Application for IFREE Small Grants Program

Date	Total Budget	Amount Requested from IFREE	
8/10/2015	5040	5040	
Proposal Title: How do we rate? An experimental study of unary versus bina		Proposed Start Date: 2/1/2016	Proposed End Date: 2/1/2017
Principal Investigator: Jonathan Lafky	Position / Title: Assistant Professor	Institution and address: 110 Simon Center, Lafayette College, Easton, PA 18042	
PI Email: jlafky@gmail.com	PI Phone: 412 849 7195		

IFREE affiliations of PI

- ☐ Graduate Student workshop. Date: 07/24/06
- ☐ Pre-doctoral or post-doctoral fellow Date: Institution:
- ☐ Visiting scholar Date: Institution:
- ☐ Other Explain:

IFREE welcomes proposals from any scholar affiliated with any research institution, regardless of geographic location or prior IFREE affiliation. However, preference will be given to individuals with a prior affiliation.

Co-PI Name(s)	Co-PI Position / Title	Co-PI Institution
Simon Halliday	Assistant Professor	Smith College
Alistair Wilson	Assistant Professor	University of Pittsburgh

Project summary (250 word limit. Must be in non-technical language that is appropriate for a general audience):

Online commerce is large and growing component of the economy that relies heavily on the existence of user-based rating systems. These systems are a mutually beneficial market-based solution to the problem of incomplete information in online markets.

Two broad classes of rating systems have become commonplace online. The first, commonly found at retailers such as amazon.com, allows consumers to evaluate a product over a variety of levels, often from 1 to 5 "stars." Other systems, especially those used by social networks, allow only unary rating, in which a user either gives a favorable rating (a "like," "+1" or "favorite"), or no rating at all. Alternatively, other sites specialize in complaints, effectively allowing only negative ratings to be given. Our goal is to understand how the choice of rating system, either allowing a full range of negative and positive ratings, or restricted to just one type, influences information transfer and market efficiency. We are interested both in experimentally observing behavior under each system, as well as

Expected benefits and broader implications (250 word limit. Must be in non-technical language that is appropriate for a general audience):

Our results will provide insight into the relative benefits of different commonly-used rating system designs. Our main results will show how different designs influence the frequency and accuracy of ratings, and how the act of restricting the types of ratings available changes consumer behavior. All these results will help us to better understand how ratings work to alleviate problems of incomplete information in markets, allowing for more mutually-beneficial exchanges between buyers and sellers.

Proposal

Online commerce is a large and rapidly growing component of the global economy, and a major factor in its continuing success is the presence of user-based rating systems. Working through decentralized, voluntary provision, online ratings allow consumers to learn about the quality of otherwise unfamiliar goods, increasing the likelihood of a purchase. There are obvious incentives for sellers to offer ratings, as increased consumer confidence can lead to increased profits. Consumer-based ratings are a mutually beneficial market-based solution to the problem of incomplete information in online markets, and may provide a channel for markets to self-regulate and create long-run value.

Despite the benefits and widespread use of rating systems, economic research on them is still relatively scarce. Existing work has focused on incentives for providing a rating (Chen et al. 2010, Lafky 2014, and Wang 2010), the prevalence of fraudulent ratings (Mayzlin et al. 2014), and the impact of ratings on sales (Chevalier and Mayzlin 2006). There is also evidence of the underprovision problem that unincentivized rating systems share with other public-good environments, as in Li and Xiao (2014) and Lafky and Wilson (2015).

Our project focuses on a design issue for rating systems, examining two broad rating designs that have become common in the field. The first, commonly used by Amazon.com, Netflix, and other major retailers, allows consumers to evaluate a product over a range of favorable and unfavorable levels, often from 1 to 5 "stars." Other systems, especially those used by social networks, allow only a unary rating, in which a user either gives a favorable rating (a "like," a "+1" or a "favorite") to a product or takes no action.

Our goal is to understand how this rating design - either allowing a full range of negative and positive ratings, or restricted to just one type - influences information aggregation and welfare. We are interested in observing rater and consumer behavior within each system, as well as how the observed act of choosing one system over another might influence consumers.

We propose to study this behavior in a laboratory setting with four experimental treatments. Each treatment will consist of repeated interactions between groups of three participants. Within each group there will be one seller, one rater and one buyer. In each period the rater learns the seller's quality and then decides whether or not to pay a small cost to send a rating to the buyer. The buyer then learns what, if any, rating was sent, and decides whether to buy from the seller or take an outside option instead. Participants are then randomly re-matched into groups of three for the next period.

For simplicity, we will collapse the usual 1-5 star rating system into a binary one, in which ratings are either favorable or unfavorable, equivalent to a user giving a "thumbs-up" or "thumbs-down" to a product. We will compare this binary setting to two unary environments, which we call Unary+ and Unary-. In Unary+ the rater chooses between sending a favorable message or taking no action, while in Unary- the only possible message is unfavorable. While Unary- is far less common in practice than Binary or Unary+, we include it both for a complete comparison to the Binary system, and as a model of a complaint-based system such as the Better Business Bureau.

In addition to the three treatments with a fixed choice of ratings, we will also consider a fourth "Choice" treatment, in which the seller selects the rating design. In each period the seller can choose to allow only favorable ratings (Unary+), only unfavorable ratings (Unary-), or give raters their choice of either type of rating (Binary). The Choice treatment therefore allows endogenous selection of the rating system used. Previous work suggests that when a principal chooses to restrict an agent's choice set, agents tend to respond with negative reciprocity (Falk & Kosfeld, 2006, Ziegelmeyer et al, 2012 and Burdin et al 2015). If these results hold in a rating environment, we would expect to see raters behave differently based on how the rating system was selected. For example, we should see more-informative ratings provided in the Unary+ treatment than under the same rating system in the Choice treatment.

Our main questions focus on changes in rater and buyer behavior between four treatments. How does the frequency of sending a rating vary by rating system? How do buyers interpret ratings and, perhaps more importantly, the lack of ratings in each treatment? How does each system influence market efficiency? How do raters respond to having their rating options restricted in the Choice treatment? We will test the following hypotheses:

- Raters are more likely to send a rating in the Unary+ treatment than when constrained to favorable messages in the Choice treatment.
- Raters are more likely to send ratings in the Unary- treatment than when constrained to unfavorable messages in the Choice treatment.
- Buyers in the Choice treatment will infer sellers to be of low quality if they constrain ratings to be favorable and no rating is sent.
- Welfare will be higher in the Binary treatment than in the Choice treatment.

Through running these experiments we will contribute to a better understanding of how the design of rating systems influences consumer behavior and market efficiency. We will provide evidence of why different designs may exist in the field as well as suggestions for implementing future systems. Our results will also give insight broadly into how consumers respond to being restricted in the types of opinions they can share. All these results will help us to better understand how ratings work to alleviate problems of incomplete information in markets, allowing for more mutually-beneficial exchanges between buyers and sellers.

Statement about how proposed research contributes to IFREE's mission

Consumer-based rating systems help markets to function more efficiently, providing a self-regulating channel for buyers and to exchange information, leading to mutually beneficial exchanges. While there are diverse rating designs employed in various marketplaces, we do not fully understand the causes or implications of this diversity. Are these different systems all solving the same problem with varying efficacy, or has each system evolved for a unique purpose? Our results will help to explain if restrictions on the type of ratings that can be sent alters the extent to which information is aggregated in online marketplaces.

Moreover, our Choice treatment will teach us whether the act of choosing a specific rating design might affect consumers' interactions within it. Our results will provide a testbed to demonstrate whether a seller changing their rating design to restrict what types of rating can be sent (both positive and negative, just positive, just negative) might influence the types of ratings chosen within it.

Studying the roles of different rating systems in online marketplaces will lead us to a more nuanced understanding of the ways in which economic exchange is evolving as it moves to the internet. It will also provide evidence of the market's ability to overcome problems of incomplete information without requiring third party intervention or long-term product branding. By examining the relative performance of the different systems we will contribute to the understanding of how each rating system can promote the functioning of freely functioning, efficient markets.

References Cited

Burdin, G., S. Halliday and F. Landini (2015), Second-party vs. Third-party Control: Disentangling the roles of autonomy and reciprocity, IZA Discussion Paper No. 9251, August, 2015.

Chen, Y., F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li. Social comparisons and contributions to online communities: A field experiment on movielens. *The American economic review* (2010): 1358-1398.

Chevalier, J. and D. Mayzlin (2006) The effect of word of mouth on sales: Online book reviews. J. Marketing Res. **43**(3), 345–354.

Falk, A. and M. Kosfeld (2006), "The Hidden Costs of Control", *American Economic Review*, 96(5): 1611-1630.

Lafky, J., (2014) Why do people rate? *Games and Economic Behavior,* **87** 554-570.

Lafky, J. and A. J. Wilson, Quality Versus Quantity in Information Transmission, University of Pittsburgh Economic Department Working Paper.

Li, L. and Xiao, E. (2014) Money Talks: Rebate Mechanisms in Reputation System Design. *Management Science* **60**:8, 2054-2072.

Mayzlin, D., Y. Dover, and J. Chevalier. (2014) Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review* **104**, 2421-2455.

Wang, Zhongmin. Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews. *The BE Journal of Economic Analysis & Policy* 10.1 (2010).

Ziegelmeyer, A, Schmelz, K., and. Ploner, M., (2012), "Hidden Costs of Control: Four Repetitions and an Extension", *Experimental Economics*, Springer, vol. 15(2), 323-340, June.

Budget

Participant fees for four treatments: Unary+, Unary-, Binary and Choice.

Unary+ treatment:

2 sessions x 24 participants per session = 48 participants

Unary- treatment:

2 sessions x 24 participants per session = 48 participants

Binary treatment:

2 sessions x 24 participants per session = 48 participants

Choice treatment:

4 sessions x 24 participants per session = 96 participants

Combined:

10 sessions x 24 participants per session = 240 participants

$20/participant x 240 participants = $4800.

Show-up fees for overbooked participants:

Assuming 20% per session overbooking.

$5.00/participant * 48 participants = $240.

Grand total:

$5040.

Timeline

February-March 2016: Run experimental sessions.

April-June 2016: Analyze data, write paper.

June-August 2016: Circulate working paper, gather seminar feedback.

August 2016: Submit for publication.

January 2017: Submit 6-month report to IFREE.

Summary of qualifications of investigators

Jonathan Lafky is an Assistant Professor of Economics at Lafayette College. His research focuses on rating systems and group dynamics. He has published experimental papers at *Games and Economic Behavior* and *Experimental Economics*. He has conducted dozens of experimental studies as a research assistant while at the University of Pittsburgh.

Simon Halliday is an Assistant Professor of Economics at Smith College. His research focuses on punishment, risk and control. He has done experimental work on taking and third-party punishment (under submission, an ERSA working paper), second-party vs. third-party control (an IZA discussion paper), and gender and risk-taking (under submission, a SALDRU working paper).

Alistair Wilson is an Assistant Professor of Economics at the University of Pittsburgh. His research focuses on ratings systems, dynamics, and market design. He has published papers in the *American Economic Review* and *Quantitative Economics* and has papers with invited revisions at *Games and Economic Behavior*.

Description of facilities available for research and statement of compliance with IRB

We will conduct the experiments at the Pittsburgh Experimental Economics Laboratory at the University of Pittsburgh. This is a well-established lab with a 40-seat capacity and approximately 2000 active participants at any time. We will run experimental sessions only once all three of our institutions' IRBs have granted approval, according to each institution's standard approval process.