -0.060 to -0.051, which implies that the product usage indeed contributes to the high review quality in the open product condition. In addition, the significance level of the coefficient of *Disclosure* has increased from 10% level to 1% level. And the marginal effect of *Disclosure* is now higher than the marginal effect of *Bonus*.

Robustness Check Using Review Length

To test the robustness of our results, we use review length as a proxy for the helpfulness of the reviews.

Mudambi and Schuff (2010) have shown that review length has a positive effect on the helpfulness of the review. Longer reviews often include more product details and more descriptions of usage experiences, so they can help potential consumers reduce the quality uncertainty of the particular products and make better purchase decisions.

We use the same linear specification for the helpfulness estimation, with Helpfulness replaced by $\text{Log}(\text{ReviewLength})$. We use logarithmic transformation here since the length of reviews follows a positively skewed distribution.

$$\log(\text{ReviewLength}_{kr}) = \alpha + \beta_1 \cdot \text{Bonus}_{kr} + \beta_2 \cdot \text{Disclosure}_{kr} + \beta_3 \cdot \text{Closed}_{kr} + \mu_k + \varepsilon_{kr}$$

The regression results are reported in Table 6 column 4. The signs on all three independent variables stay the same, which implies that our results are very robust.

Variable	Coef. (Std. Err) ^I	Coef. (Std. Err) ^{II}	Coef. (Std. Err) ^{III}
Bonus	0.046 (0.020)**	0.042 (0.020)**	0.194 (0.051)***
Disclosure	0.038 (0.020)*	0.056 (0.021)***	0.162 (0.051)***
Closed	-0.060 (0.020)***	-0.051 (0.022)**	-0.265 (0.053)***
Number of Observations	552	483	552
R-square (with fixed effects)	0.110	0.125	0.167

*** significant at 1% level; **significant at 5% level, *significant at 10% level
 I: Full sample; the dependent variable is Helpfulness.
 II: Subsample (with product usage); the dependent variable is Helpfulness.
 III: Full sample; the dependent variable is $\text{Log}(\text{ReviewLength})$.

Robustness Check Using Fractional Logit Model

Since the dependent variable *Helpfulness* is a proportion which takes values from 0 to 1, OLS estimation might not be appropriate since predicted values may fall outside the unit interval. Therefore, we use a Bernoulli (binomial) quasi-maximum likelihood method (QLME) proposed by Papke and Wooldridge (1996) to estimate the model of the fractional dependent variable. We replicate the specifications in the second and third columns of Table 6, and the QLME results are now shown in Table 7. A comparison between the two tables shows that the OLS results are pretty robust in terms of the relative magnitude, direction and statistical significance of the coefficients (the only exception is that the significance level of the coefficient on *Disclosure* changes from 10% to 5%).

Variable	Coef. (Std. Err) ^I	Coef. (Std. Err) ^{II}
Bonus	0.238 (0.096)**	0.223 (0.102)**
Disclosure	0.196 (0.096)**	0.300 (0.102)***
Closed	-0.300 (0.101)***	-0.261 (0.114)**
Number of Observations	552	483

*** significant at 1% level; **significant at 5% level, *significant at 10% level
 I: Full sample; the dependent variable is Helpfulness.
 II: Subsample (with product usage); the dependent variable is Helpfulness.

Discussions

The Marginal Effects of Bonus and Disclosure

When we run the regression using full sample, the marginal effect of bonus is bigger than that of disclosure. However, if we restrict our attention to the subsample with only reviewers who have some experience using the products, we see a larger marginal effect of disclosure. In particular, for reviewers who reported that they had used the products before, the requirement to disclose material connections increases the review helpfulness by 0.056, which is 0.018 higher than the value we obtained with full sample. This implies that, when required to add disclosure text, reviewers with product usage experience have more incentive to work harder in the review writing process. Although all the reviewers, with or without experience for product usage, desire good images in others' eyes, the reviewers who have interacted with the focal products before, are much more likely to generate higher quality reviews since their advantage in product information allows them to better describe and evaluate the products to other consumers.

Disclosure Text and User Reaction

All previous results are based on the helpful votes we got for the review text itself, with disclosure text excluded for the *Disclosure* condition. Although we have shown that the requirement to add disclosure text can in some extent improve the quality of the review, we need to be careful when applying this finding to the real online product sites. It would be interesting to see how people would react to the reviews if disclosure text is added. We examine this question using both AMT workers and users on Amazon.com.

When we asked AMT workers to evaluate the helpfulness of the reviews, we created two versions for each review: review text only, and review text plus disclosure text. We found that there are no significant differences between the helpful votes of the two versions. In other words, AMT workers are indifferent to the presence of disclosure text. (Notice that the reviews in which the reviewers were told that the disclosure will appear were still rated higher, no matter if the readers saw the disclosure or not). There are some possible explanations. First, AMT workers are, in fact, not the potential buyers who would like to purchase the products, so they are less likely to be upset about sponsored reviews. Second, workers who evaluated many reviews might get used to the disclaimer in the reviews. Therefore, they would implicitly filter out the text in disclaimer when evaluating the reviews.

On the other hand, users on Amazon.com differ in their attitudes about the disclosure. Although we don't have sufficient quantitative data due to lack of votes, we have observed a number of comments for reviews with disclosure text. In fact, most users on Amazon.com react to the disclosure of material connections very negatively, and only a few have a mixture of attitudes. We quote some comments here:

- *Negative Attitude:* "Compensated reviewer? Although their opinion may be perfectly valid, **I'm afraid I have to discount it. Shame, I look to reviews for independent insight.**"; "**I don't understand**, Amazon is paying for reviews now? Or is it the publisher? For the sake of *real* disclosure, who paid for the review and how did they pay for it/how much?"
- *Mixture of Attitudes:* "**It's great that you disclaimed your 'compensation'** but can we please please PLEASE not start having paid reviewers on Amazon. It's one of the last places you can go to hear what general consumers think. I believe everyone here should click a button before submitting reviews saying they are not affiliated in any way with the product or Amazon, and they were not paid to review. Otherwise this is going to be become one huge advertorial. And what's more, I like this product - a lot. **It doesn't need this form of cheap marketing.**"

Generally, users on Amazon.com are upset about the (disclosure of) payment. This is somewhat ironic since what triggers the "discounting" is the disclosure of the payment, while the same review would have not received negative reactions if the disclosure was not posted. Implicitly, users assume that all reviews posted on Amazon are contributed by unpaid volunteers with no connection to the product, something that is not necessarily true.

Another important message we get from here is that given that the reviews are compensated, users like the disclosure action of the reviewers. At the same time, they are willing to know more details of the compensation like sponsor identity, amount of payment. So, a clear statement of the truthfulness of the

review, together with all the details about the sponsorship might help to alleviate the offensiveness of paid reviews.

Conclusions

In this work, we conduct a behavioral experiment on AMT to examine the conditions under which workers can write higher quality paid reviews. In particular, we are interested in the role of that bonus rewards, sponsorship disclosure, and choice freedom play in review quality. We first use two-way fixed effects model to compare the quality of paid and unpaid reviews. Then, we employ a product-category-level fixed effects model to explore the effects of performance-contingent rewards, sponsorship disclosure, and restricted choice.

We find that subjects are pretty consistent in terms of their review quality in our experiment and in real life. Also, completion-contingent rewards have no significant effects on the quality of reviews. When completion-contingent rewards are present, the introduction of performance-contingent rewards can motivate workers to write high-quality reviews. When paid reviewers are required to add disclosure text, they are more likely to write reviews that are perceived helpful. The restriction for the product choice will generally undermine the quality of paid reviews. Our results still hold when we use the review length as a proxy for the review quality. Moreover, when required to disclose material connections, paid reviewers are more likely to create helpful reviews if they have used the products before.

Our work has some implications for companies and websites who aim to solicit online reviews. For example, paying an extra amount of money for reviewers who write “very helpful” reviews can potentially motivate the reviewers to exert more effort in the review creating process. Also, giving reviewers more freedom to choose the products they will review would satisfy their psychological need for autonomy and increase their intrinsic motivation. In addition, the requirement for reviewers to disclose material connections not only obeys FTC’s principles for regulating the online advertising, but also encourages reviewers to write high-quality reviews. Although the community has not appreciated the paid reviews as an important and legitimate approach to share information, companies and websites have the right and responsibility to create a healthy and honest environment for paid reviews and build trust among users and paid reviewers.

While our work has made some progress in understanding the factors that affect the quality of reviews, we acknowledge that our approach still has limitations. First, all our subjects are recruited from AMT, which might be biased in the sense that they are mostly money-driven, so the conclusion of the effects of monetary rewards on review quality may not be generalizable to the population of all online consumers. In addition, reviewers who are intrinsically motivated may be refrained from contributing after observing sponsored reviews. Whether this is true or not still needs further investigation. Second, when examining the quality differences between paid and unpaid reviews, we only have 26 paid reviews and 82 unpaid reviews, which is a relatively small sample size. Third, we use the fraction of helpful votes provided by AMT workers to measure the quality of the review but this may differ from the helpful votes assigned by the consumers, although in our tests with “gold” reviews we did not detect any such biases. Fourth, in this paper, the experiment is conducted on a third-party platform. More studies are needed to see how much these results can be generalized to the cases where sellers are the ones who reward reviewers.

There are many interesting directions for future research. We want to examine how different compensation levels, either for completion-contingent rewards or for performance-contingent rewards, would affect the quality of the reviews. We also need to do more research on how to encourage the disclosure of material connections and promote more truthful and honest reviews: if sponsored reviews are inevitable, it is better to know which reviews are paid, and devising the appropriate incentives is of importance for the future of the online reviewing communities.

References

- Alchian, A. A., and Demsetz, H. 1972. “Production, Information Costs, and Economic Organization,” *American Economic Review* (62:5), pp. 777-795.
- Alexy, O., and Leitner, M. 2007. “Norms, Rewards, and their Effect on the Motivation of Open Source Software Developers,” available online from <http://opensource.mit.edu>, *Working Paper*.

- Baumeister, R., and Leary, M. A. 1995. "The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation," *Psychological Bulletin* (117), pp. 497-529.
- Bock, G. W., Zmud, R. W., Kim, Y. G., and Lee, J. N. 2005. "Behavioral Intention Formation in Knowledge Sharing: Examining the Roles of Extrinsic Motivators, Social-Psychological Force, and Organizational Climate," *MIS Quarterly* (29:1), pp. 87-111.
- Burgess, D. 2005. "What Motivates Employees to Transfer Knowledge Outside Their Work Unit?," *Journal of Business Communication* (42:4), pp. 324-348.
- Cameron, J., Pierce, W. D., Banko, K. M., and Gear, A. 2005. "Achievement-Based Rewards and Intrinsic Motivation: A Test of Cognitive Mediators," *Journal of Educational Psychology* (97:4), pp. 641-655.
- Daugherty, T., Eastin, M. S. and Bright L. 2008. "Exploring Consumer Motivations for Creating User-Generated Content," *Journal of Interactive Advertising* (8:2), pp. 1-24.
- Deci, E. L. 1971. "Effects of Externally Mediated Rewards on Intrinsic Motivation," *Journal of Personality and Social Psychology* (18:1), pp. 20-28.
- Deci, E. L. 1975. *Intrinsic Motivation*, New York: Plenum Press.
- Deci, E. L., and Ryan, R. M. 1980. "The Empirical Exploration of Intrinsic Motivational Processes," *Advances in Experimental Social Psychology* (13), pp. 39-80.
- Deci, E. L., and Ryan, R. M. 1985. *Intrinsic Motivation and Self-Determination Theory in Human Behavior*, New York: Plenum Press.
- Deci, E. L., and Ryan, R. M. 2000. "The 'What' and 'Why' of Goal Pursuits: Human Needs and the Self-Determination of Behavior," *Psychological Inquiry* (11:4), pp. 227-268.
- Duan, W., Gu, B., and Whinston A. B. 2008. "Do Online Reviews Matter? – An Empirical Investigation of Panel Data," *Decision Support Systems* (45:4), pp. 1007-1016.
- Fama, E. F., and Jensen, M. C. 1983. "Separation of Ownership and Control," *Journal of Law and Economics* (26:2), pp. 301-325.
- Frey, B. S. 1994. "How Intrinsic Motivation Is Crowded Out and In," *Rationality and Society* (6:3), pp. 334-352.
- Ghose, A., and Ipeirotis, P. G. 2007. "Designing Novel Review Ranking Systems: Predicting Usefulness and Impact of Reviews," in *Proceeding of the 9th International Conference on Electronic Commerce*, pp. 303-310.
- Ghose, A., and Ipeirotis, P. G. 2011. "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," in *IEEE Transactions on Knowledge and Data Engineering* (23:10), pp. 1498-1512.
- Jindal, N., and Liu, B. 2008. "Opinion Spam and Analysis," in *International Conference on Web Search and Web Data Mining*, pp. 219-230.
- Kail, R. V. 1975. "Freedom of Choice, Task Performance and Task Persistence," *Journal of Experimental Education* (44:1), pp. 32-35.
- Mason, W., and Suri, S. 2011. "Conducting Behavioral Research on Amazon's Mechanical Turk," *Behavioral Research Methods* (44:1), pp. 1-23.
- Monty, R. A., and Rosenberger, M. A. 1973. "Amount and Locus of Choice as Sources of Motivation in Paired-Associate Learning," *Journal of Experimental Psychology* (97:1), pp. 16-21.
- Mudambi, S., and Schuff, D. 2010. "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com," *MIS Quarterly* (34:1), pp. 1-12.
- Papke, L. E. and J. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics* (11:6), pp. 619-632.
- Park, N., Jung, Y., and Lee, K. 2011. "Intention to Upload Video Content on the Internet: The Role of Social Norms and Ego-Involvement," *Computers in Human Behavior* (27:5), pp. 1996-2004.
- Patall, E. A., Cooper, H., and Wynn, S. R. 2010. "The Effectiveness and Relative Importance of Choice in the Classroom," *Journal of Educational Psychology* (102:4), pp. 896-915.
- Ryan, R. M., and Deci, E. L. 2000. "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development and Well-Being," *American Psychologist* (55), pp. 68-78.
- Staw, B. M., Calder, B. J., Hess, R. K., and Sandelands, L. E. 1980. "Intrinsic Motivation and Norms about Payment," *Journal of Personality*, (48:1), pp. 1-14.
- Yang, H., and Lai, C. 2011. "Understanding Knowledge-Sharing Behaviour in Wikipedia," *Behaviour & Information Technology* (30:1), pp. 131-142.
- Zhang, Z., and Varadarajan, B. 2006. "Utility Scoring of Product Reviews," in *CIKM' 06 Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, pp. 51-57.
- Zhu, J., and Tan, B. 2007. "Effectiveness of Blog Advertising: Impact of Communicator Expertise,

Advertising Intent, and Product Involvement,” in *ICIS 2007 Proceedings*. Paper 121.