HARVARD | BUSINESS | SCHOOL

Does Who Helps You Impact Your Behavior? Examining the Effects of Social Interactions on Knowledge Sharing in Online Communities

Eunkwang Seo
Frank Nagle
Sonali K. Shah

Working Paper 21-026

Does Who Helps You Impact Your Behavior? Examining the Effects of Social Interactions on Knowledge Sharing in Online Communities

Eunkwang Seo
University of Illinois at Urbana-Champaign

Frank Nagle
Harvard Business School

Sonali K. Shah
University of Illinois at Urbana-Champaign

Working Paper 21-026

Funding for this research was provided in part by Harvard Business School.

Does Who Helps You Impact Your Behavior?
Examining the Effects of Social Interactions on Knowledge Sharing in Online Communities

Eunkwang Seo
University of Illinois at Urbana-Champaign
eseo7@illinois.edu

Frank Nagle
Harvard Business School
fnagle@hbs.edu

Sonali K. Shah
University of Illinois at Urbana-Champaign
sonali@illinois.edu

Keywords: Online communities, knowledge sharing, information search, reciprocity

Acknowledgements: We thank Aravinda Garimella, Hyeonsuh Lee, Hari Sundaram, Nataliya Langburd Wright and participants at the 2019 Strategic Management Society Conference, the 2019 Boston College Digital Innovation Workshop, the 2019 Harvard Business School Strategy Alumni Workshop, and the 2020 Academy of Management Conference for feedback on these ideas. All authors contributed equally.

Does Who Helps You Impact Your Future Behavior? Examining the Effects of Social Interactions on Knowledge Sharing in Online Communities

ABSTRACT

Online communities provide vibrant forums for knowledge sharing and are increasingly being used by individual users and firms to source knowledge and create and capture value. Yet, there is much to learn about how the actions of community members affect other members, particularly new members whose continued participation is important to community vibrancy and growth. The extant literature investigates whether or not receiving a reply increases the likelihood that a new user will post again, finding mixed results. We extend this literature in two ways: we examine the effects of (1) receiving a reply on *different types of subsequent participation* (asking another question and replying to others' questions) and (2) receiving a reply from community members with different *affiliations* (user or employee). Based on data from the Statalist community, our findings show that receiving a reply from another user is positively associated with subsequent questioning, but negatively associated with subsequent replying by new users. Conversely, receiving a reply from an employee is positively associated with subsequent replying, but not with subsequent questioning by new users. To help address endogeneity concerns, we run fixed-effects models, implement an instrumental variable-based approach (using a two-stage Heckman selection model), use a bivariate probit model and coarsened exact matching, and conduct qualitative interviews with Stata employees. Our findings suggest that social interactions play an important, albeit nuanced and sometimes unexpected, role in shaping subsequent participation and highlight the importance of examining the effects of firm participation in online communities.

INTRODUCTION

Online communities serve as important forums for knowledge sharing (Lee and Cole 2003, Wasko and Faraj, 2005, Bagozzi and Dholakia 2006, Faraj et al. 2011, Zhang et al. 2013, Hwang et al. 2015, Aaltonen and Seiler 2016, Bauer et al. 2016). Within such communities, individuals voluntarily come together to ask and answer questions, using the Internet as their primary communication medium (Sproull and Arriaga 2007, O'Mahony and Lakhani 2011, Shah and Nagle 2020). Over time, a rich ongoing stream of interactions can create a vibrant community and a robust knowledge repository. Online communities represent a novel form of organizing of interest to scholars and practitioners alike: scholars are interested in better understanding social interactions in online communities (von Krogh and von Hippel 2006, Faraj, Jarvenpaa and Majchrzak 2011, Gulati, Puranam, and Tushman, 2012, Luo and Kaul 2019, He, Puranam, Shrestha, von Krogh 2020) and firms are increasingly interested in working with and harnessing online communities for their own goals (Goldman and Gabriel 2005, Dahlander 2007, Altman et al. 2015, 2021, Shaikh and Levina 2019, Shah and Nagle 2020).

Some online communities form around commercial products and services, providing value for both users and the firms around whose products they form (Franke et al. 2008, Bayus 2013, Manchanda et al. 2015, Nagle 2018, Dahlander et al. 2019, Shah and Nagle 2020).[1] In many of these communities, users tend to make up the bulk of the membership and interact primarily with other users, with occasional interactions with the firm's employees. For example, in the Apple Support community, users ask and reply to questions about various Apple products and services, with Apple employees posting on occasion. And, in the Visa Developer community, users exchange knowledge with many other users, as well as with a few Visa employees: "Find answers, ask questions, and connect with our community of Visa Developers from around the world building tomorrow's payment solutions. Let's Code Together!" (Visa Developer Community 2021). For the firms around whose product communities grow, understanding the dynamics of participation and the effects of employee participation on users is critical.

[1] Firms can, for example, derive strategic benefits by harnessing users to reduce support costs, strengthen their innovation process, and increase brand recognition (Chatterji and Fabrizio, 2012, 2014, 2016, Di Gangi and Wasko 2009, Goldman and Gabriel 2005, Jeppesen and Frederiksen 2006).

Online communities require the participation of many to form, become vibrant, and grow. Finding ways to encourage new users to continue participating is especially important, particularly in light of ample evidence suggesting that most users only make a single post to a community (Kuk 2006, Dahlander and Frederiksen 2012, Rullani and Haefliger 2013). Scholars have, therefore, stressed the importance of understanding the factors that lead new users to continue participating in the community (Bhattacherjee 2001, Butler 2001, Lampe and Johnston 2005, Joyce and Kraut 2006, Chen 2007). One factor that has been suggested as important is whether or not new users who post a question receive a reply. Several studies have investigated this question, unearthing conflicting results.[2] In a study of social interactions in online newsgroups, Joyce and Kraut (2006) find that new users who receive a reply to their initial post from another user are more likely to post again than new users who do not receive a reply. In contrast, Lampe and Johnston (2005), studying the Slashdot community, find no evidence of such a relationship.[3] One potential explanation for the conflicting findings is that these studies look at aggregate posting behavior, whereas asking questions and replying to questions are distinct behaviors that may be governed by different mechanisms and affected by different factors. This puzzle, combined with society's increasing tendency to use online communities to source and share knowledge, suggests a need to better understand the nuances of how social interactions affect knowledge sharing behaviors in online communities.

Our goal in this paper, therefore, is to deepen our understanding of the factors that affect how new users engage in the community after receiving a reply to their initial post. We do this in two ways. First, we examine the effects of receiving a reply on two *different types of subsequent participation behaviors* that frequently occur in, and are important to, knowledge sharing in online communities: asking questions and replying to the questions of other users. To our knowledge, studies of new users have not yet sought to differentiate between these different types of subsequent participation behaviors.[4] Second, we examine the

[2] These studies provide us with the initial base upon which we build our understanding of the effects of social interactions on subsequent knowledge sharing behaviors in online communities. In particular, we follow their precedent in examining *actual* behaviors in online communities.

[3] In addition, Zhang et al. (2013), studying *new and existing users* in open source software development communities also find no evidence of such a relationship.

[4] Looking at the literature more broadly, studies of *new and existing users* have looked at a broader array of behaviors, however these studies (a) often aggregate these behaviors in their reporting. Hence, we know little about differential patterns relating to

effects of the *affiliation* of the community member who replies to new users' first questions on new users' subsequent behavior: users or employees of the firm around whose product or service the community has formed. Given the distinctive affiliation of employees in these online communities (Kim 2000, Gu and Jarvenpaa 2003, Porter and Donthu 2008), new users might behave differently if they receive a reply from an employee, another user, or do not receive a reply. To our knowledge, prior studies have not considered the effects of receiving a reply from responders with different affiliations; rather, past studies have tended to focus on interactions between users. We, therefore, ask: *for new users who post a question to an online community, does receiving (versus not receiving) a reply to their initial question influence their likelihood of asking a subsequent question? Of replying to the question(s) of other user(s)? And, does the affiliation of the responder—a user or employee—affect the likelihood of engaging in each type of subsequent participation?* Table 1 summarizes the relationships we examine, highlighting that we investigate two sets of factors, one of which affects our independent variable, and the other our dependent variable.

[Insert Table 1 here]

We investigate these questions in the context of the Statalist community, an online forum dedicated to sharing knowledge related to the Stata statistical software package. We examine the behaviors of 8,157 new users who posted their first question during the 33-month time period between March 2014 and November 2016. Using logit regression analyses, we examine the aggregate behaviors of individuals participating in the community. We find that receiving a reply from another user *increases* the likelihood of a new user asking another question by 13% and *decreases* the likelihood of replying to another user's question by 83% compared to if they had not received a reply. In contrast, receiving a reply from an employee affects subsequent behavior very differently: receiving a reply from an employee does not appear to affect the likelihood of a new user asking another question, but *increases* the likelihood of replying to another user's question by 50% compared to if they had not received a reply. To help address endogeneity concerns related to how employees choose the questions to which they reply, we run fixed-

different types of subsequent participation. And, (b) measure users' *intent* to post rather than actual posting behavior (*e.g.,* Tiwana and Bush 2005, Chen 2007, Wu et al. 2007, Jin et al. 2010, Sun et al. 2012, Jin et al. 2013). Intentions are important, but espoused behaviors can be very different from actual behaviors (Wicker 1969, Ajzen 1991).

effect analyses, conduct qualitative interviews with Stata employees, and implement an instrumental variable-based approach (using a two-stage Heckman selection model). We also analyze the data utilizing a bivariate probit model and coarsened exact matching to provide evidence to support a more causal interpretation of the results.

These findings contribute to the literature in three ways. First, we extend our understanding of the effects of initial interactions in online communities. Specifically, we examine how receiving a reply affects new users' likelihood of participating again, adding nuance to the question by disaggregating future contributions by type (asking versus replying behaviors) and considering the affiliation of the respondent. Our findings reveal some surprising patterns that suggest that online social interactions are complex and, just as in the real world, it is important to take into account with whom the interactions are occurring and the types of subsequent behavior displayed. Second, our findings contribute to the literature on knowledge sharing in online communities by suggesting that reciprocity alone is unlikely to be enough to encourage new users to answer others' questions. Rather, our findings suggest that information search—where individuals find questions to which they might reply while searching for information to support their own work (and where individuals stop searching when they find the information they need)—may be at play. Finally, we contribute to the growing literature on firm participation in online communities by providing empirical evidence that small levels of effort by firms can contribute to the health and continued vibrancy of the community. These findings have practical implications for community members, firms, and policy-makers interested in seeding and maintaining vibrant communities.

THEORY & HYPOTHESES DEVELOPMENT

Participation is critical to the health and success of online communities (McGrath 1984, Preece 2000). Only when users contribute knowledge to the community does the community continue to function (Rullani and Haefliger 2013). Encouraging new users to participate again is critical (Bhattacherjee 2001, Butler 2001, Chen 2007), particularly as the majority of individuals who visit online communities contribute little and leave quickly (*e.g.,* Ren et al. 2012, Worthen 2008). For example, more than two thirds (68%) of new users to Usenet groups were never seen again after their initial post (Arguello et al. 2006).

And, over half (56%) of the new users who posted once to one of the six public newsgroups studied did not participate again (Joyce and Kraut 2006). Scholars have begun to investigate the effects of initial social interactions on subsequent participation in general, with some scholars finding evidence of a positive relationship and others finding no evidence of a relationship (*e.g.,* Lampe and Johnston 2005, Joyce and Kraut 2006). We seek to contribute to this dynamic and growing literature by examining the effects of two factors that have received little, if any, attention: the type of subsequent participation behavior and the affiliation of the community member who replies to a new user's first question.[5]

Below, we draw on existing theory to posit the expected effects of variations in from whom a response is received versus not receiving a response on each of the two types of subsequent participation behaviors. Doing so requires us to introduce heterogenous and, sometimes, multiple theories. *Heterogeneous theories* are needed, because both the independent and dependent variables studied vary across our four hypotheses. We draw on existing theory from several fields—management and organizations, marketing, social psychology, and information systems—as each has something to tell us about the relationships being investigated.[6] For several hypotheses, we draw on *multiple theories* as it is not clear which mechanism(s) are at play. Our focus is not on disentangling these mechanisms; when possible, however, we use a combination of logic and examination of additional patterns in the data to suggest which mechanisms are most likely at play.[7] Table 2 provides an overview of the relationships studied (columns 1, 2), potential mechanisms and the literature behind them (columns 3-6), our findings (column 7), and our reflections on the mechanisms most likely at play given what we can observe with our data (column 8).

[Insert Table 2 here]

[5] The community we study has coalesced around a commercially produced product, the Stata statistical software package. We use the term community *member* to refer to all participants in a community. We differentiate between participants who are not formally affiliated with the company (*users*) and participants who are employed by Stata (*employees*).
[6] Interest in these relationships from across fields is indicative of widespread interest in the effects of social interactions on knowledge sharing in online communities.
[7] A single mechanism may dominate across individuals, different mechanisms may apply to different individuals, and/or the behavior of some individuals may be shaped by multiple mechanisms. Empirically, however, we can only observe the aggregated effects of the multiple potential mechanisms over multiple individuals.

To our knowledge, none of our four hypotheses has been specifically examined, however scholars have examined loose variants of our hypotheses pertaining to the effects of receiving a reply from another user on new users' subsequent participation (Hypotheses 1 and 3). Specifically, related to Hypothesis 1, scholars have examined the effects of receiving a reply on subsequent participation in general (that is, not differentiating between asking and replying behaviors); and, related to Hypotheses 1 and 3, scholars have documented *intended* rather than *actual* behaviors of all (existing and new) users in a community. We, following the example set by Lampe and Johnston (2005) and Joyce and Kraut (2006), examine patterns using data on actual behavior. The hypotheses pertaining to the effects of receiving a reply from an employee on a new user's subsequent participation have, to our knowledge, not been examined (Hypotheses 2 and 4).

Effects of Receiving a Reply from a User on Asking Another Question

Drawing on expectation-confirmation theory, we expect that receiving a reply from a user(s)—versus not receiving a reply—will increase the likelihood that new users will ask another question. Expectation-confirmation theory was initially created to explain the psychological underpinnings of consumers' repeat purchase behavior and has been applied to knowledge sharing in online communities. The theory posits that consumers form an initial expectation about a specific product or service prior to using it. As they use the product or service, they assess the extent to which their initial expectation is confirmed or disconfirmed. Satisfied consumers are likely to repurchase or continue to use the product or service, whereas dissatisfied consumers are not (Oliver 1980).

Scholars have found evidence in support of expectation confirmation theory in the online community context, using detailed survey data to show that users (both new and existing) who receive a response to their post intend to continue participating in the community (Chen 2007, Cheung and Lee 2007, Jin et al. 2009, Jin et al. 2010). However, these studies measure the intent to participate rather than actual participation. In addition, these studies do not disaggregate between subsequent participation in the form of asking questions and replying to questions (or other behaviors), focusing instead on continued use of the community in general. Finally, these studies examine the behaviors of all—both new and existing—

users. Hence, while these studies give us reason to suspect that expectation-confirmation theory is at play, additional research examining actual behavior is warranted. Here we focus on whether or not new users engage in a specific type of participation: asking another question. Following the logic embedded in expectation-confirmation theory, we suggest that new users might observe the question replying and asking activity occurring in the community and form an initial expectation that they will receive a reply should they post a question. If they do not receive a reply, their initial expectation is disconfirmed and they will be less likely to post another question. If they receive a reply, their expectation is confirmed, and they will be more likely to post another question. We therefore predict that receiving a reply from a user will increase the likelihood of new users asking another question.

Hypothesis 1. New users who post their first question and receive a reply from a user (vs. receiving no reply), will be more likely to ask another question.

Effects of Receiving a Reply from an Employee on Asking Another Question

We also use expectation-confirmation theory to predict the effect of receiving a reply from an employee(s)—versus not receiving a reply—on the likelihood that new users will ask another question. New users might observe the question replying and asking activity occurring in the community and form an initial expectation that they will receive a reply if they post a question. Receiving a reply from an employee could confirm, or even surpass, the new user's expectations, particularly if the new user perceives employees to be highly knowledgeable about the product. As a result, the new user might post another question. We therefore expect that receiving a reply from an employee will increase the likelihood of new users asking another question.

Hypothesis 2. New users who post their first question and receive a reply from an employee (vs. receiving no reply), will be more likely to ask another question.

Effects of Receiving a Reply from a User on Replying to Questions from Other Users

We now turn to understanding how receiving a reply from a user—versus not receiving a reply—will affect new users' likelihood of *replying to the questions of others*. This question has been heavily theorized, but competing theories have been proposed and extant empirical evidence is mixed and

measures intended rather than actual behaviors. We discuss what is known about each potential explanation—a norm of reciprocity and information search—below.

A Norm of Reciprocity

Many scholars have credited norms of reciprocity as helping to sustain replying behavior in online communities (Kankanhalli et al. 2005, Cheung and Lee 2007, Chiu et al. 2006, Faraj and Johnson 2011, Ren et al. 2012, Jin et al. 2013). A social norm around reciprocity dictates that individuals treat others as they have been treated, and repay in kind what has been done for them (Gouldner 1960). Extant research shows that some users believe that a norm of reciprocity—that others will help them and that they should help others—exists in their communities (Wasko and Faraj 2000, Faraj and Johnson 2011, Jin et al. 2013). Research also finds that individuals who strongly believe that a norm of reciprocity exists in their online community *intend* to share more knowledge (Cheung and Lee 2007, Jin et al. 2013). While these studies document the existence of norms and intentions, they do not examine actual behaviors that may or may not result from those norms, nor do they examine how *actually receiving help* affects subsequent participation behavior.

In contrast, other research suggests that norms of reciprocity might *not* play a role in online communities. Wasko and Faraj (2005) find no significant relationship between a belief that norms of reciprocity exist and subsequent contributions in an online community of legal professionals.[8] Moreover, scholars have pointed out that in many communities, only a few individuals both ask and reply to questions, interpreting this pattern as evidence of a lack of reciprocity around knowledge sharing in online communities (Nam et al. 2009). For example, Nam et al. (2009) examine the largest knowledge sharing community in South Korea and find that only 5.4% of users both ask and reply to questions. Taken together, empirical evidence for norms of reciprocity acting to encourage users to reply to others' questions is, at best, limited. However, if norms of reciprocity were at play, we would expect that receiving

[8] Instead, Wasko and Faraj (2005) find that reputation enhancement, domain expertise, and network position are the primary drivers of contribution. This finding resonates with other research suggesting that some individuals are more motivated than others to participate in online communities (Lerner and Tirole 2002, Roberts et al. 2006, Shah 2006). Motivation plays a critical role is shaping participation in online communities; here, however, our focus is on social interaction.

a reply from a user would lead the new user to search for questions to reply and then reply to another user's question.

Information Search

In contrast to norms of reciprocity, a mechanism based on simply searching for information– which we refer to as information search—could make new users who receive a reply *less likely* to continue reading others' questions and thus less likely to reply to others' questions in the community. According to research on information-seeking behavior, the need for a specific fact to solve a problem drives people to search for information until they acquire relevant information that satisfies their need (Wilson 1999). Research on mailing list participation in open source communities and online health communities provides observations that support such a mechanism. In a qualitative study of open source software projects, Shah (2006, p. 1006) notes: "participants regularly monitor mailing list postings for information related to their own needs. As they scan, they observe the questions of others and—due to their relatively deep understanding of one or more software modules—they find that they can provide others with assistance with very little effort. They report providing assistance due to reciprocity and/or a desire to cultivate more developers who might be able to assist them in the future" This suggests that it may be useful to think of the work involved in replying to others' questions as having two parts—finding a question to answer and actually answering the question—searching for information propels the former and a touch of reciprocity may propel the latter. Along similar lines, in a study of an online community focused on providing health related information and support, Wang et al. (2012, p. 839) find that individuals exposed to more informational support were more likely to drop out of the community. They speculate that "it may be that information needs are short term. As a result, people who have information needs and receive informational help… have these immediate needs met and have little reason to stay in the group, just as one might not continue perusing a dictionary after looking up a definition." This leads us to suggest that as users monitor the community for a reply to their question and search the community archives, they may encounter questions where their knowledge is useful and reply to those questions. Then, when the user acquires the information they are looking for, their search—and discovery of opportunities to contribute—

stops. Hence, if information search is at play, receiving a reply from a user could decrease the likelihood of subsequently replying to a question(s) posed by another user(s).

If norms of reciprocity are at play, receiving a reply from a user would *increase* the likelihood that new users subsequently reply to a question(s) posed by another user(s). If information search is at play, receiving a reply from a user would *decrease* the likelihood that new users subsequently reply to a question(s). Given the absence of direct empirical support and the presence of contradictory evidence for explanations focused on norms of reciprocity (Wasko and Faraj 2005, Nam et al. 2009), we focus on the possibility that information search is at play and expect that receiving a reply from a user will decrease the likelihood of a new user subsequently replying to a question(s) posed by another user(s).

Hypothesis 3. New users who post their first question and receive a reply from a user (vs. receiving no reply), will be less likely to subsequently reply to a question(s) posed by another user(s).

Effects of Receiving a Reply from an Employee on Replying to Questions from Other Users

The final quadrant of the 2x2 in Figure 1 considers how receiving a reply from an employee—versus not receiving a reply—impacts the likelihood that new users will reply to questions. Three potential mechanisms may be at play: information search, gift exchange, and/or the Hawthorne effect.

Information Search

The information search mechanism would function as described for Hypothesis 3: once a reply is received, the user would stop searching and hence stop encountering questions to which to reply. If information search is at play, receiving a reply from an employee would *decrease* the likelihood that new users would subsequently reply to questions posed by other users.

Gift Exchange

Consistent with classic anthropological and sociological theorizing around gift exchange (Mauss 1925), it is possible that receiving a reply from an employee might encourage new users to reply to questions posed by others. New users, expecting to receive a reply from another user (Hypothesis 1), might interpret a reply from an employee as a gift—something above and beyond what they were expecting—because employees are likely to be perceived as both highly knowledgeable and credible sources of

information. In Mauss's conceptualization, when a gift is given, it creates a debt that needs to be repaid (Mauss 1925). Thus, to the extent that new users consider these replies to be gifts, they would feel obligated to return the gift or offer a gift in return.[9] We suggest that receiving a reply from an employee may be perceived as a gift and therefore lead new users to both make an extra effort to reply to others' questions and, ideally, ultimately post a reply.

Hawthorne Effect

In line with the classic Hawthorne effect, it is also possible that receiving a reply from an employee might lead a new user to feel that they are being observed by individuals with authority and therefore reply to questions posed by others. Prior research has found that people become more productive when observed by people in authority (Blalock and Blalock 1982, Festinger and Katz 1953, McCarney et al. 2007): the psychological stimulus of being singled out and made to feel important can increase people's motivation to participate.[10] In the context of knowledge sharing in online communities, new users might then focus on increasing their community contribution by replying to others' questions.

At the individual level, the information search effect is mutually exclusive from gift exchange and the Hawthorne effect,[11] while gift exchange and the Hawthorne effect could occur jointly. Aggregating across individuals, different individuals could be motivated by different mechanisms. Overall, we propose, that the strength of gift exchange and the Hawthorne effect will override the information search mechanism, because of the potential perceived value of receiving information from an expert—the employee. We therefore expect that receiving a reply from an employee will increase the likelihood of new users subsequently replying to a question(s) posed by another user(s).

Hypothesis 4. New users who post their first question and receive a reply from an employee (vs. receiving no reply), will be more likely to subsequently reply to a question(s) posed by another user(s).

[9] Realistically, returning the gift is an unlikely outcome as it would be awkward and knowledge, as a non-excludable good, cannot be returned once shared.

[10] Along similar lines, Gallus (2017) finds that new Wikipedia contributors who receive a symbolic award—another form of being singled out and made to feel important—are more likely to make additional edits.

[11] Specifically, new users whose behavior is motivated by information search would search until they received a reply—occasionally replying to a question if they found one in their domain of expertise—and then stop if they received a reply to their question. In contrast, new users whose behaviors are motivated by gift exchange and/or the Hawthorne effect would be *even more likely* to try to reply to the questions of others after receiving a reply to their question.

Effects of Receiving a Reply from a User versus an Employee on Future Behavior

Our hypotheses have focused on comparing the effects of receiving a reply from either a user or an employee to the baseline of receiving no reply. However, it is a natural extension to consider how receiving a reply from a user might have a different effect on new users' subsequent behavior than receiving a reply from an employee. In the case of the effect on future asking behavior (the outcome variable of interest in Hypotheses 1 and 2), it is unclear what this relationship might look like, because both are expected to be positive. However, in the case of future replying behavior, Hypothesis 3 (impact of a user reply) predicts a negative result and Hypothesis 4 (impact of an employee reply) predicts a positive result. Thus, if both of these predictions are correct, we can infer that receiving a reply from an employee will have a greater positive impact on new users' likelihood of subsequently replying to a question(s) posed by other user(s). Formally,

> *Corollary 1. When new users post their first question and receive a reply from an employee (vs. receiving a reply from a user), they are more likely to subsequently reply to a question(s) posed by another user(s).*

STUDY SETTING

Knowledge Sharing in Online Communities

Online communities bring together individuals with a diverse set of needs and knowledge bases (von Hippel and von Krogh 2003, Faraj et al. 2011) and tend to focus on a specific topic of interest, from software to video games to health and recreation (Dholakia et al. 2004, Ridings and Gefen 2004, Shah and Nagle 2020). Many different types of interactions occur within online communities—from expressing curiosity and excitement about a shared interest (Dewan et al. 2017) to the buying and selling of component parts (Muñiz and Schau, 2005) to innovation (Füller et al. 2007, Shaikh and Levina 2019) to the provision of emotional support (Wang et al. 2012) to knowledge sharing (Chiu et al. 2006, Faraj and Johnson 2011). Knowledge sharing through the asking and answering of questions occurs in nearly all communities and is the focus of this study.

Statalist

We examine online community interactions on Statalist, a forum dedicated to sharing knowledge related to Stata, an integrated statistical software package created in 1985. Statalist is an ideal context for our analysis for several reasons. First, Stata provides a relatively clean context for examining our questions of interest, because interactions on Statalist focus largely on information exchange, with little general chatter or emotional support. This narrow focus is in line with our interests and helps simplify our analysis. Second, many statistical and Stata-related questions have a factually correct answer, again creating a relatively clean context for our study. Third, both voluntary community members (users) and employees of Stata post on Statalist, allowing us to examine the effects of interactions with other users, as well as with employees, on new user' subsequent participation. Fourth, employee replies appear to be relatively random: StataCorp employees are not required to contribute to Statalist and do not seem to participate in a systematic manner (we discuss this further below). This is made clear on Statalist's splash page, which states: "Statalist is run and moderated by Stata users and maintained by StataCorp."[12]

METHOD

Data and Sample

Statalist has three subforums: General, Mata, and Sandbox.[13] We focus on the General subforum, which contains the majority (98%) of posts. Figure 1 shows an example of a question and reply.

[Insert Figure 1 here]

We collected all posts made to Statalist during the fifteen-year period from June 2002 to May 2017.[14] We then manually removed posts in which no question was asked, including announcements, conference advertisements, and program updates, resulting in 241,792 posts in which a question was posed by 25,985 users. We restricted the dataset to focus on users who made their *first* post between March 2014

[12] Statalist was established by users in August 1994 and was originally run as an archived email list hosted on a Harvard University server. In March 2014, Statalist was migrated to servers maintained by StataCorp due to the retirement of the administrator and accompanying aging of the workstation that housed Statalist; however, Statalist continued to be moderated by Stata users. The format of Statalist changed to a web-based forum. Our analysis focuses on activity occurring *after* the format change; we do, however, perform robustness checks using earlier data from the mailing list.
[13] The Mata and Sandbox subforums focus on Stata's matrix language and testing the forum software, respectively
[14] Although Statalist started in 1994, historical data is only available starting in June 2002.

to November 2016, resulting in 8,632 posts/participants.[15] We set November 2016 as the end of our sampling period to allow for a 6-month window in which to gauge subsequent interactions. We then identified new users whose first contribution to the community involved posting a new question, rather than responding to someone else's question, resulting in 8,326 posts/users.[16] Finally, we excluded 169 new users (out of 8,326 total) who posted another question within 10 minutes of posting their first question.[17] As a result, our final sample includes the first question posted by 8,157 new users to Statalist from March 2014 to November 2016. We capture replies to these questions and the subsequent behavior of the focal users during this time period and the following 6 months.

Measures

Dependent Variables

We measure two behaviors that are important to the health and continued success of knowledge sharing in online communities. Our first dependent variable, *Future Questioning*, equals one if a new user posted another question within 6 months of their original question and zero otherwise. Our second dependent variable, *Future Replying*, equals one if a new user posted a reply to someone else's question within 6 months of their original question and zero otherwise. We limit our timeframe to 6 months after the first question was posted to account for the possibility that any effect from getting a reply to that first question may wear off over time and/or other community interactions could play a role. Using shorter (1 month and 3 months) and longer windows (1 year and all time) shows consistent results.[18] Most subsequent behaviors occur within a month of the first posting: 31% of new users posted another question within 6 months (30 days on average), while only 6% posted a reply to someone else's question within 6

[15] To ensure these are first-time users, we compare usernames against all names posting to Statalist since June 2002.
[16] The vast majority of users'—96.46% (8,326 of 8,632)—first post involved posting a question. The 3.54% of users whose first post involved answering a question are excluded in our main analyses, as they behave differently from others and may have different underlying characteristics. We examine these outliers in our robustness checks.
[17] Upon examination, we found two reasons that such posts were made: mistakenly posting the same question twice or posting two separate questions nearly simultaneously. Both of these actions are unrelated to our core research interest (understanding the impact of getting a reply to your question) as the additional questions they posted were either a mistake or independent of whether or not they got a reply to their first question.
[18] Not shown due to space constraints; available from the authors upon request.

months (36 days on average). As a robustness check, we use the number of subsequent questions and replies (versus a binary) as the dependent variable in a Poisson regression and find consistent results.

Explanatory Variables

Received Reply is a binary explanatory variable indicating whether or not a new user's first question received a reply from either a user or a Stata employee. It is coded as zero if the first question received no reply and coded as one if the question received a reply (from either a user or an employee) within six months.[19] Of the 8,157 new users, 78% (6,382 new users) received a reply. *Received User Reply* is a binary explanatory variable indicating whether or not a new user's first question received a reply from a user. It is coded as one if the post received a reply from a user and zero otherwise. *Received Employee Reply* is a binary explanatory variable indicating whether or not a new user's first question received a reply from a StataCorp employee.[20] It is coded as one if the post received a reply from an employee and zero otherwise. The data show that only 2% of new users (173 out of 8,157 new users) received a reply from a Stata employee. Across the time period analyzed, 29 unique employees participated in the forum,[21] predominately during working hours (8am to 6pm local time).

A single question can receive replies from multiple people, therefore, we consider *all* replies to a question up until the point the new user participates again by asking another question or replying to the question of another user, or six months, whichever comes first. As a result, the *Received User Reply* and *Received Employee Reply* variables are *not* mutually exclusive. The data show that only 1% of new users (112 out of 8,157 new users) received replies from *both* a user and an employee. To simplify this analysis, we also run a robustness check where we only consider the first reply to a question (rather than all replies),

[19] In the rare occasion where a user asks a question, but does not receive a reply to that question until *after* they have asked another question, they are included in the control group and not the treatment group. This is because, they performed the outcome measure of "subsequent participation"—asking a question—before being "treated." Therefore, their behavior of asking an additional question (without having received a reply) contributes to the baseline likelihood of asking a second question.
[20] As shown in Figure 1, users can easily recognize employee responses, because these are clearly marked as coming from a "StataCorp Employee."
[21] Out of an estimated 51-200 employees, as estimated by LinkedIn (https://www.linkedin.com/company/statacorp/about/, accessed July 6, 2021).

and find consistent results.[22] Finally, *Elapsed Time* is a continuous variable that captures the time (in minutes) from when the question is posed to the first reply.

Control Variables

We controlled for five characteristics of the first question—total word count, emotional tone, analytic score, proportion of words that are six letters or more, and the proportion of words in the post that are dictionary words—using the Linguistic Inquiry & Word Count (LIWC) text analysis program.[23] These linguistic attributes have been shown to be useful for capturing differences across written texts (Joyce and Kraut 2006). In addition, we controlled for topic-specific heterogeneities by using Latent Dirichlet Allocation (LDA) topic modeling: an analysis of 26,638 questions posted in Statalist from March 2014 to May 2017 resulted in the identification of 10 topics. Including the probability of each post fitting into a given topic in our regression models allows us to account for their potential impact on both the likelihood of receiving a reply and of subsequent participation. Lastly, we controlled for temporal shocks by including year-fixed effects in our regression models.

Estimation

We estimate the probability of our binary outcomes using Logit models, specified as follows:

$$Prob(Y = 1|\mathbf{x}) = \frac{\exp(\mathbf{x}'\beta)}{1 + \exp(\mathbf{x}'\beta)}$$

where Y refers to our binary dependent variable (either future questioning or future replying), β denotes the coefficient vector, and $\mathbf{x}$ represents the set of explanatory and control variables including year-fixed effects. To account for heteroskedasticity, we use robust standard errors.

[22] Excluding the observations that have both user and employee replies results in broadly consistent results, with the one exception of an insignificant correlation between receiving an employee reply and our instrumental variable (*i.e.,* posting during working hours) in the first stage. The insignificant correlation is likely due to reduced statistical power due to the low number of employee replies.

[23] The emotional tone is a measure of how positive or negative the sentiments in the text are, based on dictionaries of positive and negative words. The measure ranges from 0 (very negative) to 100 (very positive). The analytic score captures the degree to which the post demonstrates logical, formal, or hierarchical thinking and ranges from 0 (low analytical thinking) to 100 (high analytical thinking).

RESULTS

Table 3 provides summary statistics and correlations. Looking at whether or not new users received a reply to their first question, 78% of new users received a reply(ies) (and 22% did not receive a reply). Digging deeper, 76.12% received a reply from only users, 0.74% received a reply from only Stata employees, and 1.37% received replies from both users and Stata employees. The average time to receive a reply was about one day (28 hours and 21 minutes), although the median time was only about one hour (62 minutes).[24] Comparing the average elapsed times of getting a reply from a user versus an employee, the difference is less than 1 minute. This tells us that we can rule out the possible explanation that employees choose to respond to questions not answered by users. Looking at subsequent participation, only 34% of new users participate again, with 31% of new users posting another question within 6 months and 6% posting a reply to someone else's question within 6 months. For those who do participate again, the average time to a second contribution was 28 days and the median time was 8 days. Table 1 summarizes our hypotheses and the main results of our regression analyses.

[Insert Table 3 here]

We perform a logit analysis predicting the likelihood that a new user's post will receive a reply from other users or Stata employees (see Appendix: Table A1). The results suggest that post characteristics are related to the likelihood of receiving a reply, so we control for these characteristics in our primary regressions. Users and Stata employees respond similarly to most post characteristics, but notably differently to one characteristic: specifically, a higher percentage of *Analytic* words increases the likelihood of getting a reply from a Stata employee, while it decreases the likelihood of getting a user's reply. This result suggests that Stata employees might behave slightly differently from users when it comes to selecting posts to reply to. Because of the potential endogeneity concern resulting from this difference, we not only control for these characteristics, but also conduct a number of robustness checks to allow for a more causal interpretation of our results. These include conducting interviews with Stata employees to

[24] This variable is extremely right-skewed: 1st quartile = 20 mins, 2nd quartile = 62 mins, 3rd quartile = 256 mins.

better understand when they reply to questions on Statalist and how they choose questions to which to reply, a fixed effects analysis, an instrumental variable based approach (via a two-stage Heckman selection model), and coarsened exact matching (described below).

In Table 4, we test our primary hypotheses. As a baseline, receiving a reply from *anyone* is positively related to future questioning (Columns 1, 2, 3). Column 4 disaggregates this finding to show the effects of receiving a reply from users and employees. Consistent with Hypothesis 1 and the mechanism described by expectation-confirmation theory, receiving a reply from a user increases subsequent questioning behavior: according to the analysis of average marginal effects (Column 4), receiving a reply from a user increases the likelihood of asking another question from 31% (the sample average) to 35%, a 13% increase (*p-value*=0.004). The data do not, however, confirm Hypothesis 2: getting a reply from a Stata employee does *not* increase the likelihood of subsequent questioning in a statistically significant manner (*p-value*=0.601). Although this result does not add support for Hypothesis 2 the coefficient for the employee reply (β_{employee}=0.092) has an overlapping confidence interval with the user effect (β_{user}=0.184) and could be of a similar size, but might suffer from a power problem due to the small sample size of replies coming from employees (2% of the sample).[25]

Next, we use our data to see if we can find additional support that expectation-confirmation theory is at play. We examine whether the time elapsed between posting an initial question and receiving a response affects whether or not new users subsequently ask another question. We thus limit our analysis to those posts that received a reply. In Column 5, we find that the elapsed time between the first question and the first reply is negatively associated with future questioning (Column 5). When the elapsed time increases by 1%, the likelihood of subsequently asking another question is shown to decrease from 31% to 28%, a 10% decrease (*p-value*<0.001). The observation that new users who receive a reply sooner are more likely to post another question is consistent with the idea that expectation-confirmation drives subsequent participation.

[25] β_{employee} is not statistically different from β_{user} (*p-value*=0.617).

We now shift our focus to examining the effects of receiving a reply on future replying to someone else's question. Receiving a reply from *anyone* is negatively related to replying to others' questions (Columns 6, 7, and 8). Digging deeper, when we analyze *who* the reply comes from (Column 9)—users versus employees—we observe a stark difference in subsequent behavior. Consistent with Hypothesis 3, receiving a reply from a user has a strong negative effect. An analysis of average marginal effects (Column 9) shows that getting a user reply decreases the likelihood of responding to the question from 6% (the sample average) to 1%, about an 83% decrease (*p-value*<0.001).

Next, we use our data to conduct a supplementary analysis to reveal whether our information search mechanism or norms of reciprocity appear to be shaping users' behavior. We examine whether the *time elapsed* between posting an initial question and receiving a response is related to the probability that the user will reply to another user's question before receiving a reply to their own. We find that the greater amount of time elapsed, the greater the probability that the user will reply to another user's question before receiving a reply to their own (β=0.356; *p-value*<0.001).[26] Combined with our main finding that after getting a reply, new users are less likely to reply to a question, it is reasonable to conclude that the information search mechanism is at play: as new users wait for a reply to their first question, it appears that they continue monitoring and searching the community, replying to others' question(s), but, once they receive a reply, it appears that they stop monitoring the community and are less likely to reply to others' posts. Although it is possible that the norm of reciprocity also drives behavior for some users, if it is at play, then the effect is overshadowed by the information search mechanism.

In contrast, receiving a reply from an employee, consistent with Hypothesis 4, has a strong, positive, and statistically significant impact on new users' subsequent replying behavior. Specifically, receiving a reply from a Stata employee increases the likelihood of subsequently replying to another user from 6% to 9%, a nearly 50% increase (*p-value*=0.032).[27] Unlike the significant negative effect of elapsed

[26] Table not shown due to space constraints (available from authors).
[27] Given that only 2% of the sample receives a reply from a Stata user, the fact that this result shows up as strongly significant implies that the effect is indeed quite strong.

time on future questioning (Column 5), the time elapsed between asking a question and receiving a reply does not have a significant impact on replying to the questions of others (Column 10).[28] In aggregate, this indicates that if the information search mechanism is at play, it is overshadowed by the gift exchange and Hawthorne effect mechanisms.

[Insert Table 4 here]

We also use linear probability models for our binary dependent variables (see Appendix: Table A2) and find results consistent with the Logit models presented above. In addition, we examine the number of total posts (asking and replying to questions) that a new user makes within 6 months of posing their first question as a continuous dependent variable rather than considering this activity in only a binary manner.[29] Poisson models with robust standard errors (see Appendix: Table A3) provide results consistent with our main findings.

Addressing Endogeneity Concerns: Fixed-Effects, Interviews & Instrumental Variable Analyses

Because our explanatory variables are not randomly assigned, our regression analysis might produce biased estimates, particularly if Stata employees purposefully or inadvertently reply to questions from users who are predisposed to continue contributing to the community by replying to questions (*e.g.,* perhaps these users are more knowledgeable, confident, and/or community-oriented users and this is somehow signaled in their post) (Hypotheses 3 and 4). However, if employees reply to questions in a manner that is *not* correlated with a particular user's likelihood of subsequent participation, then our results can be interpreted in a more causal manner. To help address this concern, we employed multiple tools: we analyzed users who received a reply from a Stata employee to a later question, but not their first, using user fixed-effects; we conducted qualitative interviews with Stata employees; and we implemented an instrumental variable based approach. In addition, we utilized coarsened exact matching.

[28] The key difference of this analysis and the analysis in the previous paragraph is that this analysis examines the probability of replying to others *after* receiving a reply whereas the previous analysis examines the probability *before* receiving a reply.
[29] Three observations have extreme outcomes (greater than 200 posts). To mitigate the effect of these outliers, we excluded these observations in our analyses of future replying.

To the extent that endogeneity is an issue in that Stata employees are somehow replying to users who are predisposed to contributing more, we investigate the behaviors of users who did not receive an employee reply to their first question, but did receive an employee reply to a later question (N=2,152). If endogeneity is a substantial cause for concern, we would expect that these individuals would participate more than other users in the interim time period. Looking at the within-individual comparison using individual-fixed effects, we find results that are consistent with our main findings, thereby helping to alleviate endogeneity concerns.[30] Specifically, we find that, when a user who received only user replies to prior questions then receives a reply from an employee for the current question, the effect on future replying is positive and significant (β_{employee}=0.078; *p-value*=0.044).[31]

In addition, to help address this endogeneity concern, we corresponded with two Stata employees who replied to user questions on Statalist to better understand why they replied. We learned that there are several reasons why employees reply to community questions, however none of these reasons appear to be tied to the likelihood of the user's subsequent participation. One employee reported answering questions on Statalist as part of his onboarding activities (See Appendix: Table A4, Quote 1). Another noted that his participation tends to focus on looking for bug reports (particularly those that only a Stata employee can address in the underlying codebase), questions that cannot be answered based on existing Stata-provided documentation, or new commands; he does this in his free-time between projects. Both clearly indicated that Statalist is a user forum and that employee participation is not required and is a voluntary act above and beyond their official responsibilities (See Appendix: Table A4, Quotes 2 and 3).

Based on these replies, we believe that the endogeneity concern (that Stata employees were systematically answering the questions of users more likely to participate again) is partially allayed. However, to further address this concern, we use an instrumental variable approach. Our instrument for receiving a reply from a Stata employee consisted of whether or not a question was posted during standard

[30] To estimate the coefficients, we use linear panel models because we can difference out the fixed effects without loss of any observations in the data.

[31] Consistent with our main results, receiving a reply from a user is negatively related to future replying (β_{employe}=-0.025; *p-value*=0.008). Full results are not shown due to space constraints and are available from the authors upon request.

working hours at Stata Corp's headquarters in College Station, Texas (*i.e.,* a dummy variable for whether or not the post was made between 8am and 6pm Central Time on weekdays that were not a national holiday). We chose this instrument because it is unlikely that a user would specifically wait until these hours to post a question on Statalist; rather, users would post their question when it arose (noting that users are spread out across the world and therefore time zones). Further, our discussions with Stata employees indicated that they primarily answer questions while they are at work, rather than on their own time (*e.g.,* nights, weekends, and holidays). Therefore, whether or not the question is posted during working hours for Stata employees can be seen as uncorrelated with the likelihood of a new user's subsequent participation on Statalist. Although this instrument is not perfect, it allows us to exploit a plausibly exogenous variation in the likelihood that a user receives a reply from a Stata employee.

To implement this instrument, we use a two-stage Heckman selection model. In the first stage, a strong correlation is found between the instrument and the chance of getting a reply from an employee. As shown in Column 1 of Table 5, a participant posting their first question during Stata working hours increases the likelihood of getting a reply from a Stata employee from the baseline of 2% to 3%, about a 50% increase (*p-value*=0.002). In the second stage, we then estimated the impact of receiving a reply from an employee (Wooldridge 2010). Consistent with previous models, getting a reply from a Stata employee increases the likelihood of subsequently replying to another user(s) from 6% to 41%, about a 583% increase (*p-value*=0.001) (Column 3). Our analyses using instrumental variables suggests that not accounting for the possibility of endogeneity in our initial estimates (Table 4, Columns 4 and 9) may underestimate the impact of receiving a reply from Stata employees on new users' future replying. This attenuation towards zero in the original estimates may occur due to measurement error, which the instrumental variable can help address in a manner similar to Bloom et al. (2013).

[Insert Table 5 here]

Robustness Checks

We performed a series of robustness checks (see Appendix: Table A5). First, we re-ran our main regression analysis with an expanded sample period, from June 2002 to November 2016, including the posts in both the old mailing-list format and the new forum format. We found consistent results with 18,866 observations (Columns 1 and 7). Second, we used alternative measures for our explanatory variables. Specifically, focusing on the first reply, rather than all replies, we measured *Received Employee Reply* and *Received User Reply* depending on whether the first reply was made by a Stata employee or a user. The first reply has been shown to have an outsized effect in online discussion boards (Muchnik et al. 2013). Among 6,382 posts that received a reply, 87 posts received their first reply from a Stata employee and 6,295 received their first reply from a user. With these measures, we still found qualitatively consistent findings (Columns 2 and 8). Third, we conducted analyses only using new users who had posted a reply to other's question before they posted their first question, which we dropped in previous models as these are a small number of users and can be considered outliers. These users did indeed behave differently from other new users on multiple dimensions (Columns 3 and 9). However, there are power issues in our analyses, likely due to the small number of observations. Fourth, we conducted a survival analysis, which differentiates censored and uncensored observations to estimate the conditional probability of an event at a certain time (Columns 4 and 10). The survival analysis allows us to estimate *how quickly* subsequent participation occurs after a user receives a reply. The Cox-Hazard models were consistent with our main findings. The results suggest that getting a reply from other users is associated with new users asking questions sooner, and getting a reply from an employee is associated with new users responding to others' questions sooner. Fifth, we used bivariate probit models to conduct the instrumental variable analysis. The positive impact of Stata employees' replies on new users' subsequent participation remained consistent in these models (Columns 5 and 11). Lastly, we further investigated the impact of Stata employees' replies by using coarsened exact matching (Iacus, King, and Porro 2012). Among 6,382 posts that received a reply, we matched 173 posts replied to by Stata employees to 5,214 posts replied by users, based on the variables

that were shown to affect the likelihood of receiving a reply from Stata employees differently than the likelihood of receiving a reply from users.[32] We found consistent results (Columns 6 and 12).

We also considered additional aspects of timing and found results consistent with our primary analysis.[33] Specifically, we expanded and contracted the six-month window after a question was asked to one month, three months, and twelve months allowing shorter and longer times for replies and future behaviors. We also looked for any difference in the results over the years and although we found that the impact of getting a reply from a user on the likelihood of asking another question got slightly stronger over the years, the results were broadly consistent with our primary analysis.

We also explored whether results from different types of users are consistent. Specifically, we took advantage of the fact that participation in online communities is highly uneven (Faraj and Johnson 2011, Dahlander and Frederiksen 2012, Rullani and Haefliger 2013, Jabr and Rahman 2018), with many users contributing just once or a few times and a handful of "super-users" contributing often. In the web-forum version of Stata-list, super-users can be readily identified by novice users, because the number of posts they have made is prominently displayed with their reply. We ran an additional analysis to see if receiving a reply from a super-user has an effect more similar to regular users or employees. We identified the top 15 users in our sample; in aggregate, they responded to 2,926 total questions (35.87%). Results show that receiving a reply from a super-user had qualitatively the same impact as receiving a reply from any user when considering both future questioning and future replying behavior.[34]

Finally, we also examined if the tone of the reply had any impact on future behavior. One might anticipate that replies that were more positive would lead to higher levels of subsequent participation than negative replies, but we found that the tone of the reply did not have any significant impact on subsequent

[32] As shown in Table 4, *Topic 2, Topic 4, Topic 6, Topic 8,* and *Analytic* have different (significant) impacts on the likelihoods. For the topic variables, we use the cutpoint of 0.5 to define the coarsening because the value of those variables represents the probability of each topic.
[33] Results available from the authors upon request (not shown due to space constraints).
[34] Results available from the authors upon request (not shown due to space constraints). We structure our arguments in Hypothesis 4 based on perceived, rather than actual, response quality and source credibility and these results suggest that it may well be *perception* that is driving these effects. Specifically, if one assumes that super users provide responses of equal quality and are equally credible sources of information as employees, one would expect new users to react equivalently to super-users and employees; instead, we see new users reacting equivalently to users and super-users.

participation.[35] This finding is consistent with prior research that has also shown the tone of replies does not influence subsequent participation in a different setting (Joyce and Kraut 2006).

DISCUSSION

Online communities are dependent upon users' voluntary contributions to thrive (Jin et al. 2013, Rullani and Haefliger 2013), but few new users continue participating (Arguello et al. 2006; Joyce and Kraut 2006, Ren et al. 2012, Worthen 2008). It is therefore important to understand what drives new users to continue participating (Bhattacherjee 2001, Butler 2001, Chen 2007). Building upon extant work that recognizes the importance of social interactions as a key driver of subsequent participation (Lampe and Johnston 2005, Joyce and Kraut 2006, Zhang et al. 2013), this study digs deeper into the effects of receiving a reply on new users' subsequent participation.

Findings summary: We find that receiving a reply from another user versus not receiving a reply results in a 13% *increase* in the likelihood of new users *asking* another question, whereas receiving a reply from an employee results in a small, but insignificant increase (possibly because of the small sample size). This suggests that receiving a reply—regardless of from whom—creates satisfaction with the community as a help-seeking tool and leads to continued use of the community as such.

We find that receiving a reply from another user versus not receiving a reply results in an 83% *decrease* in the likelihood of new users *replying* to another user's question, while, conversely, receiving a reply from a Stata employee results in a 50% *increase*. Comparing these values to each other, rather than to the baseline of not receiving a reply at all, shows receiving a reply from a Stata employee versus another user leads to an 890% increase in the likelihood of responding to a users' question. Reactions to receiving help from other users versus employees are remarkably different. We discuss this difference below.

The Effects of Social Interactions on Knowledge Sharing in Online Communities

This study contributes to the literature on the effects of social interactions in online communities. The existing literature examines the effects of receiving help *from users* and on *all* future contributions to

[35] Results available from the authors upon request (not shown due to space constraints).

the community (Lampe and Johnston 2005, Joyce and Kraut 2006, Zhang et al. 2013). This literature has mixed results, possibly because of the mixing of various types of subsequent contributions.[36] We add nuance to the literature by differentiating between the affiliation of who provides help (users and employees) and the effects on different types of subsequent contribution behaviors (asking and answering questions). This nuance provides support for existing findings (Hypothesis 1) and delivers some new (Hypotheses 2 and 4) and surprising (Hypothesis 3) findings. Perhaps, most importantly, our findings suggest that online social interactions are complex and, just as in the real world, it is important to take into account with whom the interactions are occurring and the types of subsequent behavior displayed.

Revisiting Reciprocity as a Support for Knowledge Sharing in Online Communities

We contribute additional empirical evidence related to the possible effects of a norm of reciprocity on knowledge sharing (specifically on replying to questions posed by others) in online communities. Past studies have found mixed results regarding the effects of reciprocity on actual contributions (Kankanhalli et al. 2005, Wasko and Faraj 2005, Joyce and Kraut 2006), and strong results regarding the effects of a belief in the norm of reciprocity on *intended* contributions (Cheung and Lee 2007, Jin et al. 2013). This presents something of a puzzle. Examining actual contributions, we find that receiving a reply from another user *decreases* the likelihood of new users replying to someone else's question (Hypothesis 3), suggesting that norms of reciprocity alone are not enough to explain replying behavior in online communities. This negative relationship, combined with our supplemental analysis, suggests that a mechanism based on information search—where individuals find questions to which they might reply while searching for information to support their own work (and where individuals stop searching when they find the information they need)—may be at play. The information search mechanism also suggests that a touch of reciprocity is needed to compel individuals to take the time to actually answer the question they encounter; this is consistent with prior scholarship and plausible given our data.

Can Employee Participation Encourage Users to Reply?

[36] Importantly, our results show that effects may be directionally different depending on the type of subsequent behavior, and therefore studies that grouped all future behaviors may miss effects that cancel each other out in the aggregate.

This study also contributes to the small, but growing literature on firm engagement in online communities (Dahlander et al. 2008, Colombo et al. 2013, Altman et al. 2015, Nagle 2018, Shah and Nagle 2020). Prior research suggests that firms benefit from working within or sponsoring online communities (Goldman and Gabriel 2005, Jeppesen and Frederiksen 2006, Franke et al. 2008, Manchanda et al. 2015, Nagle 2018, Shah and Nagle 2020), and scholars are actively investigating how particular firm behaviors can affect user participation within communities (*e.g.,* Jeppesen and Frederiksen 2006, Shah 2006, Shaikh and Levina 2019). We find that when an employee responds to a user's question, the user's likelihood of responding to others' questions increases (Hypothesis 4). This may provide a small lever for firms interested in seeding new communities or invigorating existing ones. A caveat, however, is in order: our theorizing suggests that this benefit might hold only when receiving replies from employees is somewhat *unexpected;* future experimental studies might alter the frequency of employee participation in the community to further investigate this effect.

Practical Implications

These findings have practical implications for community members, firms, and policy-makers interested in seeding and maintaining vibrant communities: specifically, patterns of participation in online communities are nuanced and creating levers to support participation will require an understanding of these patterns. Participation by employees of a firm around whose product a community is structured appears to have a positive effect on user participation, at least at the relatively low levels present in our context. Our findings further indicate that firms cannot simply "set it and forget it" with a user community, they should play an active role in maintaining the health of the community.

Future Research

Given the important role that online communities play as forums for knowledge sharing across a wide variety of topical areas, we believe that both theory and practice will benefit from increasingly nuanced studies examining the effects of social interactions on behaviors in online communities. At a high level, our findings suggest that future research examine how various factors combine in unique and nuanced ways to generate behavior (*e.g.,* the information search mechanism combines an active search for

information with a desire to act in accordance with norms of reciprocity). Our findings, combined with those of other scholars (Kankanhalli et al. 2005, Wasko and Faraj, 2005, Shah 2006, Wang et al. 2012), also suggest that future research may want to look at factors beyond norms of reciprocity in the search for social-interaction based-levers for encouraging replying behavior.

Limitations & Generalizability

As with most studies, our study has limitations that leave open avenues for future research. Our study is, in effect, a case study of a single community. Future research is needed to see whether our results and insights are generalizable to other knowledge-sharing communities. That said, Statalist has structural similarities to many other online communities—for example, it is question and answer based, it is focused on a particular topic and most questions are answered by users—suggesting that our findings may generalize to other communities. In addition, we are unable to comment on the percentage of employee replies to questions that would be optimal in encouraging subsequent contribution in a community (the percentage in our study is 2%). Finally, although the use of observational data does not allow us to claim a fully causal relationship, we run multiple analyses in an effort to provide evidence that such a relationship is plausible.

CONCLUSION

Knowledge is increasingly being shared through online communities; and, as individuals and firms become increasingly reliant on online communities as sources of knowledge, the need to understand the dynamics of participation in these communities grows. We contribute to understanding how interactions between community members of various types shape subsequent participation, showing both that receiving a reply has different effects on new users' subsequent asking and replying behaviors, and that the affiliation of the community member who provides help—employee or user—also matters. There is much that we have to learn about communities and much that communities can teach us about organizing and human behavior (*Cf.* He, Puranam, Shrestha, von Krogh 2020); we hope that scholars continue to dig deeper into the participation dynamics of online communities.

REFERENCES

Aaltonen A, Seiler S (2016) Cumulative growth in user-generated content production: Evidence from Wikipedia. *Management Science* 62(7):2054–2069.

Ajzen I (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50(2):179–211.

Altman E, Nagle F, Tushman M (2015) Innovating without information constraints: Organizations, communities, and innovation when information costs approach zero. Hitt MA, Shalley C, Zhou J, eds. *Oxford Handbook of Creativity, Innovation, and Entrepreneurship* (Oxford University Press, Oxford, UK).

Altman E, Nagle F, Tushman ML (2021) The translucent hand of managed ecosystems: Engaging communities for value creation and capture. *Academy of Management Annals*, in press.

Arguello J, Butler BS, Joyce E, Kraut R, Ling KS, Rosé C, Wang X (2006) Talk to me: Foundations for successful individual-group interactions in online communities. *CHI 2006 Proceedings*: 959–968.

Bagozzi RP, Dholakia UM (2006) Open source software user communities: A study of participation in Linux user groups. *Management Science* 52(7):1099–1115.

Bauer J, Franke N, Tuertscher P (2016) Intellectual property norms in online communities: How user-organized intellectual property regulation supports innovation. *Information Systems Research* 27(4):724–750.

Bayus BL (2013) Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Science* 59(1):226–244.

Bhattacherjee A (2001) Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly* 25(3):351–370.

Blalock AB, Blalock HM (1982) *Introduction to Social Research*. (Prentice-Hall, Englewood Cliffs, NJ).

Bloom N, Schankerman M, Van Reenen J (2013). Identifying technology spillovers and product market rivalry. *Econometrica* 81(4):1347–1393.

Butler BS, Sproull L, Kiesler S, Kraut RE (2007) Community effort in online groups: Who does the work and why? Weisband S, eds. *Leadership at a Distance: Research in Technologically-Supported Work* (Lawrence Erlbaum, Mahwah, NJ), 171–194.

Chatterji, A.K. & Fabrizio, K.R. (2012). How do product users influence corporate invention?. *Organization Science, 23(4)*, 971-987.

Chatterji, A.K. & Fabrizio, K.R. (2014). Using users: When does external knowledge enhance corporate product innovation?. *Strategic Management Journal, 35(10)*, 1427-1445.

Chatterji, A.K. & Fabrizio, K.R. (2016). Does the market for ideas influence the rate and direction of innovative activity? Evidence from the medical device industry. *Strategic Management Journal, 37(3)*, 447-465.

Chen IY (2007). The factors influencing members' continuance intentions in professional virtual communities—a longitudinal study. *Journal of Information Science* 33(4):451–467.

Cheung CM, Lee MK (2007) What drives members to continue sharing knowledge in a virtual professional community? The role of knowledge self-efficacy and satisfaction. Zhang Z, Siekmann J, eds. *Knowledge Science, Engineering and Management. KSEM 2007. Lecture Notes in Computer Science*. (Springer, Berlin, DE), 472–484.

Chiu CM, Hsu MH, Wang ET (2006) Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems* 42(3):1872–1888.

Colombo MG, Piva E, Rossi-Lamastra C (2013) Authorising employees to collaborative with communities during working hours: When is it valuable for firms? *Long Range Planning* 46(3):236–257.

Dahlander L (2007) Penguin in a new suit: a tale of how de novo entrants emerged to harness free and open source software communities. *Industrial and Corporate Change* 16(5):913–943.

Dahlander L, Frederiksen L (2012) The core and cosmopolitans: A relational view of innovation in user communities. *Organization Science* 23(4):988–1007.

Dahlander L, Frederiksen L, Rullani F (2008) Online communities and open innovation: Governance and symbolic value creation. *Industry and Innovation* 15(2):115–123.

Dahlander L, Jeppesen LB, Piezunka H (2019) How organizations manage crowds: Define, broadcast, attract, and select. Sydow J, Berends H, eds. *Managing Inter-organizational Collaborations: Process Views (Research in the Sociology of Organizations Volume 64)* (Emerald Publishing Limited, Bingley, UK), 239–270.

Dewan S, Ho Y-J, Ramaprasad J (2017) Popularity or proximity: Characterizing the nature of social influence in an online music community. *Information Systems Research* 28(1):117–136.

Dholakia UM, Bagozzi RP, Pearo LK (2004) A social influence model of consumer participation in network-and small-group-based virtual communities. *International Journal of Research in Marketing* 21(3):241–263.

Di Gangi PM, Wasko M (2009) Steal my idea! Organizational adoption of user innovations from a user innovation community: A case study of Dell IdeaStorm. *Decision Support Systems* 48(1):303–312.

Faraj S, Johnson SL (2011) Network exchange patterns in online communities. *Organization Science* 22(6):1464–1480.
Faraj S, Jarvenpaa SL, Majchrzak A (2011) Knowledge collaboration in online communities. *Organization Science* 22(5):1224–1239.
Festinger LE, Katz DE (1953) *Research Methods in the Behavioral Sciences* (Dryden Press, New York).
Franke N, Keinz P, Schreier M (2008) Complementing mass customization toolkits with user communities: How peer input improves customer self-design. *Journal of Product Innovation Management* 25(6):546–559.
Goldman R, Gabriel RP (2005) *Innovation Happens Elsewhere: Open Source as Business Strategy* (Elsevier/Morgan Kaufman, San Francisco, CA).
Gallus J (2017) Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia. *Management Science* 63(12):3999–4015.
Gouldner AW (1960) The norm of reciprocity: A preliminary statement, *American Sociological Review* 25(2):161–178.
Gu B, Jarvenpaa S 2003. Online discussion boards for technical support: the effect of token recognition on customer contributions. *ICIS 2003 Proceedings* 10:110–120.
Gulati R, Puranam P, Tushman, M (2012) Meta-organization design: Rethinking design in interorganizational and community contexts. *Strategic Management Journal.* 33(6):571–586.
He VF, Puranam P, Shrestha YR, von Krogh G (2020) Resolving governance disputes in communities: A study of software license decisions. *Strategic Management* Journal 41(10):1837–1868.
Hwang EH, Singh PV, Argote L (2015) Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries. *Organization Science* 26(6):1593–1611.
Iacus SM, King G, Porro G (2012) Causal inference without balance checking: Coarsened exact matching. *Political Analysis* 20(1):1–24.
Jabr W, Rahman MS (2018) Can 'top reviews' save online review systems? Evidence from empirical analyses and a quasi-natural field experiment on Amazon. *Working Paper*. Available at: https://ssrn.com/abstract=3200803 or http://dx.doi.org/10.2139/ssrn.3200803.
Jeppesen LB, Frederiksen L (2006) Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science* 17(1):45–63.
Jin XL, Cheung CM, Lee MK, Chen HP (2009) How to keep members using the information in a computer-supported social network. *Computers in Human Behavior* 25(5):1172–1181.
Jin XL, Lee MK, Cheung CM (2010) Predicting continuance in online communities: model development and empirical test. *Behaviour & Information Technology* 29(4):383–394.
Jin XL, Zhou Z, Lee MK, Cheung CM (2013) Why users keep answering questions in online question answering communities: A theoretical and empirical investigation. *International Journal of Information Management* 33(1):93–104.
Joyce E, Kraut RE (2006) Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication* 11(3):723–747.
Kankanhalli A, Tan BC, Wei KK (2005) Contributing knowledge to electronic knowledge repositories: an empirical investigation. *MIS Quarterly* 29(1):113–143.
Kim AJ (2000) *Community Building on the Web* (Peachpit Press, Berkeley, CA).
Kuk G (2006) Strategic interaction and knowledge sharing in the KDE developer mailing list. *Management Science* 52(7):1031–1042.
Lampe C, Johnston E (2005) Follow the (slash) dot: Effects of feedback on new members in an online community. *Proceedings of the 2005 international ACM SIGGROUP Conference* :11–20.
Lee GK, Cole RE (2003) From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organization Science* 14(6):633–649.
Lerner J, Tirole J (2020) Some simple economics of open source. *The Journal of Industrial Economics* 50(2):197–234.
Luo J, Kaul A (2019). Private action in public interest: The comparative governance of social issues. *Strategic Management Journal* 40(4): 476–502.
Manchanda P, Packard G, Pattabhiramaiah A (2015) Social dollars: The economic impact of customer participation in a firm-sponsored online customer community. *Marketing Science* 34(3):367–387.
Mauss M (2002) *The Gift: Forms and Functions of Exchange in Archaic Societies*. (Routledge, London, UK).
McCarney R, Warner J, Iliffe S, Van Haselen R, Griffin M, Fisher P (2007). The Hawthorne Effect: A randomised, controlled trial. *BMC Medical Research Methodology* 7:30.
McGrath JE (1984) *Groups: Interaction and Performance* (Prentice-Hall, Englewood Cliffs, NJ).

Muchnik L, Aral S, Taylor SJ (2013) Social influence bias: A randomized experiment. *Science* 341(6146):647–651.
Muñiz Jr AM, Schau HJ (2005) Religiosity in the abandoned Apple Newton brand community. *Journal of Consumer Research* 31(4):737–747.
Nam KK, Ackerman MS, Adamic LA (2009) Questions in, knowledge in? A study of Naver's question answering community *CHI 2009 Proceedings* :779–788.
Nagle F (2018) Learning by contributing: Gaining competitive advantage through contribution to crowdsourced public goods. *Organization Science* 29(4):569–587.
Oliver RL (1980) A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research* 17(4):460–469.
O'Mahony S, Lakhani KR (2011) Organizations in the shadow of communities. *Research in the Sociology of Organizations* 33:3–36.
Porter CE, Donthu N (2008) Cultivating trust and harvesting value in virtual communities. *Management Science* 54(1):113–128.
Preece J (2000) Online communities: Designing usability, supporting sociability.
Ren Y, Harper FM, Drenner S, Terveen L, Kiesler S, Riedl J, Kraut RE (2012) Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. *MIS Quarterly* 36(3):841–864.
Ridings C, Gefen D (2004) Virtual community attraction: Why people hang out online. *Journal of Computer-Mediated Communication* 10(1): JCMC10110, https://doi.org/10.1111/j.1083-6101.2004.tb00229.x.
Roberts JA, Hann IH, Slaughter SA (2006) Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management Science 52*(7): 984–999.
Rullani F, Haefliger S (2013) The periphery on stage: The intra-organizational dynamics in online communities of creation. *Research Policy* 42(4):941–953.
Shah SK (2006) Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science* 52(7):1000–1014.
Shah SK, Nagle F (2020) Why do user communities matter for strategy? *Strategic Management Review* 1(2):305.
Shaikh M, Levina N (2019) Selecting an open innovation community as an alliance partner: Looking for healthy communities and ecosystems. *Research Policy* 48(8):103766.
Sproull L, Arriaga M (2007) Online communities. Bidogli H eds, *Handbook of Computer Networks*, *Vol. 3* (John Wiley & Sons, New York, NY).
Sun Y, Fang Y, Lim KH (2012) Understanding sustained participation in transactional virtual communities. *Decision Support Systems* 53(1):12–22.
Tiwana A, Bush AA (2005) Continuance in expertise-sharing networks: A social perspective. *IEEE Transactions on Engineering Management* 52(1):85–101.
Visa Developer Community (2021) Available at: https://community.developer.visa.com.
von Hippel E, von Krogh G (2003) Open source software and the "private-collective" innovation model: Issues for organization science. *Organization Science* 14(2):209–223.
von Krogh G, von Hippel E (2006) The promise of research on open source software. *Management Science* 52(7):975–983.
Wang YC, Kraut R, Levine JM (2012) To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* :833–842.
Wasko MM, Faraj S (2000) "It is what one does": Why people participate and help others in electronic communities of practice. *The Journal of Strategic Information Systems* 9(2-3):155–173.
Wasko MM, Faraj S (2005) Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly* 29(1):35–57.
Wicker AW (1969) Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues* 25(4):41–78.
Wilson TD (1999) Models in information behaviour research. *Journal of Documentation* 55(3):249–270.
Wooldridge JM. (2010). *Econometric Analysis of Cross Section and Panel Data (2nd Edition)* (The MIT Press, Cambridge, MA).
Worthen B (2008) Why most online communities fail. *Wall Street Journal*. (July 16), http://blogs.wsj.com/biztech/2008/07/16/ why-most-online-communities-fail/.
Wu CG, Gerlach JH, Young CE (2007) An empirical analysis of open source software developers' motivations and continuance intentions. *Information & Management* 44(3):253–262.
Zhang C, Hahn J, De P (2013) Continued participation in online innovation communities: Does community response matter equally for everyone? *Information Systems Research* 24(4):1112–1130.

Table 1. Summary of Hypotheses & Findings

		Effect on New Users' Subsequent Behaviors	
		Asking Additional Questions	*Replying to Others' Questions*
Who Responds to New Users' First Question? (baseline = no reply)	*User*	H1: Positive ✓	H3: Negative ✓
	Employee	H2: Positive *not significant*	H4: Positive ✓

Table 2. Theory & Findings: Effects of Receiving vs. Not Receiving a Response on New Users' Subsequent Participation Behaviors

Response to First Question Received From:	**Subsequent Participation Behavior**	**Potential Mechanisms & Their Effects**			**Evidence from This Study**	**Reflections on Potential Mechanisms**
		Theory	**Existing Empirical Support in the Context of Knowledge Sharing in Online Communities**	**Predicted Effect**		
Another *user*	*Asking additional questions*	Expectation-Confirmation (*e.g.,* Oliver 1980)	Chen 2007, Cheung and Lee 2007, Jin et al. 2009, Jin et al. 2010	Positive	Positive (H1)	Evidence supports mechanism
An *employee*		Expectation-Confirmation (*e.g.,* Oliver 1980)	Chen 2007, Cheung and Lee 2007, Jin et al. 2009, Jin et al. 2010	Positive	Positive, but insignificant (H2)	May be at play
Another *user*	*Replying to others' questions*	Norm of Reciprocity (*e.g.,* Gouldner 1960)	Joyce and Kraut 2006, Cheung and Lee 2007, Jin et al. 2013	Positive	Negative (H3)	Unlikely to be at play. However, could be at play and overshadowed by information search
		Information Search (*e.g.,* Wilson 1999)	NA	Negative		Evidence supports mechanism
An *employee*		Information Search (*e.g.,* Wilson 1999)	NA	Negative	Positive (H4)	Logically likely to be at play, but overshadowed by gift exchange and/or reciprocity due to the unexpected and/or valuable nature of receiving knowledge from an employee
		Gift Exchange (*e.g.,* Mauss 1925)	N/A	Positive		Unable to distinguish between these theories based on the data, but one or both likely at play
		Hawthorne Effect (*e.g.,* Festinger and Katz 1953)	N/A	Positive		

Table 3. Descriptive Statistics and Correlations

Variable	Obs.	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12
Future Questioning (Binary)	8,157	0.310	0.462	0	1												
Future Questioning (Count)	8,157	0.860	2.415	0	74	0.532											
Future Replying (Binary)	8,157	0.065	0.246	0	1	0.202	0.245										
Future Replying (Count)	8,157	0.316	5.377	0	306	0.065	0.176	0.224									
Received Reply	8,157	0.782	0.413	0	1	0.061	0.050	-0.077	-0.006								
Received User Reply	8,157	0.775	0.417	0	1	0.060	0.049	-0.087	-0.015	0.979							
Received Employee Reply	8,157	0.021	0.144	0	1	-0.007	-0.013	0.037	0.058	0.078	-0.045						
Word Counts	8,157	176.84	145.92	14	4,198	0.012	-0.004	0.017	-0.001	-0.062	0.011	-0.064					
Analytic	8,157	78.545	15.563	3.55	99	-0.026	-0.023	0.001	-0.004	-0.083	0.003	-0.082	0.155				
Tone	8,157	53.728	25.893	1	99	0.009	0.016	-0.003	0.003	0.025	-0.035	0.031	-0.185	0.005			
Six Letter Words (%)	8,157	19.431	5.911	0	42.61	-0.069	-0.066	-0.008	-0.017	-0.136	0.013	-0.135	0.040	0.114	-0.057		
Dictionary Words (%)	8,157	65.754	11.066	4.19	89.51	-0.008	0.010	-0.010	-0.001	0.068	-0.017	0.070	-0.231	-0.245	0.128	0.195	
Elapsed Time (min)	6,382	1,701	13,969	0	248,080	-0.121	-0.110	-0.028	-0.025		0.046	-0.065	0.113	0.051	-0.059	0.168	-0.032

Note: *Elapsed Time* is calculated based on 6,382 observed posts that received a reply (from either a user or employee), and thus the correlation between *Elapsed Time* and *Received Reply* is undefined.

Table 4. Impact of Receiving a Reply on Subsequent Participation (Logit Models)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DV:	*Future Questioning*					*Future Replying*				
Received Reply	0.332***	0.289***	0.195***			-0.665***	-0.686***	-0.757***		
	(0.061)	(0.062)	(0.064)			(0.096)	(0.098)	(0.104)		
Received User Reply				0.184***					-0.815***	
				(0.063)					(0.102)	
Received Employee Reply				0.092					0.515**	
				(0.176)					(0.239)	
Elapsed Time (logged)					-0.126***					-0.055
					(0.016)					(0.034)
Word Count		0.000*	0.000	0.000	0.000		0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
Analytic		-0.003	-0.005***	-0.005***	-0.003*		-0.002	-0.000	-0.000	0.000
		(0.002)	(0.002)	(0.002)	(0.002)		(0.003)	(0.003)	(0.003)	(0.004)
Tone		0.001	0.001	0.001	0.001		-0.000	-0.000	0.000	0.001
		(0.001)	(0.001)	(0.001)	(0.001)		(0.002)	(0.002)	(0.002)	(0.002)
Six Letters		-0.022***	-0.009*	-0.009*	-0.009		-0.013	-0.000	-0.000	0.001
		(0.004)	(0.005)	(0.005)	(0.005)		(0.008)	(0.009)	(0.009)	(0.011)
Dictionary Words		0.000	-0.004	-0.004	-0.004		0.001	-0.000	-0.000	0.001
		(0.002)	(0.003)	(0.003)	(0.003)		(0.004)	(0.005)	(0.005)	(0.006)
Constant	-0.994***	-0.412*	-0.323	-0.315	0.502	-2.028***	-1.713***	-2.616***	-2.611***	-2.990***
	(0.068)	(0.235)	(0.293)	(0.292)	(0.331)	(0.104)	(0.437)	(0.561)	(0.560)	(0.564)
Topic Effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Year-Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,157	8,157	8,157	8,157	6,382	8,157	8,157	8,157	8,157	6,382

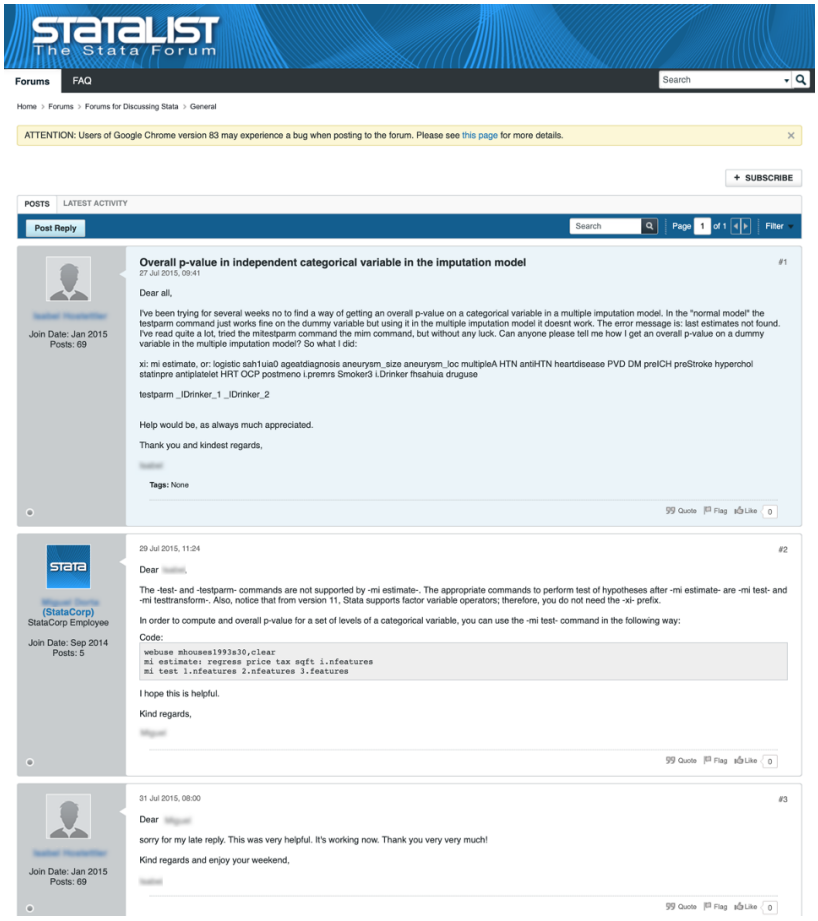
Note: ***p<.01, **p<.05, *p<.1. All standard errors are heteroskedastic robust.
Columns 5 and 10 are a sub-sample analysis with the observations of new users who received a reply from either another user(s) or an employee(s).

Table 5. Impact of Receiving a Reply on Subsequent Participation (IV Regression)

	(1)	(2)	(3)
DV:	*Received Employee Reply*	*Future Questioning*	*Future Replying*
Posted in Working Hours	0.217***		
	(0.069)		
Received Employee Reply		0.124	0.346***
		(.213)	(0.107)
Constant	-1.991***	0.446***	0.016
	(0.366)	(0.061)	(0.033)
Controls	Yes	Yes	Yes
Topic Effects	Yes	Yes	Yes
Year-Fixed Effects	Yes	Yes	Yes
Wald Chi2	186.14	366.39	222.90
Observations	8,157	8,157	8,157

Note: ***p<.01, **p<.05, *p<.1. All standard errors are heteroskedastic robust.

Figure 1. An Example of a Question and Reply in the Statalist Forum



APPENDIX

Table A1. Predicting the Likelihood of Getting a Reply (Logit Models)

	(1)	(2)	(3)
DV:	*Received Reply*	*Received User Reply*	*Received Employee Reply*
Year 2015	-0.104 (0.072)	-0.093 (0.071)	-0.048 (0.176)
Year 2016	-0.087 (0.072)	-0.038 (0.071)	-0.941*** (0.227)
Topic 1	-0.068 (0.247)	-0.013 (0.244)	-1.514** (0.650)
Topic 2	2.082*** (0.260)	2.206*** (0.258)	-4.204*** (0.809)
Topic 3	-1.478*** (0.247)	-1.397*** (0.244)	-2.184*** (0.743)
Topic 4	0.906*** (0.254)	0.922*** (0.251)	-1.049* (0.602)
Topic 5	-0.309 (0.243)	-0.286 (0.240)	-0.758 (0.564)
Topic 6	0.793*** (0.248)	0.861*** (0.245)	-3.436*** (1.003)
Topic 7	0.290 (0.269)	0.397 (0.266)	-3.770*** (0.837)
Topic 8	0.906*** (0.272)	0.918*** (0.268)	-1.764** (0.715)
Topic 9	0.890*** (0.255)	0.729*** (0.249)	0.637 (0.495)
Word Count	-0.001*** (0.000)	-0.001*** (0.000)	0.001 (0.000)
Analytic	-0.013*** (0.002)	-0.013*** (0.002)	0.010* (0.005)
Tone	-0.000 (0.001)	0.000 (0.001)	-0.009*** (0.003)
Six Letters	-0.026*** (0.005)	-0.025*** (0.005)	0.018 (0.014)
Dictionary Words	0.002 (0.003)	0.002 (0.003)	0.008 (0.008)
Constant	2.543*** (0.332)	2.445*** (0.328)	-3.519*** (0.841)
Observations	8,157	8,157	8,157

Note: ***p<.01, **p<.05, *p<.1. All standard errors are heteroskedastic robust.

Table A2. Impact of Receiving a Reply on Subsequent Participation (Linear Probability Models)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DV:	*Future Questioning*					*Future Replying*				
Received Reply	0.068***	0.059***	0.038***			-0.046**	-0.048***	-0.051***		
	(0.012)	(0.012)	(0.012)			(0.008)	(0.008)	(0.008)		
Received User Reply				0.036***					-0.056***	
				(0.012)					(0.008)	
Received Employee Reply				0.019					0.047*	
				(0.036)					(0.025)	
Elapsed Time (logged)					-0.025***					-0.003*
					(0.003)					(0.002)
Word Count		0.000*	0.000	0.000	0.000		0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
Analytic		-0.001	-0.001***	-0.001***	-0.001*		-0.000	-0.000	-0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
Tone		0.000	0.000	0.000	0.000		-0.000	-0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
Six Letters		-0.005***	-0.002*	-0.002*	-0.002		-0.001	-0.000	-0.000	0.000
		(0.001)	(0.001)	(0.001)	(0.001)		(0.000)	(0.001)	(0.001)	(0.001)
Dictionary Words		-0.000	-0.001*	-0.001*	-0.001		0.000	-0.000	-0.000	0.000
		(0.001)	(0.001)	(0.001)	(0.001)		(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.272***	0.397***	0.415***	0.417***	0.585***	0.111***	0.130***	0.079**	0.081***	0.049
	(0.014)	(0.051)	(0.061)	(0.061)	(0.061)	(0.009)	(0.028)	(0.031)	(0.031)	(0.032)
Topic Effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Year-Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,157	8,157	8,157	8,157	6,382	8,157	8,157	8,157	8,157	6,382

Note: ***p<.01, **p<.05, *p<.1. All standard errors are heteroskedastic robust.
Columns 5 and 10 are a sub-sample analysis with the observations of new users who received a reply from either another user(s) or an employee(s).

Table A3. Impact of Receiving a Reply on Subsequent Participation (Poisson Models for Count DVs)

	(1)	(2)	(3)	(4)	(5)	(6)
DV:	*Future Questioning*			*Future Replying*		
Received Reply	0.180*			-0.909***		
	(0.103)			(0.268)		
Received User Reply		0.165*			-1.195**	
		(0.099)			(0.241)	
Received Employee Reply		-0.088			0.994**	
		(0.173)			(0.486)	
Elapsed Time (logged)			-0.138***			-0.100
			(0.018)			(0.083)
Constant	-0.055	-0.037	0.908***	-3.664	-3.599	-4.836***
	(0.325)	(0.324)	(0.346)	(1.318)	(1.320)	(1.371)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Topic Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,157	8,157	6,382	8,154	8,154	6,379

Note: ***p<.01, **p<.05, *p<.1. All standard errors are heteroskedastic robust.
Three extreme outliers are excluded for the analyses of Column 4, 5, and 6.

Table A4: Interview Quotes

Quote #	Quote
1	*Part of the training for statistical developers is to participate in answering tech-support questions. Our replies are reviewed. After our training is completed, we are encouraged to participate on Statalist. We usually ask a senior developer to review a reply before posting it to Statalist (Developer 1).*
2	*I mainly look for posts that point out when Stata is not working properly. I used to try to answer more involved programming questions, but they tend to take too much time away from some of my other responsibilities…. **I tend to participate more on Statalist when I'm waiting for a task or when I am between projects**… All StataCorp developers are encouraged to participate on Statalist. Some developers are more systematic about when they browse the forum. Others are like me, peeking in from time to time (Developer 1).*
3	*We have a company policy that Statalist is supposed to be a user forum and not a StataCorp directed forum. While many of us look at the posts to see if there is anything interesting, **we do not respond to posts unless there is a bug report, a report of Stata doing something correctly, but in an odd fashion, or if there is a question about a computation which cannot be answered from our documentation**… We **also respond if we have written unofficial commands** (a.k.a. community-contributed software) about which someone has a question. Most of my replies have fallen into this last case. Otherwise **there isn't much of a decision process** that goes on (Developer 2).*[*]

[*] To clarify, this statement implies there are certain types of questions this Stata employee is more likely to answer: namely questions where it appears that a user would have a difficult time implementing a solution if there was indeed a problem (and not a misunderstanding). This makes sense as a way for this employee to focus on questions that s/he, as a Stata employee, can uniquely answer. This decreases our worries of there being an endogeneity concern related to which questions Stata employees answer. For this fact to increase rather than decrease the endogeneity concern, one would have to believe that such questions were more frequently asked by users who are more likely to reply to the questions of others. However, this is unlikely to be the case (*e.g.,* unofficial commands are widely used by Stata users of all levels, which limits the likelihood that when a Stata employee answers a user's question, that user is more or less likely to answer the questions of others. Likewise, if there was an underlying problem with the Stata software itself, encountering such a problem and posting a question about it is unlikely to be correlated with an increased likelihood of replying to others' questions). Our user fixed-effects analysis helps to further rule out this concern.

Table A5. Impact of Receiving a Reply on Subsequent Participation (Robustness Checks)

	(1)	(2)	(3)	(4)	(5)	(6)
DV:	*Future Questioning*					
Received User Reply	0.203***	0.523***	-0.536**	0.122**		
	(0.053)	(0.070)	(0.272)	(0.055)		
Received Employee Reply	0.001	0.516**	1.852**	0.021	0.692	0.011
	(0.155)	(0.246)	(0.766)	(0.139)	(1.571)	(0.036)
Constant	-0.240	-0.641**	0.611		-0.122	0.494***
	(0.211)	(0.293)	(1.509)		(0.179)	(0.082)
Model	Logit	Logit	Logit	Cox-Hazard	Bivariate Probit	OLS (CEM)
Controls & Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,866	8,157	301	8,157	8,157	5,387

Note: ***p<.01, **p<.05, *p<.1. All standard errors are heteroskedastic robust.

	(7)	(8)	(9)	(10)	(11)	(12)
DV:	*Future Replying*					
Received User Reply	-0.744***	-0.144	-1.222***	-0.804***		
	(0.082)	(0.116)	(0.285)	(0.098)		
Received Employee Reply	0.539***	0.937***	0.379	0.463**	1.433**	0.069**
	(0.202)	(0.299)	(0.684)	(0.220)	(0.601)	(0.025)
Constant	-1.526***	-3.148***	-0.240		-1.847***	0.017
	(0.306)	(0.563)	(1.645)		(0.258)	(0.037)
Model	Logit	Logit	Logit	Cox-Hazard	Bivariate Probit	OLS (CEM)
Controls & Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,886	8,157	301	8,157	8,157	5,387

Note: ***p<.01, **p<.05, *p<.1. All standard errors are heteroskedastic robust.