

Stimulating Online Reviews by Combining Financial Incentives and Social Norms

Gordon Burtch,^a Yili Hong,^b Ravi Bapna,^a Vladas Griskevicius^c

^aInformation and Decision Sciences Department, Carlson School of Management, University of Minnesota, Minneapolis, Minnesota 55455;

^bDepartment of Information Systems, W.P. Carey School of Business, Arizona State University, Tempe, Arizona 85287;

^cMarketing Department, Carlson School of Management, University of Minnesota, Minneapolis, Minnesota 55455

Contact: gburtch@umn.edu (GB); hong@asu.edu (YH); rbapna@umn.edu (RB); vladasg@umn.edu (VG)

Received: November 9, 2015

Accepted: November 1, 2016

Published Online in *Articles in Advance*:
March 3, 2017

<https://doi.org/10.1287/mnsc.2016.2715>

Copyright: © 2017 INFORMS

Abstract. In hopes of motivating consumers to provide larger volumes of useful reviews, many retailers offer financial incentives. Here, we explore an alternative approach, social norms. We inform individuals about the volume of reviews authored by peers. We test the effectiveness of using financial incentives, social norms, and a combination of both strategies in motivating consumers. In two randomized experiments, one in the field conducted in partnership with a large online clothing retailer based in China and a second on Amazon Mechanical Turk, we compare the effectiveness of each strategy in stimulating online reviews in larger numbers and of greater length. We find that financial incentives are more effective at inducing larger volumes of reviews, but the reviews that result are not particularly lengthy, whereas social norms have a greater effect on the length of reviews. Importantly, we show that the combination of financial incentives and social norms yields the greatest overall benefit by motivating reviews in greater numbers and of greater length. We further assess treatment-induced self-selection and sentiment bias by triangulating the experimental results with findings from an observational study.

History: Accepted by Chris Forman, information systems.

Funding: The first author acknowledges funding from the 3M Foundation.

Supplemental Material: The online appendix is available at <https://doi.org/10.1287/mnsc.2016.2715>.

Keywords: randomized experiment • social norms • financial incentives • online reviews • intrinsic motivation

1. Introduction

Online reviews can serve as an excellent source of information for consumers to learn about peer opinions regarding various products and services (Dellarocas 2003). Although subject to certain biases (Luca 2016), online reviews are particularly important in online markets, which are characterized by a great deal of information asymmetry (Dimoka et al. 2012). Unfortunately, like many voluntarily provided public goods (Gallus 2016), online reviews may be acutely underprovisioned (Avery et al. 1999, Anderson 1998, Levi et al. 2012). And, when consumers do provide reviews, they are often brief, limiting their helpfulness to other consumers (Cao et al. 2011, Liu et al. 2007, Mudambi and Schuff 2010). To address this problem, many retailers employ strategies intended to boost the volume and length of reviews, most commonly by offering consumers a small financial incentive in exchange for a review.

However, using financial incentives to solicit online reviews has some drawbacks. For example, a great deal of research has highlighted that offering individuals payments can undermine their intrinsic motivation to perform a task (Deci et al. 1999). This suggests that paying consumers for feedback may lead to a reduction in the effort they exert, resulting in short, uninformative

reviews. Paying for feedback may also unduly bias opinion, leading consumers to write more positive reviews (Khern-am-nuai and Kannan 2016). Furthermore, consumers tend to react negatively if they learn that a review author was paid for his or her opinion, discounting the review and inferring negative product quality (Avery et al. 1999, Stephen et al. 2012).

Given the multiple downsides of offering payment for writing reviews, here we consider the following questions: What are alternative ways to stimulate consumers to write more online reviews, and what are alternative ways to stimulate consumers to write lengthier online reviews? First, we consider the use of financial incentives to generate online reviews, a relatively common strategy among industry practitioners (Cabral and Li 2015, Fradkin et al. 2015, Stephen et al. 2012, Wang et al. 2012), wherein the retailer offers consumers a small payment in exchange for providing a review.¹ Second, we consider an alternative approach, using social norms to stimulate reviews (Chen et al. 2010), wherein we provide a consumer with information about the volume of reviews recently authored by his or her peers. For example, a social norms approach might inform consumers that in the last month, 2,345 shoppers provided an online review for a given retailer. Providing normative information

is essentially costless, and a large body of literature in psychology suggests that social norms are an effective way to encourage desirable behavior (Gerber and Rogers 2009, Ferraro and Price 2013, Allcott 2011, Goldstein et al. 2008). Finally, we test the effectiveness of using a combination of social norms and financial incentives. This influence strategy combines a small monetary payment with information about the behavior of peers into a single, joint intervention.

Our primary goal in this work is to go beyond pure financial incentives by considering alternative approaches that businesses and review aggregation platforms might employ to stimulate online reviews. By going beyond simply designing a better payment scheme, we seek to identify effective interventions that can overcome some or all of the limitations of financial incentives, which may substitute or complement existing approaches that retailers make use of today.

We employ a multimethodological research design, testing our questions via two randomized experiments and an econometric analysis of online reviews from Amazon.com, which together enable the estimation of causal effects (Aral and Walker 2011) and help to triangulate a number of interesting findings. We conduct one experiment in the field, in partnership with an online clothing retailer based in China, and a second on Amazon Mechanical Turk (AMT). In each experiment, we solicited consumers to provide online reviews by either offering a financial incentive, supplying them with normative information regarding the number of others who had written reviews, or using a combination of the two. We then assessed the effectiveness of these approaches in terms of two different outcomes: (1) *review volume*: how effective they were in motivating minimal effort by increasing the proportion of consumers who authored a review, and (2) *review length*: how effective they were in motivating more intense effort by writing reviews of greater length.

Our results suggest that financial incentives and social norms are differentially effective in motivating review volumes versus length. In both experiments, we found that financial incentives were more effective at motivating larger volumes of online reviews; individuals were more willing to exert at least minimal effort to write a review when they were promised a small payment. However, this pattern was reversed when it came to the length of reviews, where social norms were more effective. Individuals wrote longer reviews when they were informed that others had written reviews.

Perhaps most interestingly, we find in both experiments that the combined application of financial incentives and social norms delivers the greatest overall benefit, stimulating reviews both in greater volume and of greater length. The effects we observe are economically significant; the joint application of payment and a social norm in our field experiment nearly tripled

the volume of consumers who authored a review² and the reviews that resulted were approximately 50% longer.³ Our findings therefore suggest that relying on financial incentives or on social norms alone to stimulate reviews is suboptimal. To our knowledge, our research is the first to consider the effects of using both social norms and financial incentives in tandem to encourage desirable behaviors.

We also assess (i) treatment-induced sentiment bias in the reviews consumers author, as well as (ii) the relative roles of self-selection and changes in intrinsic motivation in response to our treatments to induce reviews. Regarding the former, although we observe no treatment-specific differences in consumer sentiment in our experiments, our large-scale analysis of Amazon reviews does indicate that financial incentives lead to more positive reviews in that setting. Regarding the latter, although our analyses do not enable us to draw strong conclusions, we do observe some evidence that suggests that our results are attributable to some combination of self-selection and behavioral change on the part of subjects.

This research contributes to the literature on online reviews by comparing the effectiveness of using the common approach of financial incentives to stimulate online reviews with the less common alternative of social norms. We also contribute to the broader literature on social norms by lending greater consideration to the distinction between the propensity of individuals to engage in a behavior and the intensity of engagement by a given individual. Whereas the vast majority of past work on social norms has considered behaviors that are singular in nature—that is, where the measured outcomes jointly reflect some combination of volume of participants and intensity of participation—we consider a behavior that includes two sequential decisions: agreeing to participate in the task followed by performing the task. This allows us to theorize and disentangle the effects of financial incentives and social norms on each outcome. This is important, because our findings suggest that the effects of financial incentives and social norms on each outcome are asymmetrical.

In the next section, we review prior work on the use of financial incentives and social norms to motivate behavior. We then sequentially present our two experiments, detailing the research contexts, experimental designs, and empirical results. Finally, we offer an interpretation and discussion of our findings, discuss the limitations of our work, and suggest a number of avenues for future work.

2. Hypothesis Development

2.1. Financial Incentives

One way to motivate behavior relies on offering financial incentives for desired actions. Research on this approach dates back more than 60 years (e.g., Barnes

1949, Jacques et al. 1951). Economic theory holds that rational individuals are utility driven, meaning that financial incentives should matter for individuals' behavior. A number of studies have empirically verified this prediction. For example, Volpp et al. (2009) showed that paying individuals to quit smoking increased their likelihood of doing so, and Fryer (2011) found that students could be induced to attend school more regularly with the promise of financial compensation.

Consistent with the expected impact of financial incentives, multiple studies have shown that financial incentives are effective at stimulating behavior online. For example, experimental work has found that financial incentives are effective in motivating individuals to write reviews on Airbnb.com (Fradkin et al. 2015), provide feedback on eBay (Cabral and Li 2015), and provide reviews for Best Buy (Khern-am-nuai and Kannan 2016). It is important to note that the financial incentives offered in most such studies have generally been quite small (with the exception of Fradkin et al. 2015, who offered \$25 in Airbnb credit). Cabral and Li (2015) paid rebates to subjects of just \$1 or \$2. Similarly, Khern-am-nuai and Kannan (2016) considered Best Buy's offer of 25 Best Buy reward points in exchange for each review, with a monetary value of \$0.50.

Bearing in mind that one of our experiments is situated in the context of AMT, it is useful to consider those studies that have specifically examined how financial incentives influence the supply of labor in that context. For example, Horton and Chilton (2010) asked workers on Amazon Mechanical Turk (Turkers) to click on a pair of rectangles in a specific order, offering them payment for each completed series of clicks. The authors observed that a larger volume of tasks were completed when greater payment was offered. Mason and Watts (2009) conducted a similar experiment, asking Turkers to complete "ordering" tasks, in which they were required to arrange images into a particular sequence. The authors observed that more tasks tended to be completed when pay was higher. Studies in this space have also considered relatively small financial incentives. Mason and Watts (2009) paid less than \$0.10 per task, while Horton and Chilton (2010) paid participants according to a concave function of the number of tasks completed—completing 5 tasks earned a Turker \$0.29 cents, and completing 25 tasks earned a Turker \$0.82. Taken together, studies in both AMT and other types of settings have shown that offering small financial incentives can motivate individuals to engage in a desired behavior. From this past work, we hypothesize the following.

Hypothesis 1 (H1). *Providing financial incentives leads to an increase in the volume of reviews that are provided compared with simply asking.*

2.2. Social Norms

A different way to motivate behavior relies on providing social norms. Social norms refer to the prevalence of a behavior in a relevant population, such as the number of individuals who already have written reviews. This type of social norm is known as a *descriptive social norm* (Cialdini et al. 1991). Social norms have been shown to be effective in a wide range of contexts, from motivating voter turnout (Gerber and Rogers 2009), to encouraging the reuse of hotel towels (Goldstein et al. 2008), to reducing energy consumption (Allcott 2011, Nolan et al. 2008, Schultz et al. 2007), to reducing water use (Ferraro and Price 2013), to increasing consumption of healthy foods (Robinson et al. 2014). For example, Robinson et al. (2014) tested how social norms influenced the consumption of fruit and vegetables. They exposed individuals to social norm-based messages indicating the eating behavior of others and found that those individuals ate more fruit and vegetables when they were led to believe that their peers had eaten a large amount of fruit and vegetables.

Social norms influence behavior because seeing what others have done provides information about what is socially "normal" in a given context. The greater the number of individuals who respond to the same situation in the same way, the more they will perceive the behavior to be correct (Thibaut and Kelley 1959). Individuals therefore use social normative information to determine the most appropriate course of action in a given situation (Cialdini and Trost 1998, Cialdini and Goldstein 2004).

To our knowledge, only one study has explored the use of social norms to stimulate the production of online reviews. Chen et al. (2010) conducted an experiment on a movie reviewing website, MovieLens. They informed a random set of subjects via email about the median number of reviews recently authored by their peers. They found evidence that this approach increased rates of reviewing among treated subjects, on average. Taken together, past findings give us reason to believe that social norms can have a positive influence on the production of online reviews.

Hypothesis 2 (H2). *Providing social norms leads to an increase in the volume of online reviews that are provided compared with simply asking.*

2.3. Stimulating Volume vs. Length of Reviews

The central aim of the current research is to test the effectiveness of providing financial incentives, social norms, and the combination of the two in order to stimulate a greater number of longer reviews. Note that each influence strategy could operate on two different aspects of reviewing behavior. On the one hand, we might simply seek to persuade more consumers to submit a review, increasing review volumes. However, persuading a person to write a review does not imply

that he or she will necessarily invest the effort to write a lengthy, informative review. Indeed, the vast majority of online reviews are brief and lack useful information (Cao et al. 2011, Liu et al. 2007, Mudambi and Schuff 2010). Thus, the generation of *lengthy* reviews involves motivating a second and critical aspect of behavior: after a person has decided to write a review, he or she must also be persuaded to expend additional effort to write a longer review. An effective strategy for stimulating lengthy online reviews, which are expected to be more helpful to other consumers, must motivate individuals to both choose to write a review and choose to invest effort in doing so.

The distinction between review volume, which depends simply on choosing to participate (minimal effort), and review length, which depends on intensity of effort, is important because there is reason to believe that financial incentives and social norms might operate differentially on each outcome. When individuals are provided with a financial incentive to perform a behavior, they are likely to perform that behavior for extrinsic reasons rather than intrinsic reasons (Heyman and Ariely 2004). Offering a financial incentive to write a review is likely to lead individuals to write the review because they seek to receive the financial reward rather than because of some intrinsic desire to be helpful. The presence of financial incentives therefore shifts individuals into an effort-for-payment mind-set, increasing the probability that they will provide the minimal effort that is warranted, given the level of payment (Heyman and Ariely 2004).

A large body of research indicates that offering financial incentives can change the nature of an individual's task performance by undermining their intrinsic motivation (e.g., Frey 1994, Deci et al. 1999, Jenkins et al. 1998). This suggests that paying individuals to write a review is likely to undermine their intrinsic motivation to expend much effort on writing reviews. Moreover, offering financial incentives might elicit reviews specifically from the type of individuals who lack a preexisting motivation to write reviews. Sensitivity to financial rewards is a characteristic of individuals that remains relatively stable over time (Rick et al. 2008) and is predictive of selfish behavior in social dilemmas (Seuntjens et al. 2015). Accordingly, individuals who are particularly attracted by the presence of the financial incentive may be predisposed to exert little effort in performing the task, producing short reviews. Taken together, this suggests that although the common approach of offering financial incentives might be effective at motivating many individuals to write a review, those reviews are likely to be short and relatively uninformative because the individuals writing them will expend only the minimal effort needed to obtain the reward.

Indeed, research dealing with payment for online task performance shows this exact pattern. Although multiple studies have shown that offering small financial incentives can motivate individuals to complete higher volumes of tasks (Mason and Watts 2009), offering payments does not increase the intensity of effort that individuals dedicate in any given task. For example, Wang et al. (2012) found that offering a financial incentive of the sort we consider here could induce more Turkers to write reviews, but that it had no impact on the quality of those reviews, suggesting no differences in effort intensity under payment. Stephen et al. (2012) likewise found that payment had no effect on the intensity of effort that subjects put into writing evaluations. Finally, Khern-am-nuai and Kannan (2016) observed that, even though consumers began to author larger volumes of reviews following Best Buy's introduction of redeemable reward points, the average length of reviews declined.

Because online reviews must be sufficiently lengthy to convey meaningful information, and because financial incentives appear to undermine individuals' intrinsic motivation to write longer reviews, we consider whether providing social norms might help fix this problem. Whereas the presence of a financial incentive provides individuals with an explicit extrinsic reason for why they engaged in a particular behavior ("I wrote the review to receive money"), social norms are more closely linked to intrinsic than extrinsic motivation (Henrich et al. 2006). In fact, individuals specifically do not view social norms as an extrinsic driver of their behavior (Nolan et al. 2008). Taken together, when individuals are motivated to write a review as a result of receiving normative information, they likely experience intrinsic motivation in doing so.

The provision of normative information, as with financial incentives, has the potential to induce selection effects, attracting individuals who are predisposed to exert greater levels of effort in the task. As with greed and sensitivity to money, past research has found that a tendency toward altruism and prosocial behavior is a stable trait of the individual (Brief and Motowidlo 1986). This line of reasoning indicates that those individuals who are most likely to be motivated by normative information are those individuals who might also be predisposed toward contributing to the public good. If so, such individuals might also write lengthier and more useful online reviews for the benefit of other consumers. Taken together, the provision of a social norm might result in lengthier online reviews because it may stimulate subjects' intrinsic motivation and may induce participation by individuals who are predisposed to help others. This suggests that social norms are likely to be most effective at stimulating lengthy online reviews.

Hypothesis 3 (H3). *Providing social norms leads to an increase in the length of online reviews compared with providing financial incentives or simply asking for a review.*

2.4. Simultaneously Stimulating Volume and Length of Reviews

The ultimate goal of the current research is to identify an influence strategy that motivates both larger volumes of reviews and lengthier reviews. Considering the discussion thus far, financial incentives or social norms alone may be suboptimal at achieving this goal, especially because financial incentives might undermine the motivation to exert more than minimal effort. We, therefore, consider a third approach: the combined application of social norms and financial incentives. Although no prior work has, to our knowledge, sought to combine social norms and financial incentives to motivate behavior, there is reason to believe that the combined approach might be superior to either approach alone.

We believe that the key to the effectiveness of a combined approach lies in using financial incentives to increase an individual consumer's likelihood of writing a review *without* undermining the intensity of effort that each consumer exerts when writing a review. Evidence for this possibility comes from research in child psychology, which shows that it is possible to circumvent the undermining effects of extrinsic rewards. Cialdini et al. (1998) tested how promising a child an extrinsic reward influences the child's desire to practice writing skills. Consistent with the classic undermining effect, they found that although promising a reward motivated individuals to practice writing, the reward undermined children's intrinsic motivation to practice writing when the children were no longer being rewarded for the behavior. However, the research found that, despite the promise of an extrinsic reward for participating, children's intrinsic motivation to practice writing remained high if the kids were subsequently led to believe that they were the sort of children who would want to write well. When the children could attribute their behavior to an internal reason, rather than to an extrinsic reward, they continued to be intrinsically motivated to expend effort on the task. In the same vein, Hennessey and Zbikowski (1993) reported that if children were taught to focus on their own interests as their primary reason for learning, and to treat external incentives as secondary, they were more likely to maintain intrinsic motivation.

We consider the possibility that combining financial incentives with social norms can serve to undermine the undermining effect of external rewards (Cialdini et al. 1998). We hypothesize that the presence of information about the social norm may enable individuals to rationalize their decision to write a review as one of goodwill or a personal desire to do what is appropriate, rather than as one of effort for payment. This

means that the presence of a financial incentive would serve to motivate individuals to write the review, but the presence of the social norm would provide individuals a reason to believe that they chose to write a review for some intrinsic reason rather than solely for financial gain. This leads to our final hypothesis.

Hypothesis 4 (H4). *Providing social norms and financial incentives, together, leads to the greatest volume and length of reviews, in tandem.*

3. Empirical Approach

3.1. Research Design

To test our four hypotheses, we first conduct two randomized experiments, one in the field and one on AMT. In each experiment, we compare the effectiveness of providing consumers with (1) a financial incentive, (2) a social norm, (3) a combination of a financial incentive and a social norm, and (4) simply asking them to write a review (our control). To assess effectiveness, we measured how these approaches influence our three outcomes of interest: volume of reviews, length of reviews, and a combination of volume and length.

We evaluate H1 (financial incentives lead to an increase in the volume of reviews) by comparing the volume of reviews produced in our financial incentive condition with that in the control condition, and we evaluate H2 (social norms leads to an increase in the volume of reviews) in a similar fashion, comparing the social norm condition with control. To evaluate our third hypothesis, H3 (social norms leads to an increase in the length of reviews compared to financial incentives or simply asking), we compare the length of reviews authored in our social norm condition with those authored in our control and financial incentive conditions. Finally, to evaluate H4 (the combined treatment will have the largest joint effect on review volumes and review length), we follow the approach of Burtch et al. (2015) and construct a third outcome measure, unconditional length (populating a length of 0 for those subjects who did not author a review), which jointly captures the combination of quantity and length of reviews, and thereby allows us to evaluate the total influence of our treatments on both outcomes, in tandem.

Beyond testing our main hypothesis in the first experiment, we conduct a variety of analyses to better understand these findings and to triangulate our results. In particular, we replicate our experiment, testing each hypothesis again in Study 2 via a similar combination of treatments. We then explore two additional treatments intended to help identify the specific role of changes in intrinsic motivation while eliminating the possible confounding influence of self-selection. In each new experimental condition, we reinforce the

social norm or the financial incentive only *after* a subject has agreed to author a review. Under this setup, any differences in review length that might arise could only be attributed to changes in a subject's behavior, conditional on agreeing to participate.

Finally, following the experiments, we further triangulate and clarify our findings in a number of ways. First, we hand-code additional outcome measures from the reviews obtained in Study 1 and Study 2 (namely, helpfulness and diagnosticity), which we then analyze to draw a connection between "review quality" and review length. Second, we collect and analyze a large volume of paid and unpaid online reviews from Amazon.com, by which we demonstrate that financial incentives, in particular, can impact consumers' intrinsic motivation when writing reviews. Finally, we draw on data from both experiments and our sample of archival data from Amazon to explore possible biases in consumer sentiment that may arise as a result of the treatments.

3.2. Power Analysis

To determine the number of participants needed to have sufficient power to detect our hypothesized effects in our two experiments, we conducted an a priori power analysis based on the average effect size obtained in past work most relevant to the current study (Cabral and Li 2015, Goldstein et al. 2008, Mason and Watts 2009, Nolan et al. 2008). The average effect size among these past studies is a Cohen's d of 0.47, which would lead us to require a minimum sample of 73 subjects per condition using two-tailed t -tests, maintaining a power of 0.80. If we anticipate a relatively more conservative Cohen's d of 0.30, we would require a minimum of 176 participants per condition. As we detail below, each of our experiments involved a number of subjects per condition that is well in excess of this threshold. At the same time, it is important to note that our analysis of review length is ultimately conducted only on the subset of subjects who authored a review under each treatment. As a result, a small sample of converts in any given treatment condition, in either of our experiments, may inhibit our ability to reliably detect significant effects of the different treatments on review length.

3.3. Study 1

We partnered with a large online retailer located in China that sells children's apparel via TMall, an online platform for business-to-consumer retail. TMall hosts retailers' online sales operations and also allows customers to write and post online reviews about products they purchase. TMall is owned by Alibaba and hosts large businesses, which utilize its marketplace to advertise, promote, and sell their products. Businesses on TMall can engage with customers in many

ways, such as by offering product promotions and 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.⁴ 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.⁵

3.3.1. Experiment Design and Procedure

Participants. Our participants included 2,000 customers of our retail partner, well above the threshold specified by our power analysis. Each participant was sequentially entered into the sample over a two-day period. With each sequential transaction, the associated customer was entered into our sample and randomly assigned to one of five conditions: No Message, Control, Money, Social Norms, and Money + Social Norms.⁶ Our five treatment conditions are summarized in Table 1. For reference, here we provide an English translation of the SMS message, which was confirmed by three coders fluent in both languages.

Experimental Manipulations. In our No Message condition, subjects were not contacted at all. This condition served as a baseline. In our Control condition, subjects received a generic SMS message, asking that they write a review for their recent product purchase ("Please write a review for this product"). In our Money condition, subjects were asked to write a review using the same language as in the Control condition, and they were also told that they would be paid ¥10 after doing so (approximately US\$1.50).⁷

When making the review request and offering financial compensation, we did not impose any conditions

Table 1. Study 1: Treatment Conditions

Condition	Description
No Message	No SMS was issued.
Control	"Dear Buyer, you purchased «product» from our online store on «date». Please write a review for this product."
Money	"Dear Buyer, you purchased «product» from our online store on «date». Please write a review for this product. You will receive a ¥10 coupon for use in our online store for your effort."
Social Norms	"Dear Buyer, you purchased «product» from our online store on «date». Last month, 3,786 other buyers have submitted product reviews for our store! Please write a review for this product."
Money + Social Norms	"Dear Buyer, you purchased «product» from our online store on «date». 3,786 other buyers have submitted product reviews for our store last month! Please write a review for this product. You will receive a ¥10 coupon for use in our online store for your effort."

on the content or quality of the consumer's review (e.g., a minimum length). The decision to not enforce a minimum review length was made to ensure that the design of our financial incentive closely matched with the most commonly implemented financial payment schemes that are used by retailers and review aggregation sites today. For example, *AMZ Review Trader* (now *Vipon.com*), previously enabled Amazon retailers to offer consumers a product discount in exchange for committing to provide a review after receipt of the product; *Review Trader* would encourage lengthy, thoughtful reviews, but it would not impose any conditions on compensation beyond the mere posting of a review. Many other major reviewing platforms do not enforce minimum review lengths, including Amazon and TMall, the setting for our experiment.

In our Social Norms condition, participants were asked to write a review and were informed that 3,786 different customers had written a product review for the retailer in the prior month (a true value reflecting actual reviewing volumes in the 30 days prior to our experiment). Finally, in our combined Money + Social Norms condition, participants were asked to write a review, were told that they would be compensated ¥10 after doing so, and were informed of the number of other recent customer reviewing volumes. When implementing our treatments, we used a "push" approach, delivering communications to participants via SMS messaging. We did not ask buyers to offer a good or positive review; we simply asked that they provide feedback. In addition, we considered the potential for interference to manifest in our experiment (e.g., communication or interaction between subjects in different treatment groups). We examined the geographic distribution of consumers across districts (the Chinese equivalent of a ZIP code) and observed that our 2,000 subjects were spread across 883 different districts, a wide geographic area. As such, interference is of little concern.

Dependent Variables. Two weeks after the treatment began, the retailer supplied us with information about which customers ultimately authored a review for their product purchase, as well as the textual content of each review. This allowed us to construct measures of a consumer's effort intensity and the resultant review quality. We examined three primary outcomes in our analyses: (1) participation in review authorship, measured as the quantity of reviews authored; (2) intensity of review authorship, measured as the length of the reviews; and (3) a combined measure, the product of the two, intended to capture the joint impact on quantity and quality.

The logic underlying the third, combined measure is based on recent experimental work that has dealt with a similar two-stage decision-making process. Burtch et al. (2015) considered the effect of a randomized

treatment on conversion and contribution among visitors to online crowdfunding campaigns. In that context, contribution amounts can be observed only if conversion takes place. This is similar to our setup, wherein review length can be observed only if authorship takes place. Burtch et al. (2015) evaluated the net effect of their treatment on the unconditional expected contribution (i.e., contribution per campaign visitor) by taking the product of binary conversion and continuous contribution amount. The analog in our context is to equate the lack of a review to a review with no content—that is, a review of 0 length (or minimal helpfulness and diagnosticity, in the case of our quality measures, which we discuss below). Although we recognize that providing a review valence without text is of strictly greater value than not providing a review at all, we make this simplifying assumption for the purposes of assessing the overall benefit of each treatment, as manifest in the resulting overall body of review content.

Additional Measures. In addition to evaluating our main hypotheses on the three key outcome measures, we explore the relationship between proxies for review quality and review length. We operationalize quality in terms of content coded helpfulness and diagnosticity. From 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 2007). By 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. These dimensions were manually coded for each review by two research assistants, reporting Likert scale values in each case, ranging from 1 to 7, labeled extremely unhelpful (undiagnostic) and extremely helpful (diagnostic) at the endpoints.

To ensure consistent coding, 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. 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. The coders were blind to condition, meaning that they did not know which review came from which experimental condition. We assessed

the 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. These values are well in excess of standard cut-offs for acceptable use in the literature (Kline 2000). Online Appendix B provides additional details about the coding procedure and the coding instructions.

3.3.2. Experiment Findings. Our empirical analysis begins with a model-free consideration of any average differences in our three dependent measures: the volume of reviews, the length of reviews, and the combination of volume and length. We then consider the relationship between review helpfulness, diagnosticity, our treatments, and review length. 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. The No Message group attracted 16 reviews, the Control group attracted 27 reviews, the Money group drew 73 reviews, the Social Norms group drew 41 reviews, and the Money + Social Norms group drew 70 reviews.

Figures 1(a)–1(c) graphically depict the differences across conditions in average review volumes (Figure 1(a)) and review length (Figure 1(b)); note that our y axis reflects the number of Chinese characters: one Chinese character roughly translates to one English word). We tested H1 and H2 using pairwise comparisons of group means. Hypothesis 1 predicted that offering financial incentives should lead to an increase in the volume of reviews that are provided compared to simply asking consumer to provide reviews. Consistent with H1, we observe that participants in the Financial Incentives condition were more likely to author a review than the Control condition ($p < 0.001$).

We next tested H2, which predicted that providing social norms should lead to an increase in the volume of reviews that are provided compared to simply asking consumer to provide reviews. We found only

Table 2. Study 1: Descriptive Statistics

Variable	Mean	Std. dev.	Min	Max	N
Authorship	0.114	0.317	0.000	1.000	2,000
Length	13.410	8.051	1.000	42.000	227 ^a
log(Length)	2.376	0.730	0.000	3.738	227 ^a
Perceived Helpfulness	2.901	1.461	1.000	6.500	227 ^a
Review Diagnosticity	2.892	1.362	1.000	7.000	227 ^a

^aOf 2,000 subjects, 227 subjects wrote a review; this value reflects only authored reviews.

Figure 1(a). Study 1: Percentage of Subjects Who Wrote a Review in Each Condition

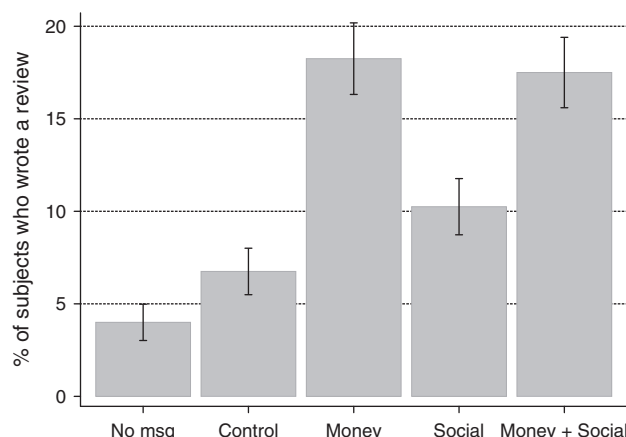


Figure 1(b). Study 1: Average Length of Reviews Written in Each Condition (Number of Chinese Characters)

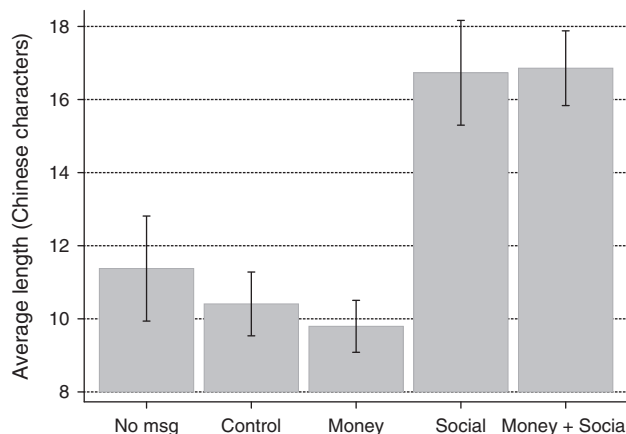
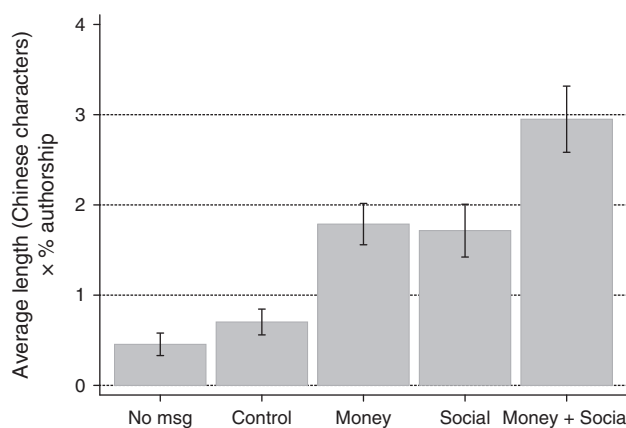


Figure 1(c). Study 1: The Combined Effect on Authorship and Length Across Conditions (i.e., Length = 0 If No Review Was Written)



marginal support for H2, that participants in the Social Norms condition were more likely to author a review than the Control condition ($p = 0.076$). In addition to testing H1 and H2, we also observed that participants

in the Social Norms group were also more likely to author a review than in the No Message group ($p < 0.001$), that the Money group was more likely to author a review than the Social Norms group ($p = 0.001$), and that the Money and Money + Social Norms groups appear roughly equivalent in their likelihood of authoring a review ($p = 0.782$).

Next, we consider the average length of reviews to test H3, which stated that providing social norms should lead to an increase in length of online reviews compared with providing financial incentives or simply asking consumers to write a review. Consistent with H3, participants in the Social Norms condition wrote reviews that were longer than those in Control condition ($p = 0.001$) or those in the Financial Incentives condition ($p < 0.001$). In addition to testing H3, we also observed that the length of reviews in the Control condition were roughly similar to those in the No Message group ($p = 0.544$) and the Money group ($p = 0.634$). Moreover, the length of reviews in the Money + Social Norms group authors was longer than in the Control group ($p < 0.001$), but there was no discernible difference in the length of reviews in the Social Norms condition and the Money + Social Norms condition ($p = 0.942$).

Finally, we test H4 by considering the net effect of our treatments on volume and length, which stated providing social norms and financial incentives, together, should lead to the greatest combination of volume and length of reviews. To measure the “net” effect, we constructed a combined measure, unconditional length, where we assign a 0 value to length whenever a review was not authored (see Figure 1(c)). Consistent with H4, we find that the Money + Social Norms group drives total review output in excess of the Social Norms group ($p = 0.009$), the Money group ($p = 0.007$), and the Control group ($p < 0.001$). Thus, we find support for H4. In addition to testing H4, we observed that the Social Norms and the Money groups each had a greater combined effect than the Control group ($p = 0.001$ and $p = 0.002$, respectively) and that Social Norms and the Money were similar in their combined effect ($p = 0.845$). Finally, we observe that the combined output of the Control group and the No Message group were also similar ($p = 0.192$). Here, we should note that if we were to consider conventional thresholds for statistical significance—for example, an α of 0.05—applying a Bonferroni correction to account for multiple comparisons would have little impact on our hypothesis tests; H1, H3, and H4 would continue to be strongly supported, while support for H2 would be weak, at best.⁸

To obtain more efficient estimates of the treatment effect, we next consider econometric estimations. 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 relate our binary outcome (review) to

dummy indicators of each of our treatment conditions (Equation (1)) using a linear probability model (LPM). Subsequently, we employ ordinary least squares (OLS) to analyze the relationship between review length and treatment (Equation (2)). As noted above, we also report subsequent regression analyses of our review quality measures (helpfulness and diagnosticity) on our treatments and our effort measure, length, in an attempt to validate the connection between downstream benefits to other consumers that result from increased rates of reviewing and greater intensity of effort exerted by reviewers (Equations (3a) and (3b)). These latter regressions are estimated via ordered logit regression. Here, subjects are indexed by i , and our various treatments are indexed by p :

$$Authorship_i = \alpha + \sum_p Treatment_i^p + \varepsilon_i, \quad (1)$$

$$\log(Length_i) = \alpha + \sum_p Treatment_i^p + \varepsilon_i, \quad (2)$$

$$Helpfulness_i = \alpha + \sum_p Treatment_i^p + \log(Length_i) + \varepsilon_i, \quad (3a)$$

$$Diagnosticity_i = \alpha + \sum_p Treatment_i^p + \log(Length_i) + \varepsilon_i. \quad (3b)$$

The results of our authorship and length regressions align with the findings reported in our model-free descriptive analyses and graphical depictions with pairwise comparisons of group means (see the first and second columns of Table 3).

In this regression analysis, we also examine the relationship between our treatments, review length, and our measures of review helpfulness and diagnosticity. We first estimate a pair of ordered logit regressions, taking our coded Likert-scale measures of helpfulness and diagnosticity as dependent variables (fourth and fifth columns). We then repeat the process in an unconditional manner, assigning values of 1 to helpfulness and diagnosticity in those cases in which no review was supplied (sixth and seventh columns). In each case, we find that quality is primarily associated with review length. In summary, Study 1 provides support for each of our four hypotheses. First, individuals were more likely to write a review either when they were promised a small payment or when they were informed of the social norm. The regression results indicate once again that financial incentives (H1) can be very effective in stimulating larger volumes of reviews, while social norms (H2) may be effective.

It should be kept in mind that, although we observe that financial incentives were more effective than social norms at eliciting review volumes in our particular experiment, we would be hesitant to conclude that financial incentives are generally more effective

Table 3. Study 1: Regression Results

Explanatory variable	Authorship	Conditional log(Length)	Unconditional log(Length)	Conditional Helpfulness	Conditional Diagnosticity	Unconditional Helpfulness	Unconditional Diagnosticity
No Message	−0.028 ⁺ (0.016)	−0.009 (0.201)	−0.064 ⁺ (0.038)	0.265 (0.605)	0.398 (0.585)	0.238 (0.602)	0.308 (0.556)
Money (M)	0.115 ^{***} (0.023)	−0.184 (0.122)	0.245 ^{***} (0.054)	0.326 (0.368)	0.316 (0.356)	0.320 (0.371)	0.349 (0.363)
Social (S)	0.035 ⁺ (0.020)	0.389 ^{**} (0.137)	0.120 [*] (0.052)	0.056 (0.407)	0.752 ⁺ (0.448)	−0.016 (0.410)	0.598 (0.455)
M + S	0.108 ^{***} (0.023)	0.382 ^{**} (0.123)	0.317 ^{***} (0.062)	0.298 (0.356)	0.488 (0.367)	0.233 (0.359)	0.357 (0.376)
log(Length)	—	—	—	4.135 ^{***} (0.355)	3.063 ^{***} (0.273)	4.717 ^{***} (0.311)	3.921 ^{***} (0.187)
Constant	0.068 ^{***} (0.010)	2.248 ^{***} (0.087)	0.159 ^{***} (0.030)	—	—	—	—
Observations	2,000	227	2,000	227	227	2,000	2,000
F-stat.	18.03 (4, 1995)	7.73 (4, 222)	18.00 (4, 1995)	—	—	—	—
Wald χ^2	—	—	—	151.50 (5)	143.45 (5)	243.06 (5)	466.16 (5)
R-squared	0.032	0.126	0.031	0.255	0.205	0.628	0.613

Notes. Robust standard errors are provided in parentheses. Regressions based on raw length exhibit the same pattern of results in terms of significance and magnitude.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.10$.

because the relative strength of each treatment will depend on the amount of money offered, or the strength of the social norm.

Second, participants wrote longer and more useful reviews when they were informed of the social norm compared with when they were promised a small payment or received a generic request to author a review. This finding suggests that social norms are effective at stimulating lengthier reviews (H3).

Third, providing a combination of financial incentives and social norms was most effective at motivating participants to write high volumes of lengthy online reviews. This last finding suggests that combining financial incentives and social norms delivers the greatest overall benefit because it jointly stimulates greater review volumes and review lengths (H4).

3.4. Study 2

The second experiment sought to replicate the findings of Study 1 in a different context (to evaluate generalizability across cultures and to a nonpurchase setting). Study 2 was also designed to explore the mechanism underlying the observed effects (e.g., self-selection versus changes in behavior). To do so, we incorporated additional treatment conditions into Study 2, wherein we initially supply the combined Money and Social Norms treatment, but we then reinforce either the financial incentive or the social norm after the person has agreed to supply feedback. By comparing the relative efficacy of these two new reinforcement conditions, wherein subjects are retreated after selection has already taken place, we can explore the degree to which the effects we observed are driven by changes in individuals' level of intrinsic motivation versus self-selection.

Were we to observe no significant differences between the two reinforcement conditions, or between those conditions and the baseline, we might conclude

that our main effects are driven primarily by individuals selecting into each treatment, who are predisposed toward exerting greater or lesser effort and thus authoring lengthier or shorter reviews. Conversely, if we were to observe that reinforcement of the social norm does result in lengthier reviews, this suggests that individuals' behavior is modified by the treatment. In each of the new "reinforcement" conditions, subjects were exposed to a second reinforcement message immediately after they agreed to author feedback.

The reinforcement message reminded participants either of the financial benefit of writing the review (As compensation for completing this survey, you will receive \$0.04) or of the social norm ("You are now the 257th Turker to provide us with feedback"). Thus, all participants in the two new conditions were first provided with the combination of financial incentives and social norms when they were choosing whether to write a review; however, after making the decision to write the review, participants were subsequently reminded of either the social norm or the financial incentive. We expected that the two new conditions (combined + social norm reminder and combined + financial incentive reminder) would produce a similar effect on review volumes as the original combined condition, without any reminder, given that the reinforcement message would not be delivered until after a subject had agreed to write a review. However, we explored whether the two new conditions might have a different effect on review lengths. In particular, we examined whether reminding participants of the financial reward could undermine the intensity of effort, leading to shorter reviews. At the same time, we also considered whether reinforcing the social norm might have a different effect such as, for example, potentially increasing review lengths over and above the baseline combined condition.

3.4.1. Experiment Design and Procedure. In Study 2, we recruited 1,200 Turkers to respond to a survey about

the demographics and work behavior of workers on AMT (Ipeirotis 2010). Additional screening questions were embedded in the survey to identify subjects who were not paying sufficient attention (e.g., “what month is it?”). This resulted in the exclusion of seven subjects. Following a 36-hour delay after completion of the survey, we invited the subjects (via email, using the AMT Requester API) to provide an overall rating of the quality of the demographics survey, in the form of a seven-point scale response, as well as any comments, suggestions or feedback (text).

Subjects were randomized into one of six groups: Control, Money, Social Norms, Money + Social Norms, *Money + Social Norms with Money Reinforced*, and *Money + Social Norms with Social Norms Reinforced*. (Note that the italicized treatments are the new treatment conditions that we introduced, to evaluate the mechanism underlying the combined treatment effect and to assess whether the treatment effects operate via changes in intrinsic motivation or self-selection.) The first four groups were equivalent to those employed in Study 1. We maintained only one control group in this instance because organic (unprompted) feedback was not possible; some form of survey invitation was required. The presence of a financial incentive or a social norm was communicated in the email itself, in both the subject line and the body. Table A2 in Online Appendix A reports randomization (balance) tests for a set of self-reported, subject-level demographic covariates obtained from the initial recruitment (demographic) survey, where we observe no significant differences.

If a subject was assigned to the Money group, the email subject line read “Receive \$0.04 for providing feedback about our survey!” By contrast, if a subject was assigned to the Social Norms group, the subject line read “Join the 256 Turkers who have provided us with feedback about our survey!” In the combined conditions, the email subject line mentioned both treatments: “Receive \$0.04 and join the 256 Turkers who have provided us with feedback about our survey!” In the various financial incentive conditions, the email body contained a hyperlink to AMT, prepopulated with task search parameters, thereby directly navigating the subject to group-specific human intelligence task (HIT) on AMT, which in turn linked to a follow-up survey. Subjects were assigned a unique qualification on AMT prior to running the experiment, to ensure no other individuals could stumble on the HIT or survey by accident. In the unpaid conditions, the email body contained a hyperlink directly to the follow-up survey.

In each of the new treatment conditions, subjects received an additional message after navigating to the follow-up survey, reinforcing either the financial incentive or the social norm. Figures 2(a) and 2(b) provide screenshots of the reinforcement treatments. Table 4 provides a summary of our treatment conditions.

Follow-up survey responses were collected over the next 24 hours. We again measured the volume of subjects in each group that provided a follow-up response and the textual length of feedback each subject provided. Moreover, we once again hand-coded the perceived helpfulness and diagnosticity of textual feedback. Finally, we again consider unconditional

Figure 2(a). Reinforcement of Financial Incentive

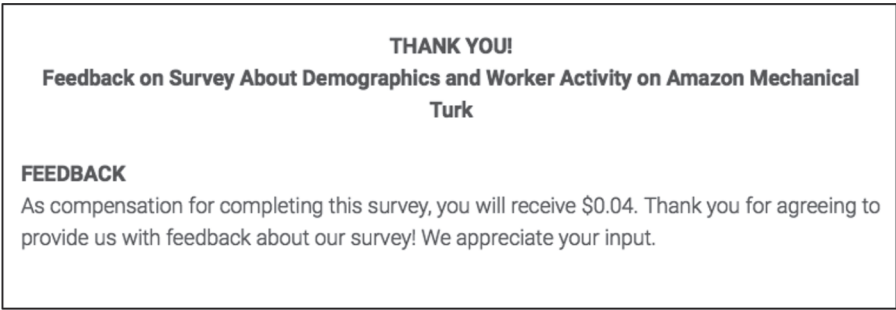


Figure 2(b). Reinforcement of Social Norm

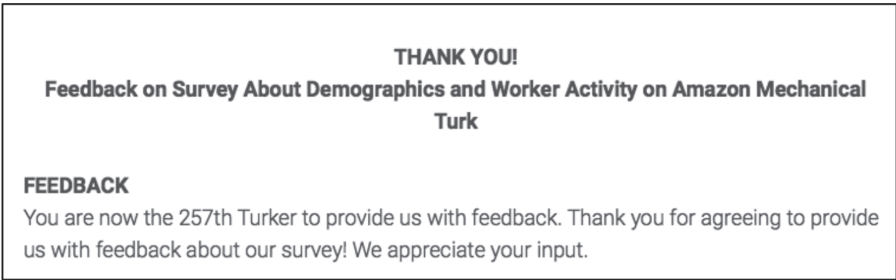


Table 4. Study 2: Treatment Conditions

Condition	Description
Control	A generic email was issued soliciting survey feedback.
Money	An email was issued soliciting feedback, offering \$0.04 as compensation, in the form of an AMT task.
Social Norms	An email was issued soliciting feedback, noting that 256 other Turkers had already provided such feedback.
Money + Social Norms	An email was issued soliciting feedback, offering \$0.04 as compensation, in the form of an AMT task, and noting that 256 other Turkers had already provided such feedback.
Money + Social Norms + Money Reinforced	An email was issued soliciting feedback, offering \$0.04 as compensation, in the form of an AMT task, and noting that 256 other Turkers had already provided such feedback. Upon arriving at the follow-up survey, the financial incentive was reinforced.
Money + Social Norms + Social Norms Reinforced	An email was issued soliciting feedback, offering \$0.04 as compensation, in the form of an AMT task, and noting that 256 other Turkers had already provided such feedback. Upon arriving at the follow-up survey, the Social Norms treatment was reinforced.

measures of length, helpfulness and diagnosticity as well, similar to Study 1, substituting a length of 0, a helpfulness of 1, and a diagnosticity of 1 when subjects did not provide any feedback. We defined diagnosticity and helpfulness in a manner analogous to Study 1 (coding details are provided in Online Appendix B).

3.4.2. Experiment Findings. Figures 3(a), 3(b), and 3(c) plot group means and standard errors, in terms of proportion of subjects providing feedback, length of text provided, and unconditional length, respectively. The Control condition, in which we simply asked for feedback, attracted 66 responses; the Money condition attracted 95 responses; the Social Norms condition attracted 69 responses; the Money + Social Norms condition attracted 95 responses; the Money + Social Norms + Money condition attracted 88 responses; and the Money + Social Norms + Social Norms condition attracted 95 responses. Table 5 presents the descriptive statistics for Study 2's sample.

As seen in Figure 3(a), we again observe that the financial incentive is effective in driving participation, the exertion of at least minimal effort (supporting H1). Whereas textual feedback is supplied approximately 30% of the time in control, in the various paid treatments, it reaches nearly 50% (Money *versus* Control: $p = 0.003$, Money + Social Norms *versus* Control: $p = 0.002$, Money + Social Norms + Money Reinforced *versus* Control: $p = 0.023$, Money + Social Norms + Social Norms Reinforced *versus* Control: $p = 0.003$). However, in contrast to Study 1, we observe no discernible differences in the volume of subjects supplying feedback between the Control and Social Norms treatment ($p = 0.832$). Thus, we do not observe support for H2. We once again observe that the presence of the social norm in tandem with the financial incentive does not appear to change feedback volumes relative to offering money by itself ($p = 0.962$).

Figure 3(b) reflects a similar pattern to that observed in Study 1 (i.e., Figure 1(b)). We do not find that simply paying for feedback produces reviews of discernibly greater length ($p = 0.274$). By contrast, treating subjects

with the social norm appears to raise the length of feedback they provide, relative to the common approaches of offering money ($p = 0.005$) or simply asking (i.e., our control) ($p = 0.090$), once again providing support for H3. Combining social norms and financial incentives, we observe that the effect of social norms on feedback length remains stable; that is, we observe no discernible differences between the Social Norms group and the Money + Social Norms group ($p = 0.891$).

Of particular interest, however, are the observed differences between the Money + Social Norms condition, and the two new conditions that involve reinforcement messages. Whereas no stark differences manifest around the number of subjects providing feedback (which is expected, given that reinforcement takes place after the decision to author a review), compared with offering money on its own, offering both money and normative information produced longer feedback than our control ($p = 0.005$), as did offering the combined treatment along with a reinforcement of the social norm ($p = 0.015$). However, when the combined treatment was offered along with a reinforcement of the financial incentive, those differences fade ($p = 0.146$). At the same time, a comparison between the three combined conditions does not indicate discernible differences (Money + Social Norms *versus* Money + Social Norms + Money: $p = 0.260$, Money + Social Norms *versus* Money + Social Norms + Social Norms: $p = 0.921$, Money + Social Norms + Money *versus* Money + Social Norms + Social Norms: $p = 0.221$).

Finally, considering Figure 3(c), which presents group averages for the combined measure (unconditional length), we observe a pattern similar to that observed in Study 1. The combined treatment outperforms the Financial Incentive condition ($p = 0.023$), as well as the Social Norms condition ($p = 0.105$), once again providing evidence in support of H4. Moreover, when the payment is reinforced, the impact on review lengths from the combined treatment fades, such that it is not discernibly different from the Money condition ($p = 0.412$), whereas reinforcing the social norm causes the increase in effort (length) to persist ($p = 0.039$).

Figure 3(a). Study 2: Percentage of Subjects Who Wrote a Review in Each Condition

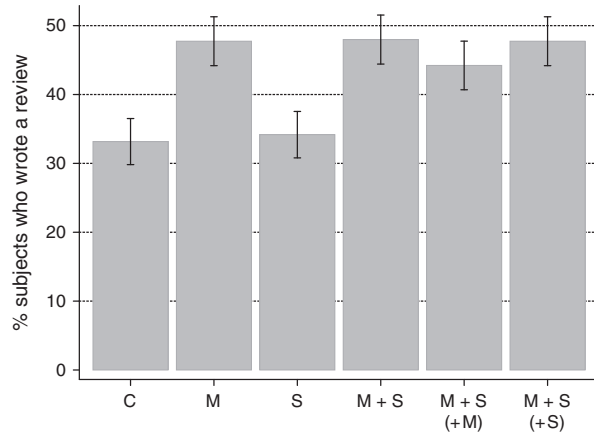


Figure 3(b). Study 2: Average Length of Reviews in Each Condition (English Characters)

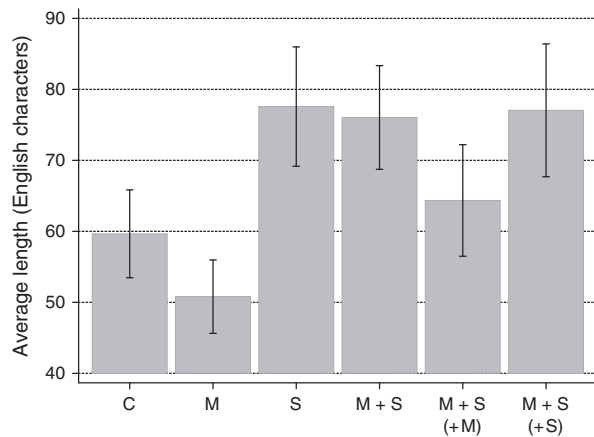
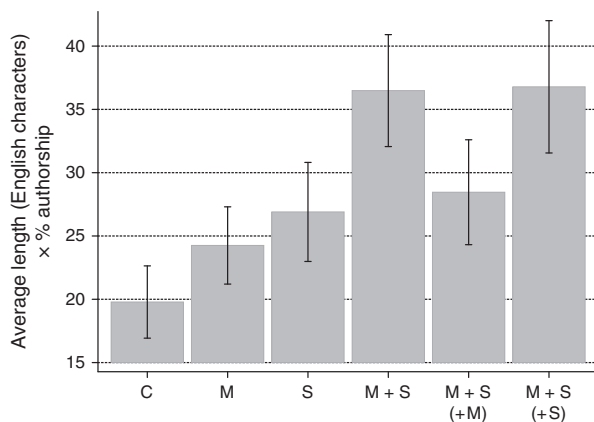


Figure 3(c). Study 2: Joint Effect on Review Quantity and Length Across Conditions (Length = 0 If No Review Was Written)



Moreover, when we compare the two reinforcement conditions with one another, we do observe relatively clear differences ($p = 0.055$). At the same time, if we compare the financial incentive reinforcement condition with the baseline combined condition, the differences

Table 5. Study 2: Descriptive Statistics

Variable	Mean	Std. dev.	Min	Max	N
Authorship	0.425	0.495	0.000	1.000	1,193
Length	67.565	70.492	0.000	776.000	508 ^a
log(Length)	3.838	1.021	0.000	6.655	508 ^a
Perceived Helpfulness	2.134	0.990	1.000	6.667	508 ^a
Feedback Diagnosticity	2.151	0.947	1.000	7.000	508 ^a

^aOf 1,193 subjects, 508 wrote a review; this value reflects only authored reviews.

in review lengths are admittedly less apparent ($p = 0.221$). Furthermore, the social norm reinforcement condition does not appear to increase review lengths over and above the baseline combined condition, as we might have expected ($p = 0.921$).

Thus, taken together, the findings in Study 2 offer some evidence in support of our expectations. It appears that the benefits of combining social and financial incentives in this context derive largely from providing subjects with a plausible rationalization for their behavior; it appears that they are choosing to write the review for reasons other than receiving a financial incentive. This, in turn, results in more intensive provision of feedback (i.e., greater length). At the same time, our results are by no means clear-cut, and some between-group differences we might have expected to observe ultimately failed to manifest.

We again conducted a formal econometric analysis of these relationships, reported in Table 6. Again, the results of our authorship and length regressions align with the findings reported in our descriptive analyses and graphical depictions with pairwise comparisons of group means (see the first and second columns of Table 6). Similar to Study 1, we also consider the net effect of our treatment conditions on review output (unconditional length). The results of this regression are reported in the third column. We once again observe positive significant effects in each of the Money + Social Norms conditions relative to control. Moreover, we find that, compared with the Money + Social Norms condition ($p < 0.001$), reinforcing financial incentives attenuates the magnitude of the differences in review output, relative to control ($p = 0.018$), whereas reinforcing the social norms does not appear to attenuate the effects ($p < 0.001$). Thus, our regression results are broadly consistent with the graphical comparison of group means noted earlier. However, we also acknowledge that if we make a direct comparison of the coefficients associated with our two reinforcement conditions in the unconditional regression, the differences are rather weak ($F(1, 1187) = 1.55$, $p = 0.214$).

Finally, we again assessed the downstream impact of our treatments and length on the helpfulness and diagnosticity of textual feedback. These results are reported in the fourth and fifth columns. We also considered

Table 6. Study 2: Regression Results

Explanatory variable	Authorship	Conditional log(Length)		Unconditional log(Length)		Conditional Helpfulness		Conditional Diagnosticity		Unconditional Helpfulness		Unconditional Diagnosticity	
<i>Money (M)</i>	0.146** (0.049)	−0.125	(0.193)	0.473* (0.190)		0.297 (0.348)		0.547 (0.333)		0.381 (0.346)		0.652 ⁺ (0.334)	
<i>Social (S)</i>	0.010 (0.047)	0.423* (0.184)		0.202 (0.194)		0.329 (0.367)		0.595 ⁺ (0.357)		0.310 (0.364)		0.570 (0.358)	
<i>M + S</i>	0.148** (0.049)	0.330 ⁺ (0.185)		0.699*** (0.200)		0.373 (0.361)		0.772* (0.337)		0.377 (0.362)		0.768* (0.340)	
<i>M + S(+M)</i>	0.111* (0.049)	0.118 (0.184)		0.456* (0.193)		0.015 (0.334)		0.482 (0.325)		0.071 (0.335)		0.562 ⁺ (0.328)	
<i>M + S(+S)</i>	0.146** (0.049)	0.373* (0.177)		0.710*** (0.199)		0.291 (0.328)		0.796* (0.337)		0.281 (0.329)		0.800* (0.338)	
log(Length)	—	—		—		1.342*** (0.145)		1.183*** (0.148)		1.645*** (0.086)		1.555*** (0.082)	
Constant	0.332*** (0.033)	3.652*** (0.154)		1.211*** (0.133)		—		—		—		—	
Observations	1,193	508		1,193		508		508		1,193		1,193	
F-stat.	4.16*** (5, 1187)	4.06** (5, 502)		3.97 (5, 1187)		98.96 (6)		74.81 (6)		393.18 (6)		364.08 (6)	
R-squared	0.017	0.040		0.016		0.077		0.071		0.345		0.351	

Notes. Robust standard errors are provided in parentheses. Regressions, based on raw length, exhibit the same pattern of results in terms of significance and magnitude.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.10$.

an unconditional analysis, wherein helpfulness and diagnosticity were coded as values of 1 when no feedback was provided, in the sixth and seventh columns. Once again, we see that quality measures are primarily driven by our proxy for the intensity of effort that a subject expends, length. One notable difference here is that we observe significant effects on diagnosticity from a number of treatments that involve the social norm. We interpret these results as an indication that effort may manifest in ways other than review length, such as more careful word choice or greater clarity of writing.

3.5. Study 3

Study 2 provides evidence that our treatments operate at least in part by changing subjects' intrinsic motivation, in that the reinforcement of financial incentives appeared to weaken any benefits of the combined condition. However, that evidence is by no means clear. Moreover, other open questions remain about the possible collateral effects of our treatments in inducing biased sentiments. To gain further clarity on these two issues, we performed a third, archival study, based on a large sample of online reviews collected from Amazon.com, wherein we were able to reliably identify whether or not the retailer had provided a financial incentive to the consumer. Online Appendix C provides details about the data collection process and sample characteristics.

We estimate a linear three-way fixed effects model (product, reviewer, and time) on a set of 90,764 reviews for 839 products, authored by 57,469 individuals. We find that the reviews are approximately 6% shorter when a discount was provided by the retailer (the first column of Table 7). This result is consistent with the findings of Khern-am-nuai and Kannan (2016). More to the point, the result provides clearer evidence of the negative impacts that financial incentives may have on consumers' intrinsic motivation to write reviews.

Additionally, this finding further supports our earlier conclusion, from Study 2, that the effects of our treatments are driven, at least in part, by changes in subjects' behavior (e.g., intrinsic motivation), and not just via treatment-induced self-selection.

Finally, we consider the question of treatment-induced bias in sentiment, a subject that has received significant attention in the literature, though with inconclusive results. For example, in the experiments by Wang et al. (2012) and Stephen et al. (2012), no apparent differences in review valence were found as a result of payment. By contrast, Khern-am-nuai and Kannan (2016) found that the text of reviews began to contain more positive words and that the average star valence increased, following Best Buy's introduction of reward points for writing reviews.

From an analysis of hand-coded sentiment in the reviews obtained from our two experiments, we also found no evidence of a treatment-induced sentiment bias. However, we also considered that these null results (and those of prior studies) may have been a result of small sample sizes and thus a lack of study power. Indeed, examining the effect of financial incentives on review valence in our much larger sample of Amazon data, we find that paid reviews are positively biased, being 0.031 stars higher ($p < 0.001$), on average.

Table 7. Regression Results (Amazon Review Trader)

Explanatory variable	log(Length)	Positivity
<i>Paid</i>	−0.06*** (0.012)	0.031** (0.010)
<i>Product Effects</i>	Yes	Yes
<i>Reviewer Effects</i>	Yes	Yes
<i>Year-Month Effects</i>	Yes	Yes
<i>Product-Specific Trends</i>	Yes	Yes
Observations	90,764	90,764
F-stat.	419.62 (730, 57,468)	145.68 (730, 57,468)
Within R ²	0.101	0.083

Note. Robust standard errors are provided in parentheses.

*** $p < 0.001$; ** $p < 0.01$.

Online Appendix C reports more details of these analyses, along with a more elaborate discussion.

4. General Discussion

We tested the effectiveness of using financial incentives, social norms, and a combination of both strategies on motivating individuals to write online reviews. In two randomized experiments, one in the field conducted in partnership with a large online clothing retailer based in China and a second on Amazon Mechanical Turk, we compared the effectiveness of each strategy at stimulating online reviews in larger numbers and of greater length. We find that financial incentives are more effective at inducing larger volumes of reviews than a simple request (consistent with H1), and it appears that social norms may be as well, to some degree (i.e., we observed partial support for H2). When it comes to review length, we find that social norms outperform both financial incentives and a simple request (consistent with H3). Finally, we show that the combination of financial incentives and social norms yields the greatest overall benefit, motivating reviews in greater numbers and of greater length (consistent with H4).

4.1. Practical Implications

Many businesses offer financial incentives to motivate consumers to write reviews. However, using such an approach to solicit reviews appears to present some problems, as discussed earlier. This research suggests that it may be optimal for firms to use financial incentives in tandem with a social norm, in order to minimize the possibility of receiving short reviews. Our results also suggest a more advanced strategy, which may be appropriate for retailers who are launching a new product offering, where little to no reviews have previously been authored by consumers (and thus where advertising a descriptive norm might not be possible). In particular, our findings suggest that firms might initially employ financial incentives to seed early (albeit potentially short) reviews and then quickly transition to sustainable (lengthier) contributions by exploiting a social norm, ultimately transitioning to higher-quality contributions. Of course, this strategy would need to be implemented with caution, because the longer-term effects of social norm treatments remain unclear. In terms of policy, our analyses (particularly those of Amazon.com reviews, reported in Online Appendix C), suggest that the current policies imposed by federal regulators (e.g., the Federal Trade Commission) and major online platforms (e.g., Amazon) are justified in their view of paid reviews as a possible form of false advertising. Our analyses indicate that paid reviews are systematically more positive than organic reviews, suggesting a bias of reciprocity. Our analyses also suggest that online retailers may be

well served to limit or avoid paying for reviews for reasons beyond simply adhering to regulators and platforms policies, given our finding that paid reviews are systematically shorter in length and, as a result, perhaps of lower quality.

4.2. Theoretical Contributions

While prior research has considered the respective effects of financial incentives (Khernamnuai and Kannan 2016, Stephen et al. 2012, Wang et al. 2012) and social norms (Gerber and Rogers 2009, Ferraro and Price 2013, Allcott 2011) in isolation, we offer a first consideration of the relative and joint effects of financial incentives and social norms on motivating prosocial behavior, and we find that combining financial incentives with social norms results in the greatest overall effect. Furthermore, our work seeks to disentangle the effect of social norms on behavior, distinguishing between the breadth and depth of engagement in an activity (participation and intensity). Although this distinction has been considered in the financial incentives literature, which has observed in many cases that financial incentives motivate individuals to do the minimum required to earn pay, no prior work to our knowledge has explored this distinction in the use of social norms to motivate behavior. We observe an asymmetrical effect, in that financial incentives are more effective at motivating participation (write a review) than intensive effort (write a lengthy review), whereas social norms are more effective at motivating intensive effort than they are at motivating participation. Finally, our work contributes to the emerging literature on design interventions that can lead to the production of user-generated content (Chen et al. 2010, Jabr et al. 2014, Goes et al. 2016).

4.3. Limitations and Future Research

Our findings around the individual effects of social norms and financial incentives are broadly consistent with the prior literature, in that both appear at least somewhat effective at motivating participation. However, we find that not all participation is equal. Some participants exert less effort than others, whether because of their inherent characteristics or because the treatments we impose cause changes in their behavior. In particular, our three studies provide evidence that, whereas social norms can drive a high level of effort, as manifest in longer reviews, financial incentives may lead to just the opposite, eliciting shorter reviews. Though the latter finding has not been observed in past experimental work (Stephen et al. 2012, Wang et al. 2012), this difference likely arises because our observational analysis benefits from a large-scale sample of more than 90,000 reviews. Most interestingly, however, our work is the first, to our knowledge, to consider and demonstrate that the joint application of financial

incentives and social norms can produce the greatest overall benefit, leading jointly to the greatest volume and length of reviews. However, our findings are subject to a number of limitations, which present opportunities for future work.

First, the pattern of effects that we observe across both experiments and the archival study is consistent with the idea that providing financial incentives can undermine intrinsic motivation, placing individuals in an effort-for-payment mind-set (Heyman and Ariely 2004). Therefore, we surmise that providing both financial incentives and normative information can circumvent the undermining effect by providing individuals with a plausible rationalization for their decision to act as one of goodwill, rather than as one of effort for payment. This, in turn, allows intrinsic motivation to persist. However, it is possible (even likely) that our results are also driven in part by subjects' self-selection into the receipt of each treatment. More work is therefore needed to better understand the relative roles of intrinsic motivation and self-selection in driving these outcomes. A replication of the reinforcement conditions from Study 2, with a larger sample, might enable this.

Second, it is also worth noting that we are incapable of comparing the relative effects of financial incentives and social norms on review volumes, because these results are quite likely to depend on the exact level of each treatment (e.g., amount of money offered, the strength of the norm, how such norms are perceived in different contexts and culture). Future work might explore varied levels of each treatment to understand how the effects vary. To this point, the effects of financial incentives are quite likely to be nonlinear in nature. Cabral and Li (2015) observed that while a \$1 rebate had a borderline effect on a consumer's willingness to provide eBay feedback, increasing the amount to \$2 produced a stronger effect. Similarly, the effects of social norms are also likely to vary nonlinearly in the level of activity among an individual's peers. Moreover, our results may also be contingent on the actual level of a subject's own reviewing activity at the time of the experiment. Notably, we have only manipulated the provision of normative information to subjects; we have not manipulated the information itself. This, too, warrants further exploration.

Finally, it would be useful to understand the dynamic nature of the observed effects—for example, whether they continue to manifest for a given subject with repeatedly treatments over time, or whether subjects become desensitized. Additionally, it may be possible to improve on our results by targeting these treatments toward individuals who are most likely to respond in a positive, desirable manner (e.g., can we deliver the financial incentive to only those individuals that exhibit no evidence of a sentiment bias?).

Acknowledgments

The authors are grateful to seminar participants at the 2015 Conference on Information Systems and Technology, the 2015 International Conference on Information Systems, the 2015 Conference on Digital Experiments (CODE), and the University of Minnesota Carlson School's Summer Applied Economics Workshop, as well as seminar participants at Temple University, HEC Paris, Michigan State University, and McGill University for their useful suggestions. The authors also thank Pedro Ferreira and Evan Polman for providing helpful comments.

Endnotes

¹ When we refer to the use of financial incentives in this paper, we mean the solicitation of truthful consumer evaluations in exchange for payment, regardless of consumer sentiment. That is, we are not referring to payment in exchange for strictly positive reviews or fake reviews.

² Compared with a baseline proportion of 4% in the control condition in which no intervention was used.

³ Compared with a baseline average length of ~52 English characters, or 10 words, in the control condition.

⁴ The merchant and product review systems are owned and maintained by TMall, not the retailers. With respect to valence, only aggregate ratings are displayed on the TMall website and provided to the individual retailers; that is, the valence of individual customer ratings is not observed. However, the platform does provide retailers (and us, by extension) with the review text. We therefore initially focus on the textual content in the analysis of our first study.

⁵ Using SMS messaging to communicate with customers has many advantages over email. Because cellular numbers are recorded as part of a buyer's shipping 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. Moreover, 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.

⁶ Customers were excluded from consideration if they had already entered the sample as a result of a prior transaction. Because of our randomization procedure, the offer of a financial incentive and the dissemination of a social norm are independent of other factors that might influence reviewing behavior. To ensure that this was the case, we performed balance tests across a number of available subject-level covariates that we obtained from the retail partner and a third-party market research firm that tracks data about TMall users. Table A1 in Online Appendix A reports the results of these balance tests. The general lack of significant differences supports the validity of our randomization procedure.

⁷ This payment amount is in line with past work in this space, which has typically offered subjects US\$1 to US\$2 in exchange for authoring an online review (Cabral and Li 2015, Stephen et al. 2012, Wang et al. 2012). We also explored an additional treatment condition, in which we offered subjects ¥5. We observed no statistically significant difference between the ¥5 and ¥10 groups for any of our dependent variables.

⁸ It is important to note that there is some disagreement in the literature about when and whether to employ multiple comparison corrections. Some textbooks advise using these adjustments when it comes to hypothesized tests, while others actually advise that these adjustments are specifically relevant only to untheorized tests (Barraguet et al. 2016, p. 321; Pagano 2012, p. 422). Applying a Bonferroni correction to the five pairwise tests that pertain to H1–H4

lowers the “critical threshold” to $0.05/5 = 0.01$. If we further consider that our four hypotheses are evaluated on two independent samples of data, we might relax this threshold to 0.02, and if we were to employ one-tailed tests, we might further relax this threshold to 0.04. Regardless, only H2 (the positive effect of social norms on review volumes, compared with the control) fails to meet this bar (two-tailed p -value = 0.076) and thus might be considered only weakly supported.

References

- Allcott H (2011) Social norms and energy conservation. *J. Public Econom.* 95(9):1082–1095.
- Anderson E (1998) Customer satisfaction and word of mouth. *J. Service Res.* 1(1):5–17.
- Aral S, Walker D (2011) Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Sci.* 57(9):1623–1639.
- Avery C, Resnick P, Zeckhauser R (1999) The market for evaluations. *Amer. Econom. Rev.* 89(3):564–584.
- Barnes R (1949) *Motion and Time Study* (John Wiley & Sons, New York).
- Barragans J, Morais A, Guisasaola J (2016) *Probability and Statistics: A Didactic Introduction* (CRC Press, Boca Raton, FL).
- Brief AP, Motowidlo SJ (1986) Prosocial organizational behaviors. *Acad. Management Rev.* 11(4):710–725.
- Burtch G, Ghose A, Wattal S (2015) The hidden cost of accommodating crowdfunder privacy preferences: A randomized field experiment. *Management Sci.* 61(5):949–962.
- Cabral L, Li L (2015) A dollar for your thoughts: Feedback conditional rebates on eBay. *Management Sci.* 61(9):2052–2063.
- Cao Q, Duan W, Gan Q (2011) Exploring determinants of voting for the helpfulness of online user reviews: A text mining approach. *Decision Support Systems* 50(2):511–521.
- Chen Y, Harper F, Konstan J, Li S (2010) Social comparisons and contributions to online communities: A field experiment on MovieLens. *Amer. Econom. Rev.* 100(4):1358–1398.
- Cialdini R, Goldstein N (2004) Social influence: Compliance and conformity. *Annual Rev. Psych.* 55:591–621.
- Cialdini RB, Trost MR (1998) Social influence: Social norms, conformity and compliance. Gilbert D, Fiske S, Gardner L, eds. *The Handbook of Social Psychology*, 4th ed., Vol. 2, (McGraw-Hill, New York), 151–192.
- Cialdini R, Kallgren C, Reno R (1991) A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. Zanna MP, ed. *Advances in Experimental and Social Psychology*, Vol. 24 (Academic Press, San Diego), 201–234.
- Cialdini RB, Eisenberg N, Green BL, Rhoads K, Bator R (1998) Undermining the undermining effect of reward on sustained interest. *J. Appl. Soc. Psych.* 28(3):249–263.
- Deci E, Koestner R, Ryan R (1999) A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psych. Bull.* 125(6):627–668.
- Dellarocas C (2003) The digitization of word-of-mouth: Promise and challenges of online feedback. *Management Sci.* 49(10):1407–1424.
- Dimoka A, Hong Y, Pavlou PA (2012) On product uncertainty in online markets: Theory and evidence. *MIS Quart.* 36(2):395–426.
- Ferraro P, Price M (2013) Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Rev. Econom. Statist.* 95(1):64–73.
- Fradkin A, Grewal E, Holtz D, Pearson M (2015) Bias and reciprocity in online reviews: Evidence from field experiments on Airbnb. *Proc. Sixteenth ACM Conf. on Econom. and Comput.* (ACM, New York), 641–641.
- Frey B (1994) How intrinsic motivation is crowded in and out. *Rationality Soc.* 6(3):334–352.
- Fryer RG (2011) Financial incentives and student achievement: Evidence from randomized trials. *Quart. J. Econom.* 126(4):1755–1798.
- Gallus J (2016) Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia. *Management Sci.*, ePub ahead of print September 30, <http://dx.doi.org/10.1287/mnsc.2016.2540>.
- Gerber AS, Rogers T (2009) Descriptive social norms and motivation to vote: Everybody’s voting and so should you. *J. Politics* 71(1):178–191.
- Goes PB, Guo C, Lin M (2016) Do incentive hierarchies induce user effort? Evidence from an online knowledge exchange. *Inform. Systems Res.* 27(3):497–516.
- Goldstein N, Cialdini R, Griskevicius V (2008) A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *J. Consumer Res.* 35(3):472–482.
- Hennessey B, Zbikowski S (1993) Immunizing children against the negative effects of reward: A further examination of intrinsic motivation training techniques. *Creativity Res. J.* 6(3):297–308.
- Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A, Cardenas J, et al. (2006) Costly punishment across human societies. *Science* 312(5781):1767–1770.
- Heyman J, Ariely D (2004) Effort for payment: A tale of two markets. *Psych. Sci.* 15(11):787–793.
- Horton J, Chilton L (2010) The labor economics of paid crowdsourcing. *Proc. 11th ACM Conf. Electronic Commerce* (ACM, New York), 209–218.
- Ipeirotis P (2010) Demographics of Mechanical Turk. Working paper, New York University, New York. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1585030.
- Jabr W, Mookerjee R, Tan Y, Mookerjee V (2014) Leveraging philanthropic behavior for customer support: The case of user support forums. *MIS Quart.* 38(1):187–208.
- Jacques E, Rice A, Hill J (1951) The social and psychological impact of a change in method of wage payment. *Human Relations* 4(4):115–140.
- Jenkins GD, Mitra A, Gupta N, Shaw J (1998) Are financial incentives related to performance? A meta-analytic review of empirical research. *J. Appl. Psych.* 83(5):777–787.
- Jiang Z, Benbasat I (2007) The effects of presentation formats and task complexity on online consumers’ product understanding. *MIS Quart.* 31(3):475–500.
- Khern-am-nuai W, Kannan K (2016) Extrinsic versus intrinsic rewards to participate in a crowd context: An analysis of a review platform. Working paper, Purdue University, West Lafayette, IN. <http://ssrn.com/abstract=2496528>.
- Kline P (2000) *The Handbook of Psychological Testing*, 2nd ed. (Routledge, London).
- Levi A, Mokryn O, Diot C, Taft N (2012) Finding a needle in a haystack of reviews: Cold start context-based hotel recommender system. *Proc. 6th ACM Conf. Recommender Systems (RecSys ’12)* (ACM, New York), 115–122.
- Liu J, Cao Y, Lin C, Huang Y, Zhou M (2007) Low-quality product review detection in opinion summarization. *Proc. Joint Conf. Empirical Methods Natural Language Processing Comput. Natural Language Learn.* (Association for Computational Linguistics, Stroudsburg, PA), 334–342.
- Luca M (2016) User-generated content and social media. Anderson SP, Stromberg D, Waldfogel J, eds. *Handbook of Media Economics*, Vol. 1 (Elsevier, Amsterdam), 563–592.
- Mason W, Watts D (2009) Financial incentives and the “performance of crowds.” *Proc. ACM SIGKDD Workshop on Human Computation* (ACM, New York), 77–85.
- Mudambi S, Schuff D (2010) What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quart.* 34(1):185–200.
- Nolan J, Schultz P, Cialdini R, Goldstein N, Griskevicius V (2008) Normative social influence is underdetected. *Personality Soc. Psych. Bull.* 34(7):913–923.
- Pagano R (2012) *Understanding statistics in the behavioral sciences*. Cengage Learning, Belmont, CA.
- Rick SJ, Cryder CE, Loewenstein G (2008) Tightwads and spend-thrifts. *J. Consumer Res.* 34(6):767–782.

- Robinson E, Fleming A, Higgs S (2014) Prompting healthier eating: Testing the use of health and social norm based messages. *Health Psych.* 33(9):1057–1064.
- Schultz P, Nolan J, Cialdini R, Goldstein N, Griskevicius V (2007) The constructive, destructive, and reconstructive power of social norms. *Psych. Sci.* 18(5):429–434.
- Seuntjens TG, Zeelenberg M, van de Ven N, Breugelmans SM (2015) Dispositional greed. *J. Personality Soc. Psych.* 108(6):917–933.
- Stephen A, Bart Y, Du Plessis C, Goncalves D (2012) Does paying for online product reviews pay off? The effects of monetary incentives on content creators and consumers. Gürhan-Canli Z, Otnes C, Zhu RJ, eds. *NA—Advances in Consumer Research*, Vol. 40 (Association for Consumer Research, Duluth, MN), 228–231.
- Thibaut J, Kelley H (1959) *The Social Psychology of Groups* (John Wiley & Sons, Oxford, UK).
- Volpp KG, Troxel AB, Pauly MV, Glick HA, Puig A, Asch DA, Galvin R, et al. (2009) A randomized, controlled trial of financial incentives for smoking cessation. *New England J. Medicine* 360(7):699–709.
- Wang J, Ghose A, Ipeirotis P (2012) Bonus, disclosure, and choice: What motivates the creation of high-quality paid reviews? *Internat. Conf. Inform. Systems (ICIS 2012)* (Association for Information Systems (AIS), Atlanta).