What Are Social Incentives Worth? A Randomized Field Experiment in User Content Generation

Completed Research Paper

Gordon Burtch
Carlson School of Management
University of Minnesota
gburtch@umn.edu

Yili Hong
W.P. Carey School of Business
Arizona State University
ykhong1@asu.edu

Ravi Bapna
Carlson School of Management
University of Minnesota
rbapna@umn.edu

Vladas Griskevicius
Carlson School of Management
University of Minnesota
vladasg@umn.edu

Abstract

Content generation is a critical aspect of user engagement in online communities, yet many platforms face a problem of under-provision. We focus here on the potential of different types of incentives (descriptive social norms and money) to stimulate the production of online reviews. Partnering with a Chinese online clothing retailer, we conducted a large-scale randomized field experiment, in which we considered the independent and joint effects of monetary payment and descriptive social norms on the quantity and quality of reviews. We found that money attracts a greater volume of reviews, descriptive social norms attract greater quality, and combining the two yields the greatest benefit in both respects. We discuss the mechanisms underlying these effects as well as the implications of our results for theory and practice. We also highlight potentially fruitful avenues of future research.

Keywords: user-generated content, social norms, online reviews, experimental economics

Introduction

The under-provision of user-generated content (UGC) has reportedly resulted in the demise of many an online community. In recent years, a number of scholars have undertaken studies of possible incentives to stimulate UGC production. The incentives that have been considered to date can generally be classified into one of two groups: *social* and *monetary*. Chen et al. (2010) provide an example of work exploring *social* incentives. Those authors found that supplying users with information about the average number of movie reviews written by peers caused a regression toward the average; treated users initially contributing below the median increased their production sharply, while those initially contributing above the median reduced their production. Wang et al. (2012) provide an example of work looking at *monetary* incentives. Conducting a study at Amazon Mechnical Turk, they reported no significant differences in the baseline quality of paid and unpaid reviews, but they did report that review quality could be improved via quality-contingent bonuses.

Studies of the sort described above ultimately contribute to a much broader literature on the use of social and monetary incentives to stimulate private contributions to public goods. A frequent consideration in that broader literature has been the application of combinations of incentives. In particular, scholars have discussed reasons why combining monetary and social incentives might not result in the additive increases to private contributions that intuition would suggest. From a psychological standpoint, it has been argued that this is because subjects tend to ignore social benefits and focus almost exclusively on monetary rewards whenever they are offered (Gneezy and Rustichini 2000; Heyman and Ariely 2004), because the presence of any remuneration causes individuals to operate in a market-based mindset. An economic rationale has also been proposed for the same pattern of influence; it has been argued that monetary incentives undercut social incentives because those individuals most inclined to respond to social benefits are those most concerned about image perception. That is, these social minded individuals will be concerned that others will think less of them, once they become aware that they have received monetary compensation for their efforts (Benabou and Tirole 2006).

This work contributes to the literature on UGC by exploring not only the degree to which social incentives and monetary compensation jointly and independently influence the quantity of UGC, but also their parallel effects on the quality of content. The latter analysis is particularly novel, as no prior work, to our knowledge, has attempted to disentangle the effect of different incentives on quantity and quality of content production.

We undertake a large-scale randomized field experiment with an online retailer, attempting to motivate consumers to provide an online review following product purchase, an approach that provides the combined benefit of realism and scale. Randomly assigning subjects to one of four conditions (control, social, monetary, and social + monetary), we subsequently examine the rate and quality of content production across groups. We find that monetary and social incentives related to review authorship impact the rate of UGC production in a manner consistent with prior theory and evidence from the literature in psychology and economics. That is, both types of incentives are effective in stimulating higher volumes of UGC production, but they produce a negative interaction, consistent with the idea that social incentives lose their effectiveness in the presence of remuneration. However, we find a quite different pattern when it comes to the quality of the UGC produced. We find that UGC produced under pure monetary incentives is of relatively low quality. In contrast, socially incentivized content is of significantly higher quality than organically produced content, and this remains true even when monetary incentives are introduced.

We attribute this overall result to the two-step nature of user content generation; the decision to participate is followed by a decision about how much effort to exert. In our experiment, social and monetary incentives are related to participation, rather than performance (i.e., quantity and not quality). Once a consumer opts to author a review, the financial incentive ceases to be in play, because it is a given. In contrast, the social norm related to authorship, though operationalized in terms of quantity, appears to spill over into quality. This is likely to manifest because individuals who perceive a social norm for review quantity then infer a correlated pro-social norm for review quality – e.g., "if you're going to do something, do it well."[1]

Generally, our results suggest that, by offering participation-based social incentives in tandem with monetary incentives, firms can stimulate a greater volume of UGC production, and also greater quality. In particular, we find that the joint provision of social and monetary incentives nearly triples the probability that a consumer will write a review, and results in reviews that are approximately 50% longer, and 25% more helpful.

The remainder of this paper is structured as follows. In the next section, we begin with a review of the literature pertaining to private contributions to public goods, and we explore the application of those notions to the context of online user-generated content (online reviews). We then detail our research context and experimental design, report our empirical results, and offer an interpretation and discussion. We then draw a series of managerial and theoretical implications, and conclude by suggesting a number of avenues for future research in this area.

Literature Review

Numerous scholars have considered the incentives underlying private contributions to public goods in recent decades (Andreoni and Bernheim 2009; Benabou and Tirole 2006; Daughety and Reinganum 2010). A common element in much of that work has been an observation that individuals consider not only the direct returns to contribution, but also social factors. This is true for at least two reasons. First, contributors may become concerned about their social image, wishing to maintain a *fair* persona, one that adheres to social norms and acts in accordance with the group (Becker 1974; Daughety and Reinganum 2010). Second, if the contributor's audience is also the potential market for their contributions, individuals may perceive that their contributions have a greater potential to deliver reputational benefits (Zhang and Zhu 2011). Moreover, contributors may experience a greater warm glow effect, anticipating that their contributions will deliver a greater overall benefit for others (Andreoni 1990).

A number of studies in recent years have considered questions related to public good contributions in contexts of interest to scholars in the information systems discipline. Examples include studies of open-source software development (Lakhani and von Hippel 2003; Lerner and Tirole 2002; Roberts et al. 2006), Wikipedia article writing (Ransbotham and Kane 2011; Zhang and Zhu 2011), contributions to crowdfunding campaigns (Burtch et al. 2013), and the writing of online product reviews (Chen et al. 2010; Wang 2010), to name a few.

Online reviews, in particular, have received a great deal of attention, because under-provision is a known issue in that context. Avery et al. (1999) describe the issue at length and propose a number of design mechanisms for overcoming the problem. More concretely, Jindal and Liu (2008) report that of the more than 3 million product listings that they consider on Amazon.com, approximately 50% have just one review, and just 19% have more than 5 reviews. Thus, the findings of Chen et al. (2010), which indicate that social incentives (providing users with information about their peers' reviewing activity to establish a social norm) can be used to overcome the problem of under-provision, are quite promising.

A number of platforms have also considered more direct approaches to stimulating review production, offering to compensate consumers directly for their efforts. Most famously, the media reported that Amazon was enticing some consumers to author reviews by sending them free merchandise.[2] Recent work has thus subsequently considered various implications of paying consumers to author reviews. For example, Stephen et al. (2012) report that, when paid, authors of reviews perceive no difference in the effort required to write a review, or the level of difficulty in doing so. Wang et al. (2012) conduct an online experiment with subjects drawn from Amazon Mechanical Turk, and find that paid and unpaid reviews are of roughly equivalent quality. However, they go on to show that the quality of reviews can be increased

---

[1] Additional experimentation is of course necessary (and we are presently conducting it) to determine the relative effects of performance-based (rather than participation-based) monetary and social incentives.

[2] http://www.npr.org/blogs/money/2013/10/29/241372607/top-reviewers-on-amazon-get-tons-of-free-stuff

by offering performance contingent bonuses, and by informing an author that their compensation will be made public.

Interestingly, no prior work has considered the joint introduction of social and monetary incentives in the context of online reviews, as we aim to do here. Although a number of studies, some noted above, suggest that social incentives may cease to operate in the presence of monetary incentives, whether because a market-based mindset takes over (Gneezy and Rustichini 2000; Heyman and Ariely 2004), or because the presence of monetary incentives mitigates reputational gains (Benabou and Tirole 2006), we consider a rather different form of social incentive (social pressure from an established norm, as opposed to the non-monetary gifts that Heyman and Ariely employ), and we explore a context in which public knowledge about compensation is not guaranteed. As such, the relevance of prior findings is not a given. Moreover, we separately consider the effects of each incentive type on i) the mere decision to contribute and ii) the effort invested conditional on contribution. As such, our work has the potential to provide novel insights into the dynamics of contributor decision making, which may have implications for the practical application of different incentive types.

Methods

Study Context

We partnered with a retailer located in China, a large online seller of children's apparel that offers its products for purchase via TMall. TMall is an online intermediary, similar to Taobao, JD and Amazon, which hosts retailers online sales operations and also allows customers to write and post online reviews about product purchases (see Figure 1). TMall is owned by Alibaba (NYSE: BABA). Unlike Taobao, which is targeted toward small individual sellers, TMall generally hosts larger businesses, which utilize its marketplace to advertise, promote and sell their products. Businesses on TMall can engage with customers in many ways, offering product promotions, discount coupons, and even issuing targeted SMS text messages. After a purchase transaction, the buyer can optionally choose to submit a review for the product[3]. Although SMS has traditionally been used by TMall's online retailers to communicate product delivery notices, buyers may also receive promotional SMSs from time to time as well.

Scholars have begun to leverage the SMS communication channel for the purposes of randomized experimentation, to answer a variety of questions related to electronic commerce. One recent example of this is the work of Andrews et al. (2015), who have delivered targeted promotions to potential purchasers via SMS messaging, exploring the moderating impact of physical context on the likelihood of subsequent purchase. Using SMS messaging to communicate with customers has a number of advantages over email. First, only certain subsets of the population regularly use email in China (youth, white collar workers); in contrast, much larger segments of the population have a cellular phone and check that phone frequently. Second, because cellular numbers are recorded as part of a buyer's shipping address information (in China it is common practice for the carrier to contact the buyer via phone call or SMS before delivering an item), a business can maintain greater confidence that communications have indeed been received by the customer. Third, and last, it has been reported that as much as 20% of all promotional emails are flagged as spam by email service providers, and thus never delivered to customers.[4] This is an issue that does not arise with SMS messaging. For the various reasons mentioned above, we chose to leverage the SMS messaging capability of our retail partner, to apply our treatment conditions to a random set of consumers.

Experiment Design

Our subject pool is comprised of 2,000 customers of our retail partner. Subjects were identified at random from the set of all customers who had completed an online product purchase within a 24-hour

---

[3] Please note that in this context, the merchant and product review systems are owned and maintained by TMall, not the retailers. Only aggregate ratings are displayed on the TMall website and provided to the individual retailers - individual customer ratings are not available. We therefore do not have measures of review valence (star ratings); we have only the text of each review for use in our analysis.

[4] See recent research report by ReturnPath: http://blog.returnpath.com/blog/return-path-2/email-deliverability-still-plagues-commercial-email-senders-worldwide-only-81-of-email-reaches-the-inbox

period prior to the time of our treatment. We randomly assigned each subject to one of five conditions (400 subjects per condition). We then used a "push" approach to deliver our treatments to the subjects via SMS messaging. In our control condition, subjects were not contacted at all.

In our generic message condition, subjects received a standard SMS message, asking that they complete an online review for their recent product purchase. In our money condition, subjects were asked to write a review, and were also told that they would be paid ¥10 upon doing so (approximately $1.50 USD).[5] In our social condition, subjects were asked to write a review, and were told the volume of reviews recently authored by other customers of the retailer within the prior month. Finally, in our interaction condition, subjects were asked to author a review, were told that they would be compensated upon doing so, and were informed of other recent customer reviewing volumes. Please note that we did not ask buyers to offer a good or positive review; we simply asked that they provide feedback.

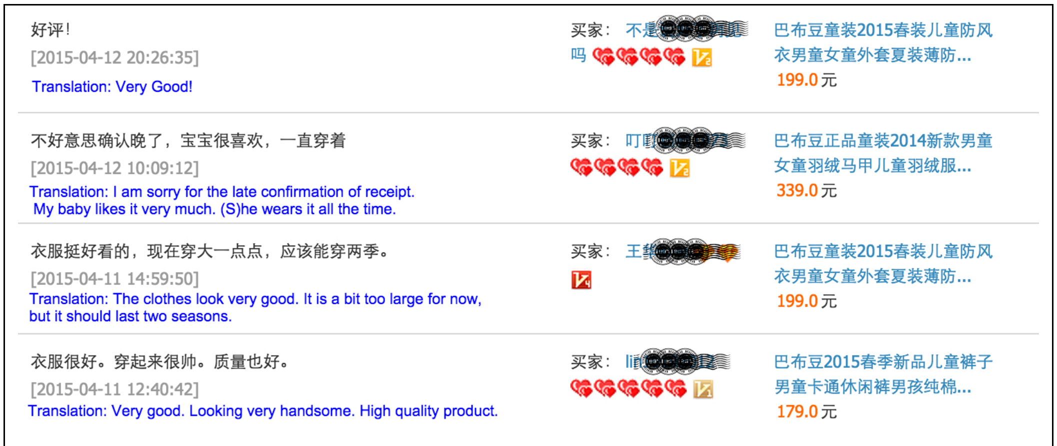

Figure 1. Sample Reviews from Customers of Retail Partner

One week after the treatment time, the retailer supplied us with information about which customers ultimately authored a review for their product purchase. Additionally, the retailer supplied us with the textual content of each review. This latter aspect is important, because it allows us to construct a measure of content quality. In this case, we operationalize quality based on three factors: the character length of a review, and measures of helpfulness and diagnosticity, based on the results of a manual coding effort (more details below). Thus, we consider four outcome variables in our eventual analysis: the probability of authoring a review, and our three measures of review quality, conditional on authorship. Our five treatment conditions are summarized in Table 1. Figure 2 presents a mockup of the SMS message content that was issued to subjects in our interaction condition. For reference, we also provide an English translation of the SMS message (note: these translations were confirmed by three coders, fluent in both languages).

[5] This payment amount is in line with the recent literature, which has typically offered subjects $1 USD in exchange for authoring an online review (Stephen et al. 2012; Wang et al. 2012). Additionally, we explored an additional treatment condition not reported here, in which we offered subjects ¥5. We observed no statistically significant difference between the ¥5 and ¥10 groups for any of our dependent variables.

Table 1. Treatment Conditions	
Condition	**Description**
Control	No SMS was issued.
Generic SMS	A generic SMS was issued, asking the subject to write an online review.
Monetary SMS	An SMS was issued asking the subject to write an online review, and promising to compensate him or her with ¥10 (~$1.50 USD).
Social SMS	An SMS was issued asking the subject to write an online review, and informing him or her of the number of reviews other customers had written within the prior month.
Monetary + Social SMS	An SMS was issued asking the subject to write an online review, promising to compensate him or her with $1.50, and informing him or her of the number of reviews written by other customers within the prior month.

Figure 2. Treatment Mockup (Monetary + Social SMS)

Results

We observe several outcomes from the experiment. First, we look at a binary indicator of whether a subject authored a review. Second, we look at the text of any authored review. We operationalize quality in terms of review length, counting the number of characters used. Third, we consider the helpfulness of a review. The helpfulness of each review was coded manually by two research assistants. These helpfulness ratings were based on two questions, relating to review diagnosticity and perceived helpfulness.

To ensure that the two coders employed a consistent approach, we conducted two instructional sessions, using 35 reviews of products sold by the same merchant (note that these 35 reviews did not come from our experimental sample). In the first instructional session, the concept of review diagnosticity was explained to the coders. The coding assistants and one of the study authors then proceeded to code 10 reviews together, to help the coders better understand the task. Based on the literature, a review may be viewed as having high diagnosticity if it helps consumers to identify product attributes, and to characterize those attributes as being either positive or negative (Jiang and Benbasat 2004; Jiang and Benbasat 2007). In contrast, perceived helpfulness is a more subjective measure, reflecting a buyer's evaluation of how useful a particular review is in coming to a purchasing decision.

In the second instructional session, the students were asked to code the remaining 25 reviews, and to then reconvene, to compare and discuss any coding discrepancies. Following the two instructional sessions, the coding assistants were asked to independently code all of the reviews generated in our experiment, in terms of review diagnosticity and perceived helpfulness. We assessed measurement validity and consistency of the coding process via Cronbach's Alpha and Krippendorff's Alpha. Constructing our composite measure of perceived helpfulness from the results reported by our two coders, we observe a Cronbach's Alpha of 0.851 and a Krippendorff's Alpha of 0.706. For our composite measure of review diagnosticity, we observe a Cronbach's Alpha of 0.884, and a Krippendorff's Alpha of 0.781. Each of these values is well in excess of standard cutoffs for acceptable use in the literature (Kline 2000 pg. 13; Krippendorff 2004).

Our empirical analysis begins with a consideration of any average differences in the probability of authorship, the length of reviews, and the coded characteristics of reviews across conditions. Table 2 presents our descriptive statistics for the variables that enter into our analysis. We first consider the impact of each treatment on the probability that a subject authors an online review within the seven days following the treatment event. We then consider the impact of each treatment on the quality of written reviews, where quality is variably measured in terms of review length, diagnosticity and helpfulness. Finally, we combine the two sets of analyses (decision to author, and quality of authorship), by exploring the impact of each treatment on review quality in an unconditional manner (i.e., we populate a length, diagnosticity and helpfulness value of 0 for those subjects that chose not to author a review). This last set of analyses allows us to identify whether and which of our treatment conditions dominate in terms of their net effect on high-quality content production (given that increases in review volume are arguably of little use if the resultant content is of low quality).

Table 2. Descriptive Statistics: Sample-Wide

Variable	**Mean**	**St. Dev.**	**Min**	**Max**	**N**
Authorship	0.114	0.317	0.000	1.000	2,000
Length	13.41	8.051	1.000	42.000	227[x]
Perceived Helpfulness	2.901	1.461	1.000	6.500	227[x]
Review Diagnosticity	2.892	1.362	1.000	7.000	227[x]

Notes: x – 227 subjects wrote a review, out of 2,000 – this value reflects only authored reviews

We begin by graphically depicting the differences in average rates of reviewing (Figure 3), and review length across our various treatment conditions (Figures 4 and 5 – note that our y-axis reflects the number of Chinese characters, and 1 Chinese character translates roughly to 5 English characters). Here, we immediately begin to see distinct differences, indicating that our treatment is having a strong effect. The monetary condition is associated with the greatest rate of authorship (regardless of the presence of social incentives). However, the combined condition is most clearly associated with unconditional review quality (the joint result of authorship and effort).

We next report more formal econometric estimations, based on Linear Probability Models (Horrace and Oaxaca 2006) and ordinary least squares. We begin by estimating the relationship between our various treatment conditions and the probability of a subject authoring an online review. To conduct this analysis, we estimate a linear probability model, relating our binary outcome (review) to dummy indicators for each of our treatment conditions (Equation 1).

$$Authorship_i = \alpha + \sum_{\rho} Treatment_i^{\rho} + \varepsilon_i \tag{1}$$

$$Quality_i^{\delta} = \alpha + \sum_{\rho} Treatment_i^{\rho} + \varepsilon_i \tag{2}$$

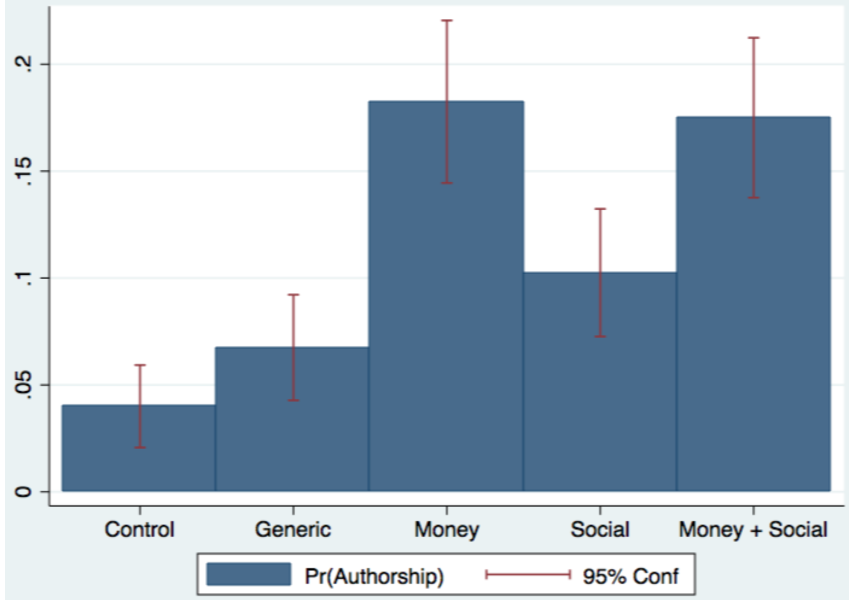

Figure 3. Mean + 95% CI of Pr(Authorship)

We then proceed with our second-stage estimation, relating the quality of a review (we have three measures of quality, conditional on authorship, indexed by δ: review length, diagnosticity and helpfulness) to a subject's treatment condition (Equation 2), employing ordinary least squares (OLS). Lastly, we repeat our second estimations, substituting a value of 0 for our length measure and 1's for diagnosticity and helpfulness, in those cases where no review was authored (minimum length, helpfulness and diagnosticity). This last estimation enables us to estimate the net effect of each treatment on ultimate content production. This approach jointly considers the subject's decision to produce a review and the effort they then put into their review, conditional on production (note: in each regression, our reference category is the control condition).

The results of our regressions align with the findings reported in our descriptive analyses and graphical depictions of group averages. Moreover, we repeated each regression incorporating additional controls for the user's purchase channel (mobile vs. desktop), location (a province fixed effect) and purchase total (total spend). Although we identified some significant effects (e.g., some provinces), the various treatment effects remain effectively the same, both in terms of sign and significance, as well as in terms of relative magnitude.

Looking first at column 1, we observe that all of our treatments have a significant, positive impact on the probability that a customer authors an online review. However, the different treatments vary quite a bit in their efficacy. Because we estimate our specification using a linear probability model, we can interpret the coefficients directly as marginal probabilistic effects. Thus, we find weak support for the notion that a simple reminder can boost content generation (the Generic condition), though the effect is relatively small compared to our treatments of interest (a boost of ~3% over the baseline probability of 4%). Offering money has a much larger impact, increasing the probability of authorship by more than 14%. In contrast, our social condition, in which we attempt to establish a social norm by advising the customer about the rate of review authorship amongst his or her peers, only increases the probability of authorship by 6%.

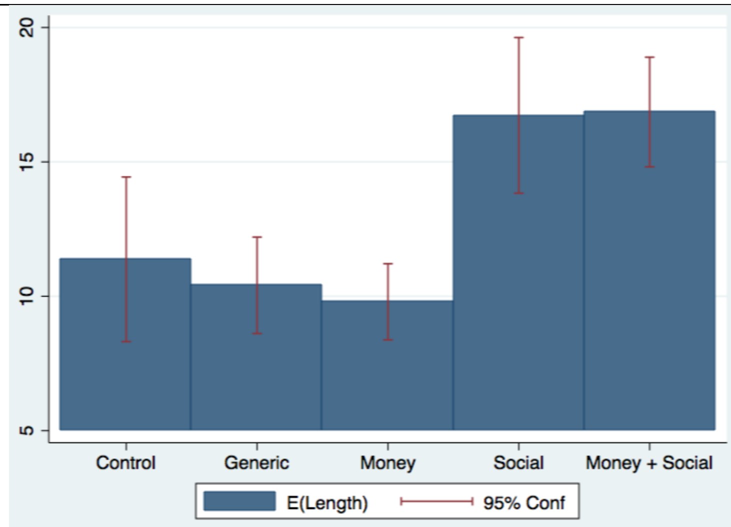

Figure 4. Mean + 95% CI of Length (Given Authorship)

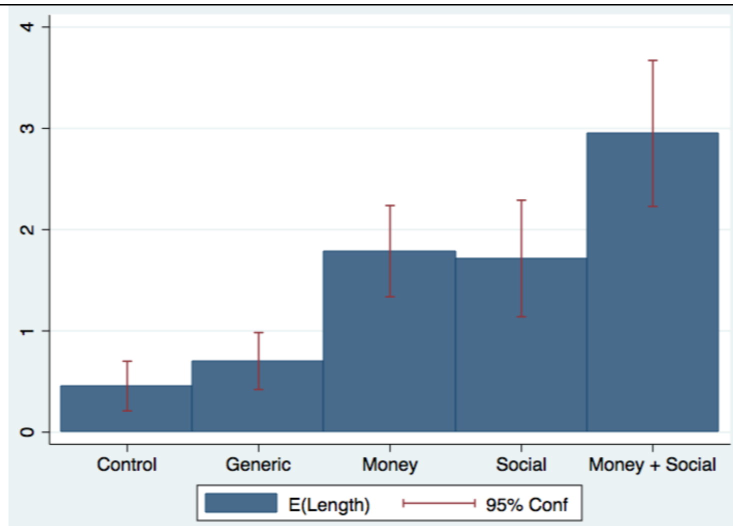

Figure 5. Mean + 95% CI of Length (Unconditional; No Review = 0)

Finally, the joint condition, in which we provide both money and information on peer authorship, results in a boost to authorship rates that is comparable to that observed in our money condition. Indeed, we are unable to reject the equivalence of the two coefficient estimates at conventional levels (F = 0.08 (1, 1995), p = 0.782), which suggests that there is no incremental impact to offering social incentives in the presence of monetary incentives.

Next, examining columns 2 through 4, we observe no significant effect from a generic reminder or offering a monetary incentive by itself on the quality of authored reviews, for all three measures. Consistent with other recent work on paid reviews, we observe that paid reviews are roughly equivalent to organic reviews in terms of quality (Wang et al. 2012; Stephen et al. 2012). Our social condition, however, results in reviews of significantly greater quality, in terms of length, diagnosticity and helpfulness. Moreover, introducing a monetary incentive does not have a significant impact on any of our quality measures. For all three outcome variables, once again, we are unable to reject a hypothesis of equivalence between the coefficient estimates for the Social and Money + Social conditions, at conventional levels

(Length: $F = 0.01$ (1, 222), $p = 0.943$; Diagnosticity: $F = 0.09$ (1, 222), $p = 0.764$; Helpfulness: $F = 0.22$ (1, 222), $p = 0.638$).

Table 3. Regression Results (Authorship & Conditional Quality)

Explanatory Variable	LPM (1) Authorship	OLS (2) Conditional Length	OLS (3) Conditional Diagnosticity	OLS (4) Conditional Helpfulness
Generic SMS	**0.028+ (0.016)**	-0.968 (1.652)	-0.206 (0.364)	-0.212 (0.340)
Monetary SMS	**0.143*** (0.022)**	-1.580 (1.578)	-0.277 (0.329)	-0.225 (0.321)
Social SMS	**0.063** (0.018)**	**5.357** (2.007)**	**0.703+ (0.369)**	**0.624+ (0.362)**
Monetary + Social SMS	**0.135*** (0.021)**	**5.482** (1.742)**	**0.620+ (0.340)**	**0.765* (0.342)**
Constant	**0.040*** (0.010)**	**11.375*** (1.407)**	**2.688*** (0.296)**	**2.656*** (0.282)**
Observations	2,000	227	227	227
F-stat	**18.03 (4, 1995)**	**11.61*** (4, 222)**	**6.52*** (4, 222)**	**6.05*** (4, 222)**
R-squared	0.032	0.174	0.106	0.099

Notes: *** $p < 0.001$, ** $p < 0.01$, + $p < 0.10$; Robust standard errors in parentheses; reference category is the control condition, where customers were not contacted via SMS.

Taken together, the impact of our treatment conditions on authorship and quality appear to run in different directions. On the one hand, money is more impactful when it comes to stimulating content production, and introducing social incentives has no incremental impact in the presence of money. On the other hand, social incentives are more impactful when it comes to stimulating high quality content, and once again, introducing money has no incremental impact. Taken together, these findings appear to suggest a strictly dominant condition: offering monetary incentives and social incentives, in tandem.

To gain a clear understanding of this net effect, and to determine whether our intuition is correct, we next repeated our quality estimations, this time in an unconditional manner (i.e., populating 0 values for length, and values of 1 for the composite scores of diagnosticity and helpfulness, whenever a subject did not author a review). This approach enables us to understand the joint impact of each treatment on both the probability of authorship and the quality of authored content. The results of these estimations are presented in Table 4.

Considering columns 1 to 3 in Table 4, we observe a pattern of estimates that is consistent with our initial interpretation. We observe that money and social incentives have independently similar net effects on the overall body of user-generated content. Money appears to operate predominantly by stimulating a larger volume of online reviews, whereas social incentives appear to operate predominantly by stimulating higher quality online reviews. Combining both incentives together appears to provide the best of both worlds, across all three of our quality measures. The coefficient estimates for the Monetary + Social condition are significantly larger than the estimates obtained for Money or Social independently, across all three of our estimations (all F statistics are significant at $p < 0.05$).

Finally, to help establish robustness of these results, we repeated our authorship estimations using logistic regression. These results are reported in Table 5, column 1. Here, we observe effectively identical results to our earlier models. Similarly, repeating our quality estimations for diagnosticity and helpfulness (measured on 1-7 Likert scales) using ordinal logistic regression (columns 2 and 3), we again obtain effectively identical results.

Table 4. Regression Results (Net Quality)

Explanatory Variable	**OLS (1) Unconditional Length**	**OLS (2) Unconditional Diagnosticity**	**OLS (3) Unconditional Helpfulness**
Generic SMS	0.248 (0.190)	0.033 (0.031)	0.031 (0.030)
Monetary SMS	**1.333*** (0.261)**	**0.190*** (0.043)**	**0.195*** (0.044)**
Social SMS	**1.260*** (0.318)**	**0.178*** (0.047)**	**0.168*** (0.046)**
Monetary + Social SMS	**2.495*** (0.387)**	**0.336*** (0.056)**	**0.358*** (0.060)**
Constant	**0.455*** (0.125)**	**1.068*** (0.020)**	**1.066*** (0.020)**
Observations	2,000	2,000	2,000
F-stat	**17.14*** (4, 1995)**	**14.17*** (4, 1995)**	**14.14*** (4, 1995)**
R-squared	0.031	0.026	0.027

Notes: *** $p < 0.001$; Robust standard errors in parentheses.

Table 5. Robustness Checks

Explanatory Variable	**Logit (1) Authorship**	**Ordinal Logit (2) Unconditional Diagnosticity**	**Ordinal Logit (3) Unconditional Helpfulness**
Generic SMS	**0.552+ (0.324)**	0.553 (0.341)	0.510 (0.344)
Monetary SMS	**1.679*** (0.286)**	**1.422*** (0.307)**	**1.420*** (0.307)**
Social SMS	**1.008** (0.304)**	**1.084** (0.321)**	**1.049** (0.322)**
Monetary + Social SMS	**1.627*** (0.287)**	**1.687*** (0.304)**	**1.657*** (0.305)**
Constant	**-3.178*** (0.255)**	--	--
Observations	2,000	2,000	2,000
Wald Chi^2	**57.55*** (4)**	**44.71*** (4)**	**43.95*** (4)**
Pseudo R-squared	0.048	0.025	0.024

Notes: *** $p < 0.001$, ** $p < 0.01$, + $p < 0.10$; Robust standard errors in parentheses.

Discussion

Our findings are important from a practical standpoint, because they demonstrate that user-generated content has the potential to be motivated in a costless manner. Based on a raw comparison of coefficient sizes between the Social and Money conditions from our estimation of unconditional length (Table 4, Column 1), the provision of social incentives in our experiment carried approximately 95% of the effect of monetary incentives (or ¥9.50). Our results are thus suggestive of a possible strategy that firms might look to employ to seed content and then transition to sustainable ongoing contributions. By first offering customers money to author reviews, firms might then eventually use the increased rate of reviewing to subsequently institute social mechanisms, ultimately transitioning to higher quality contributions.

Such a strategy would of course need to be implemented with caution, as the long-term effects of our treatments remain unclear. For example, our observation that money does not appear to crowd out effort, or intrinsic motivation (i.e., we observed no significant difference in review lengths between paid reviews and organic reviews), though consistent with prior work (Wang et al. 2012; Stephen et al. 2012), may not tell the entire story. A long stream of literature in psychology on financial incentives has repeatedly found that the effort individuals put into a task declines once the financial incentives is removed. It is therefore possible that money crowds out effort or intrinsic motivation in the long term. As such, an interesting avenue for future work would be to explore the longer-term impact of our treatments on consumer reviewing activity, after the cessation of the experiment.

Our work is of course subject to a number of limitations, aspects that can be explored or elaborated upon in future work. The effect of monetary incentives is likely to be non-linear in nature. That is, we might anticipate diminishing marginal returns to increasing levels of monetary compensation. Future work can explore this by considering varying levels of monetary compensation. Our results may also be contingent on the actual level of social proof that was present in the retail marketplace we have studied, at the time of our experiment. That is, we have not manipulated the information about prior others' reviewing activity, we have only randomized the provision of that information to new customers. It is possible that variation in the level of social proof (e.g., the rate of others' reviewing activity) would increase or decrease subject response.

Additionally, the format in which information is presented may drive different responses. It has been reported, for example, that individuals exhibit cognitive biases around raw number values. For example, the rule of 100 speaks to the idea that discounts reflecting amounts below $100 can be made more attractive to consumers if represented as percentages, because larger numerical digits tend to be perceived as representative of greater value (Berger 2013). A number of other numeracy-related cognitive biases have also been noted in the psychology and consumer behavior literatures, some of which suggest individual-level heterogeneity in response and comprehension of different forms of numerical information, presented in different formats (e.g., Dickert et al. 2011).

Our analyses might also be expanded upon in future work to control for the potential role of social impact from a contribution, as a moderator for the identified social norm effect. Past work has noted that altruism, in response to social incentives, can be moderated by the number of anticipated beneficiaries from a given contribution (Andreoni 2007). Other work has noted that the presence of a large audience may motivate contribution because it increases the potential for reputation building (Zhang and Zhu 2011). Although it is admittedly difficult for the author of an online reviewer to anticipate the number of individuals who will read his or her review, let alone factor its content into a purchase decision, certain design mechanisms could be pursued that would facilitate exactly that (e.g., indications of recent web traffic to a product URL).

Finally, it is also worth noting that our treatments may have other, orthogonal effects on the characteristics of online reviews, unrelated to quality. For example, paid reviews may result in inflated star-ratings, if consumers engage in some sort of reciprocity. However, though we do not report on the results here, our preliminary findings from a follow-on experiment (in which we replicate our treatment conditions, in addition to exploring others) exhibit no significant difference in star-ratings. Moreover, recent work around paid reviews has made a similar observation, that star-ratings are not affected by offering financial incentives (Stephen et al. 2012).

Conclusion

We have presented what is to our knowledge a first attempt to explore the independent and joint effects of monetary and social incentives as drivers of user content generation, in the context of online reviews. We have found that the influence of each incentive type varies, depending on whether one considers consumers' decision to contribute or the effort consumers invest, conditional on authorship. On the one hand, our results replicate the findings of past work, in the sense that both money and social incentives stimulate increased volumes of UGC, yet when combined, monetary incentives take over entirely. On the other hand, our results with respect to effort (proxied by various measures of review quality) differ. Here, social incentives are a primary driver of increased effort, whereas monetary incentives have little to no effect (in fact, if an effect does exist, it appears likely to be negative). When social and monetary incentives are combined, the social incentives maintain their efficacy.

This latter result seems to suggest that, at least in the context of online reviews, a consumer's authorship decision is distinct from their effort decision. This makes sense, given the finding of Wang et al. (2012)

that the quality of paid reviews grows in the presence of quality contingent bonuses. Given the two-step nature of the review authorship process, our design implies a very practical approach to seeding content in online settings: pay for reviews to establish a reviewing norm, then introduce social incentives to improve quality.

This work presents only a first step toward improving our understanding of the efficacy of different incentives in stimulating the production of high quality online reviews. We have identified a number of potentially fruitful avenues for future work, considering alternative social incentives, and alternative approaches to imposing those incentives. It is our hope that future work can develop these ideas further, to provide practical recommendations on the implementation of these recommendations, as well as theoretical insights about the psychological and economic mechanisms underlying them.

References

Andreoni, J. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal* (100:401), Dec 12, pp. 464-477.

Andreoni, J. 2007. "Giving Gifts to Groups: How Altruism Depends on the Number of Recipients," *Journal of Public Economics* (91:9), pp. 1731-1749.

Andreoni, J., and Bernheim, B.D. 2009. "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica* (77:5), pp. 1607-1636.

Andrews, M., Luo, X., Fang, Z., and Ghose, A. 2015. "Mobile Ad Effectiveness: Hyper-Contextual Targeting with Crowdedness,").

Avery, C., Resnick, P., and Zeckhauser, R. 1999. "The Market for Evaluations," *The American Economic Review* (89:3), pp. 564-584.

Becker, G. 1974. "A Theory of Social Interactions," *Journal of Political Economy* (82:6), pp. 1063-1093.

Benabou, R., and Tirole, J. 2006. "Incentives and Prosocial Behavior," *The American Economic Review* (96:5), pp. 1652-1678.

Berger, J. 2013. *Contagious: Why Things Catch On*. New York, NY: Simon & Schuster.

Burtch, G., Ghose, A., and Wattal, S. 2013. "An Empirical Examination of the Antecedents and Consequences of Contribution Patterns in Crowd-Funded Markets," *Information Systems Research* (24:3), pp. 499-519.

Chen, Y., Harper, F.M., Konstan, J., and Li, S.X. 2010. "Social Comparisons and Contributions to Online Communities: A Field Experiment on Movielens," *The American Economic Review* (100:4), Sep, pp. 1358-1398.

Daughety, A.F., and Reinganum, J.F. 2010. "Public Goods, Social Pressure, and the Choice between Privacy and Publicity," *The American Economic Journal: Microeconomics* (2:2), pp. 191-221.

Dickert, S., Kleber, J., Peters, E., and Slovic, P. 2011. "Numeracy as a Precursor to Pro-Social Behavior: The Impact of Numeracy and Presentation Format on the Cognitive Mechanisms Underlying Donation Decisions," *Judgment and Decision Making* (6:7), pp. 638-650.

Gneezy, U., and Rustichini, A. 2000. "A Fine Is a Price," *The Journal of Legal Studies* (29:1), pp. 1-17.

Heyman, J., and Ariely, D. 2004. "Effort for Payment: A Tale of Two Markets," *Psychological Science* (15:11), pp. 787-793.

Horrace, W.C., and Oaxaca, R.L. 2006. "Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model," *Economics Letters* (90:3), Apr, pp. 321-327.

Jiang, Z., and Benbasat, I. 2004. "Virtual Product Experience: Effects of Visual and Functional Control of Products on Perceived Diagnosticity and Flow in Electronic Shopping," *Journal of Management Information Systems* (21:3), pp. 111-147.

Jiang, Z., and Benbasat, I. 2007. "The Effects of Presentation Formats and Task Complexity on Online Consumers' Product Understanding," *MIS Quarterly* (31:3), pp. 475-500.

Jindal, N., and Liu, B. 2008. "Opinion Spam and Analysis," *Web Search and Data Mining (WSDM)*, Palo Alto, CA: ACM, pp. 219-229.

Kline, P. 2000. *The Handbook of Psychological Testing*, (2nd ed.). London: Routledge.

Krippendorff, K. 2004. "Reliability in Content Analysis," *Human Communication Research* (30:3), pp. 411-433.

Lakhani, K., and von Hippel, E. 2003. "How Open Source Software Works: “Free” User-to-User Assistance," *Research Policy* (32:6), pp. 923-943.
Lerner, J., and Tirole, J. 2002. "Some Simple Economics of Open Source," *The Journal of Industrial Economics* (50:2), pp. 197-234.
Ransbotham, S., and Kane, G.C. 2011. "Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia," *MIS Quarterly* (35:3), Jul 19, pp. 613-627.
Roberts, J.A., Hann, I., and Slaughter, S.A. 2006. "Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of Apache Projects," *Management Science* (52:7), pp. 984-999.
Stephen, A., Bart, Y., Du Plessis, C., and Goncalves, D. 2012. "Does Paying for Online Product Reviews Pay Off? The Effects of Monetary Incentives on Consumers' Product Evaluations," in: *Association for Consumer Research (ACR) Research Conference*. Vancouver, BC: ACR.
von Hippel, E., and von Krogh, G. 2003. "Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science," *Organization Science* (14:2), pp. 209-223.
Wang, J., Ghose, A., and Ipeirotis, P. 2012. "Bonus, Disclosure, and Choice: What Motivates the Creation of High-Quality Paid Reviews?," *International Conference on Information Systems (ICIS)*, Orlando, FL: AIS.
Wang, Z. 2010. "Anonymity, Social Image, and the Competition for Volunteers: A Case Study of the Online Market for Reviews," *The BE Journal of Economic Analysis & Policy* (10:1).
Zhang, M., and Zhu, F. 2011. "Group Size and Incentive to Contribute: A Natural Experiment at Chinese Wikipedia," *The American Economic Review* (101:4), pp. 1601-1615.