

Stack Overflow Badges and User Behavior: An Econometric Approach

Andrew Marder
Harvard Business School
Boston, Massachusetts, USA
amarder@hbs.edu

Abstract—Does gamification work? This paper examines how Stack Overflow users behave when earning badges. A regression analysis of user activity logs shows users change their contribution amounts when earning some badges but not others. This paper adds new support to the growing literature that gamification works, but its efficacy is context-dependent [1]. Alternative methods for motivating user contributions are considered.

I. INTRODUCTION

Stack Overflow is a question and answer community designed for programmers. It is the largest of 130 communities in the Stack Exchange network. Created in 2008, the knowledge organized by Stack Overflow has become a valuable resource for software developers. On January 20, 2015, Spoelsky [2] announced that Stack Exchange had raised \$40 million in venture capital funding. Stack Overflow gives users who ask questions access to expert technical help, while users who answer questions build their reputation for technical expertise.

Deterding et al. [3] define “gamification as the use of game design elements in non-game contexts.” Stack Overflow gamifies the process of asking and answering questions as follows. A user earns reputation points when another user votes on her posts (5 points when a question is voted up, 10 points when an answer is voted up, 15 points when an answer is accepted, and 2 points when an edit is approved). As a user earns reputation points she unlocks privileges on the site. For instance, a user must have at least 15 reputation points to vote up a question or answer.¹ Users are awarded badges for special achievements. For example, one receives the *Informed* badge by reading the tour page.²

This paper takes a first step along the path of applying econometric analysis to publicly available Stack Overflow data [4]. Do badges motivate users to contribute to the site? Which badges are most effective? What types of user contributions are responsive to gamification? To begin answering these questions, I study how users behave around the time they are awarded badges.

II. MOTIVATING CONTRIBUTIONS

Kraft-Todd et al. [5] survey the literature on field experiments designed to promote cooperation in social dilemmas.

¹A full list of privileges and the necessary reputation points is available at <http://stackoverflow.com/help/privileges>.

²The Stack Overflow tour can be found at <http://stackoverflow.com/tour>, and all badges are listed on <http://stackoverflow.com/help/badges>.

Examples of social dilemmas include voter turnout, environmental protection, and universal healthcare. Table I describes the four categories they use to classify interventions.

TABLE I
FOUR TYPES OF INTERVENTION USED TO PROMOTE CONTRIBUTIONS TO A PUBLIC GOOD.

	<i>Material-benefit</i>	<i>Social-benefit</i>
<i>Self-focused</i>	Material rewards	Observability
	Cash or gifts provided in exchange for contributing	Others informed about your contribution decisions
<i>Other-focused</i>	Increased efficacy	Descriptive norms
	Matching/seed funds provided, or benefit to others emphasized	You are informed about the contribution decisions of others

Focusing on the experiments surveyed by Kraft-Todd et al. [5], interventions that provided material benefits led to mixed results while interventions that provided social benefits were consistently effective at promoting cooperation. The general categories of Table I hint at some of the specific reasons why users might contribute to Stack Overflow:

- 1) Material rewards: A user might contribute to Stack Overflow to signal her technical ability to potential employers. By contributing to Stack Overflow she increases the probability of finding better employment opportunities, leading to material rewards.
- 2) Increased efficacy: By emphasizing how high quality answers help others solve similar problems, Stack Overflow can motivate other-interested individuals to contribute. [6]
- 3) Observability: Contributions to the platform are publicized through the reputation and badge system. If a user is concerned about how others perceive her on the site (generous, competent, ...), she may be motivated to contribute.

TABLE II
BADGES OF INTEREST

Name	Description	Awarded	Introduced	Dropped
<i>Strunk & White</i>	Edited 80 posts	7,073	2008-09-15	0.00
<i>Copy Editor</i>	Edited 500 posts	1,288	2010-07-09	0.04
<i>Archaeologist</i>	Edited 100 posts that were inactive for 6 months	691	2011-08-15	0.05
<i>Curious</i>	Asked a well-received question on 5 separate days ³	138,264	2014-07-02	0.86
<i>Inquisitive</i>	Asked a well-received question on 30 separate days	14,081	2014-07-02	0.92
<i>Socratic</i>	Asked a well-received question on 100 separate days	1,240	2014-07-02	0.93
<i>Explainer</i>	Edited and answered 1 question (answer score > 0)			
<i>Refiner</i>	Edited and answered 50 questions (answer score > 0)			
<i>Illuminator</i>	Edited and answered 500 questions (answer score > 0)			

The six badges considered in this paper were introduced to the Stack Overflow site at different times. The *Strunk & White* badge was first awarded on September 15, 2008, while the three badges for asking questions were all introduced on July 2, 2014. Since the badges for asking questions were added to the site so recently many users who have been awarded these badges earned them for actions taken before the badge was introduced. For instance, 86% of users who earned the *Curious* badge were awarded the badge for actions taken before July 2, 2014. I drop these users from the analysis as they have no incentive to change their behavior to earn the *Curious* badge. The three badges awarded for editing and answering multiple questions were introduced to Stack Overflow after September 14, 2014 and are not included in the Stack Exchange Data Dump updated on September 26, 2014.

- 4) Descriptive norms: If a user is concerned about how her contributions compare to the contributions of others, the reputation leaderboards may motivate her to contribute.⁴

This is not an exhaustive list of why users contribute to Stack Overflow. However, it will be useful when considering why users increase contributions when receiving some but not all badges.

III. METHODS

Grant and Betts [7] present empirical evidence that three badges awarded for editing encourage recipients to make more edits in the two months preceding receipt of the badge compared to the two months after receiving the badge. This paper extends their findings by examining all types of user activity (posting questions, posting answers, and editing posts), and exploring the impact of three new badges awarded for asking questions. Table II describes the six badges considered in this paper.

Let y_{it} be the number of edits user i makes on day t , and t_i^* denote the day user i receives the badge of interest. Following the approach of Jacobson et al. [8], I regress the number of edits user i makes on day t on a user fixed effect α_i , a set

of dummy variables indicating whether the user received the badge on day $t - k$, and day of the week effects γ_j

$$\log(1 + y_{it}) = \alpha_i + \sum_{k=-29}^{30} \mathbb{1}\{t = t_i^* + k\} \delta_k + \sum_{j=1}^6 \mathbb{1}\{t \bmod 7 = j\} \gamma_j + \epsilon_{it}. \quad (1)$$

The model parameters are estimated using an ordinary least squares regression, and standard errors are clustered at the user level.

Define $f(k)$ to be the expected number of actions taken on the k 'th day since receiving the badge

$$f(k) = E[\log(1 + y_{it}) \mid t = t_i^* + k]. \quad (2)$$

The predicted number of actions $f(\hat{k})$ is presented in Figure 1. The 95% confidence interval is depicted as a gray band around the linear prediction, standard errors were calculated using the delta method [9].

IV. RESULTS

The first three rows of Figure 1 illustrate how user activity changes around the time one earns a badge for editing. Each row is labeled with the name of the focal badge (*Strunk & White*, *Copy Editor*, and *Archaeologist*). There is one column for each type of user action (posting a question, posting an answer, or editing a post). The figure confirms the findings of Grant and Betts [7]. Editing increases gradually before receiving a badge for editing, with a large jump in activity on the award day. We also see that editing drops quickly

³A question is *well-received* if it is open and has a score greater than 0. <http://meta.stackexchange.com/questions/234259/asking-days-badges>

⁴The Stack Exchange leaderboards are online at <http://stackexchange.com/leagues>.

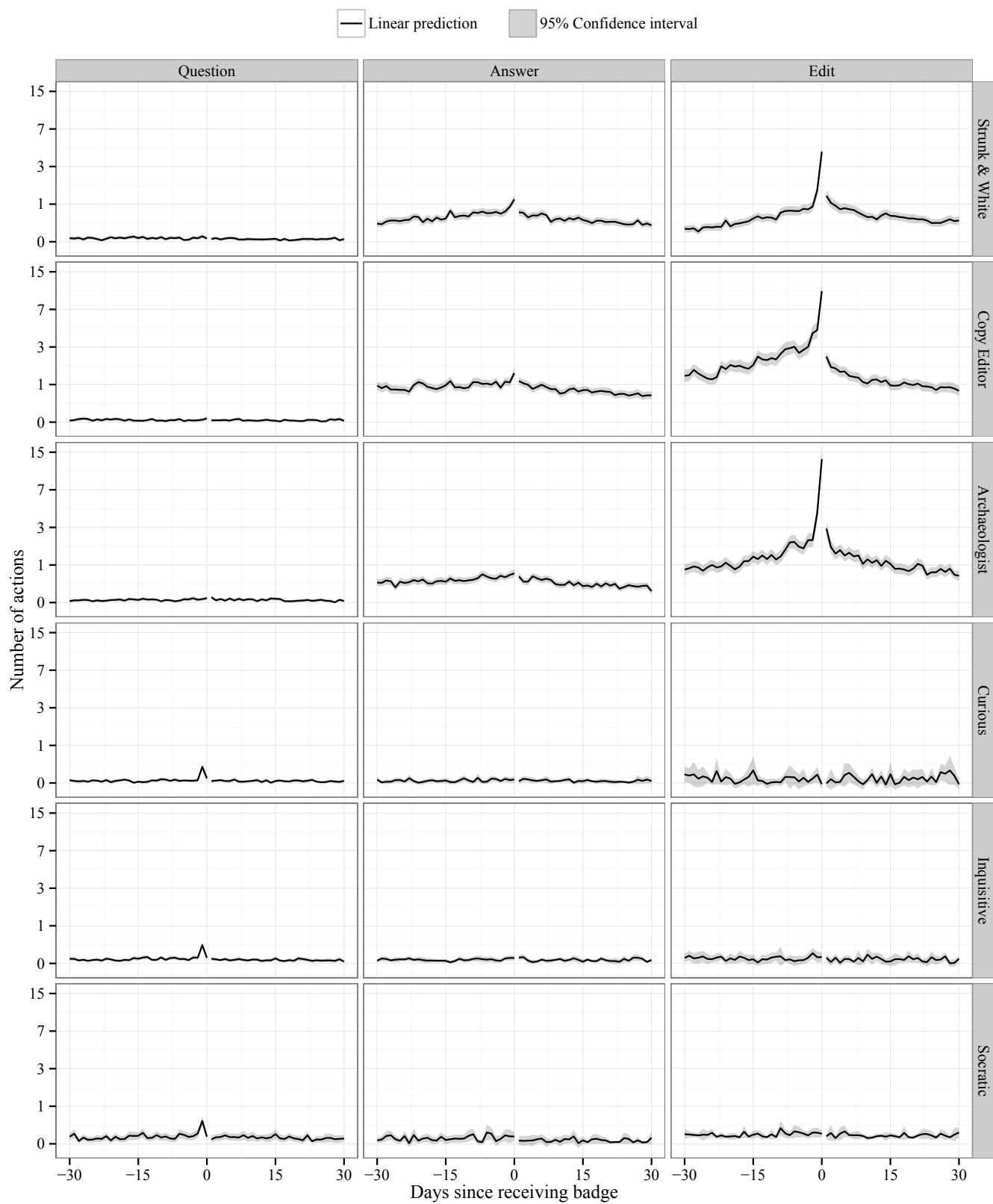


Fig. 1. User activity over time

after receiving the badge and gradually declines over time. It is interesting to see how few questions were asked by the recipients of the editing badges, and that the rate of answering questions has a very slight increase leading up to receiving the badge and a similarly slight decrease after receiving the badge.

The results for the question-focused badges, *Curious*, *Inquisitive*, and *Socratic*, are quite different. In general, recipients of these badges are not particularly active on the site. The average level of questions, answers, and edits made all hover near zero. The uptick in questions asked on the day before receiving the badge is mechanical. Many users who earn these badges ask a question the day before they earn the badge.

V. CONCLUSION

When interpreting the empirical results of this paper, please recall Holland and Rubin's motto "no causation without manipulation" [10]. There is no manipulation of the explanatory variable in this study, consequently we have not identified the causal effect of badges. To estimate the causal impact of badges on user activity we need to find a source of exogenous variation [11].

This paper confirms the empirical observation of Grant and Betts [7]. On average, users who receive a badge for editing make more edits in the 30 days prior to receiving the badge compared to the 30 days after receiving the badge. In addition, we find that recipients of the editing badges ask almost no questions, and answer about one question each day. Finally, we show that users who receive badges for asking questions behave differently. In particular, we found that users do not appear motivated to change their activity levels to earn badges for asking questions.

Although Stack Overflow has design elements that exemplify each of the four motivators outlined by Grant and Betts [7], badges awarded for asking questions appear unable to increase the number of questions posted to the site. Asking a good question is difficult, it requires "search and research" [12]. One feature that might increase the number of questions posted is adding support for anonymous questions. Allowing a user to remove the observability of her action would help her avoid the social cost of posting a poorly researched question. A careful system of anonymous questions would have to balance the benefit of increasing the number of questions asked, with the cost of lower quality questions on average.

ACKNOWLEDGEMENTS

The author would like to thank three anonymous reviewers for helpful comments and suggestions, and Harvard Business School for financial support.

REFERENCES

- [1] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? - A literature review of empirical studies on gamification," *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2014. [Online]. Available: <http://dx.doi.org/10.1109/HICSS.2014.377>
- [2] J. Spoelsky, "Andreessen Horowitz Invests in Stack Exchange," 2015. [Online]. Available: <http://blog.stackoverflow.com/2015/01/andreessen-horowitz-invests-in-stack-exchange/>
- [3] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining "Gamification"," in *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11*, 2011. [Online]. Available: <http://dx.doi.org/10.1145/2181037.2181040>
- [4] S. Exchange, "Stack Exchange Data Dump," 2014. [Online]. Available: <https://archive.org/details/stackexchange>
- [5] G. Kraft-Todd, E. Yoeli, S. Bhanot, and D. Rand, "Promoting cooperation in the field," *Current Opinion in Behavioral Sciences*, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.cobeha.2015.02.006>
- [6] M. E. Gerbasi and D. A. Prentice, "The Self- and Other-Interest Inventory," *Journal of Personality and Social Psychology*, 2013. [Online]. Available: <http://dx.doi.org/10.1037/a0033483>
- [7] S. Grant and B. Betts, "Encouraging user behaviour with achievements: An empirical study," in *IEEE International Working Conference on Mining Software Repositories*, 2013. [Online]. Available: <http://dx.doi.org/10.1109/MSR.2013.6624007>
- [8] L. S. Jacobson, R. J. LaLonde, and D. G. Sullivan, "Earnings losses of displaced workers," *The American Economic Review*, 1993. [Online]. Available: <http://www.jstor.org/stable/2117574>
- [9] R. Williams, "Using the margins command to estimate and interpret adjusted predictions and marginal effects," *Stata Journal*, 2012. [Online]. Available: <http://econpapers.repec.org/RePEc:tsj:stataj:v:12:y:2012:i:2:p:308-331>
- [10] P. W. Holland, "Statistics and Causal Inference," *Journal of the American Statistical Association*, 1986. [Online]. Available: <http://dx.doi.org/10.1080/01621459.1986.10478354>
- [11] A. Miller, "Principal turnover and student achievement," *Economics of Education Review*, 2013. [Online]. Available: <http://dx.doi.org/10.1016/j.econedurev.2013.05.004>
- [12] S. Overflow, "How do I ask a good question?" [Online]. Available: <http://stackoverflow.com/help/how-to-ask>
- [13] A. T. T. Ying, "Mining Challenge 2015: Comparing and combining different information sources on the Stack Overflow data set," in *The 12th Working Conference on Mining Software Repositories*, 2015.