Stages of Motivation for Contributing User-Generated Content: A Theory and Empirical Test

Kevin Crowston

Syracuse University School of Information Studies

348 Hinds Hall

Syracuse, NY 13244 USA

crowston@syr.edu

Isabelle Fagnot

Audencia Business School

8 Route de la Jonelière

44300 Nantes France

ifagnot@audencia.com

Stages of Motivation for Contributing User-Generated Content: A Theory and Empirical Test

> "I've always been only a Wikipedia reader, never a Wikipedia editor. Over the years, Wikipedia has greatly benefitted me with scads of information about every topic under the sun. However, the prospect of editing the thing seemed scary and mysterious—I mean, who are these people anyway? How does one become an encyclopedia editor? —but there it was, a big honkin' typo staring at me. I was suddenly seized by the responsibility—obligation, really—to fix it. So I took the plunge and hit that edit button.
>
> So began my love affair with editing Wikipedia. It turns out editing an article isn't scary at all. It's easy, surprisingly satisfying and can become obsessively addictive."
>
> – Gina Trapani, editor of Lifehacker[1]

1 INTRODUCTION

Internet-based information and communication technologies (ICT) supporting online community spaces and shared information resources have made possible a new mode of coordinated effort, open online communities for user-generated content (UGC). Signal features of this phenomenon include:

1. large numbers of distributed contributors, commensurate with the popularity of the activity, ranging from dozens to tens of thousands or more;
2. mostly unpaid contributions; and
3. jointly-focused activity, in which contributors collectively develop new content (e.g., text, images or software) of value to a larger audience.

[1] From http://lifehacker.com/133747. Included with permission from the author

Wikipedia is the most dramatic though not unique example of UGC. This online encyclopedia has expanded rapidly to more than 40 million articles in more than 290 languages, with a huge number of contributions from voluntary contributors who develop and edit content for the site: more than 10 million edits from over 2 million active contributors in September 2016 alone[2].

The purpose of this research is to propose and test a novel theory of the motivation of contributors of UGC projects to contribute to a project. By *motives*, we mean factors that increase the probability that an individual will make a contribution. By *contribution*, we mean the effort that is given by individual volunteers to create the collective good produced by the project, such as articles or text for Wikis and blogs; software, documentation, bug reports or tests results for free/libre open source software (FLOSS) development; or videos or photos on sites such as YouTube or Flickr. The focus of the paper is on positive contributions that occur either by adding to a collective output or editing contributions for the benefits of the project. We do not address the question of motives for (or ways to discourage) negative contributions, such as Wikipedia vandalism. Nor do we theorize about the quality of contributions, i.e., to distinguish why some contributions may be more or less popular.

The main contribution of the paper is to argue and empirically show that what was previously considered a single, static and individual phenomenon, namely motivation for contribution to UGC, is in fact three separate but interrelated phenomena with separate motives for initial, sustained and meta-contribution (i.e., contributions that structure and enable further contributions (Bryant et al., 2005)).

[2] From http://stats.wikimedia.org/EN/TablesDatabaseEdits.htm and http://stats.wikimedia.org/EN/TablesWikipediansContributors.htm

2 CONCEPTUAL FRAMEWORK

In this section, we develop our theory by identifying and organizing motives for contribution to user generated content. The model is based on one proposed by Crowston and Fagnot (2008). Our first and primary contribution is to distinguish motives that operate at different individual stages of contribution to UGC projects. Distinguishing different stages of individual contribution acknowledges the common observation that the distribution of contributions to UGC is quite skewed, with a few people doing most work, and most people doing little or none. For example, Mockus et al. (2000), in their study of the development of the Apache web server, observed that the top 15 contributors (out of 388 total) contributed over 83% of modification requests and 66% of problem reports. On Wikipedia, only 25% of registered users have edited 10 times or more, and 2.4% of users have contributed 80% of the edits[3]. Arazy et al. (2017) found that 89% of Wikipedia editors were active only in a single article. Skewed distributions are not restricted to online settings: Reed and Selbee (2001) state that "in Canada in 2000, 18% of adults were responsible for 80% of all money donated to organized charities, 9% accounted for 80% of hours volunteered and 21% accounted for 65% of civic participation."

However, despite its ubiquity, this skewed pattern of contribution seems not to have been considered in prior work on motivations in voluntary collaborations. An exception is Preece and Shneiderman (2009), who noted a possible progression of participation in online groups from "reader to leader" characterized by different activities and motives at each stage. Studies of motivation generally assume that all contributors are alike, either in theorizing about motivations

[3] http://stats.wikimedia.org/EN/TablesWikipediaEN.htm

or in empirical study, e.g., statistical analyses of motivation that expect a volunteer with thousands of contributions to simply have more of the motives than a volunteer with one.

To address these skewed distributions, our model distinguishes three stages of contributions, which we label initial, sustained and meta-contribution. We propose an overall framework for synthesizing diverse motives for contribution, but then differentiate motives that are relevant for the individual at the different stages, resulting in three distinct models of motivations.

Of course, the volume of contribution varies continuously across members of a project, so any grouping into distinct categories is a theoretical abstraction. However, we argue that the three proposed stages of contribution do exhibit distinct patterns of involvement with different motivations, making the theoretical abstraction meaningful. That is, we explicitly argue that the motivations to make a first contribution are not the same as the motivations to make additional contributions: it is not simply the case that sustained contributors have higher levels of the motivations that impel an initial contribution. Similarly, the motivations for making meta-contributions are not just more of the motivations to contribute in other ways. In line with our basic argument—that motivation for contribution to UGC is actually a set of interrelated phenomena—we draw on different theories to explain contributions at different stages. Specifically, we incorporate theories of helping behaviour (Schwartz and Howard, 1982) and social movements theory (Klandermans, 1997).

We start with helping behaviour. As noted above, contributions to UGC are mostly unpaid. As a result, we are interested in the phenomenon of voluntary participation from virtual team members in UGC and view UGCs as a form of voluntary organization, that is to say, "an activity that produces goods and services at below market rate" (Wilson, 2000). Wilson (2000) describes

volunteering as "part of a cluster of helping behaviours, entailing more commitment than spontaneous assistance but narrower in scope than the care provided to family and friends" (p. 215). Given this view, we use the theory of helping behaviours to structure our analysis of motives for contribution. Research on helping behaviours suggests that such behaviour results from the satisfaction of four precursor conditions (Schwartz and Howard, 1982):

1. First, an individual must recognize a need in the others to be helped. This condition, called *attention*, focuses on recognizing situational cues that suggest the need for a helping response. These situational cues vary in salience and seriousness.

2. Second, an individual must have *an impetus* to respond, arising from a combination of feelings of social obligation and/or responsibility together with a self-perceived capability to respond. The capability to respond arises from the volunteer's resources (Uslaner, 2003) and skills and knowledge relevant to the voluntary role (Wilson, 2000).

3. Third, individuals weigh the obligation and capability of helping against the social and tangible costs of doing so in a phase called *evaluation* (Schwartz and Howard, 1982). Helping has some costs but may also have benefits to the volunteer. Unlike much of the literature on helping behaviours that has examined crisis situations requiring quick decisions, evaluation of volunteering can be done deliberately over time.

4. Finally, in cases where individuals opt not to help the person in need, a series of psychological *defence mechanisms* occur in which the individual self-justifies why a helping response was not needed (Schwartz and Howard, 1980). Given our focus on

motives that distinguish those who decide to contribute, we do not examine this stage further in our theorizing.

2.1 Stage theories

We develop our theory as a stage theory. Most commonly used theories in group research are continuum theories rather than stage theories. Continuum theories are expressed as a set of factors that collectively predict an outcome, e.g., the probability that a person will enact a specific behaviour. Examples of such theories are the Theory of Reasoned Action (Fishbein and Ajzen, 1975) or the Theory of Planned Behaviour (Ajzen and Madden, 1986). Continuum theories are useful in explaining behaviour or in suggesting which interventions (changes in input factors) will be effective in achieving a desired outcome (e.g., a particular behaviour). However, Weinstein et al. (1998) identify several limitations of continuum theories: they do not account for the fact that variables have limits (i.e., once a threshold in some input is reached, further increases may have no further effect); they assume there is no need to match interventions to the specific situations of different people; and they assume there is no need to sequence interventions. In contrast, stage theories assume that people move through distinct stages of behaviour, with different factors being important in different stages. For instance, in a well-known stage theory, Tuckman and Jensen (1977) suggested that small group development goes through five distinct stages. In our theory, UGC team members are seen as moving from one stage of contribution to another, with different motivation relevant as they change stages.

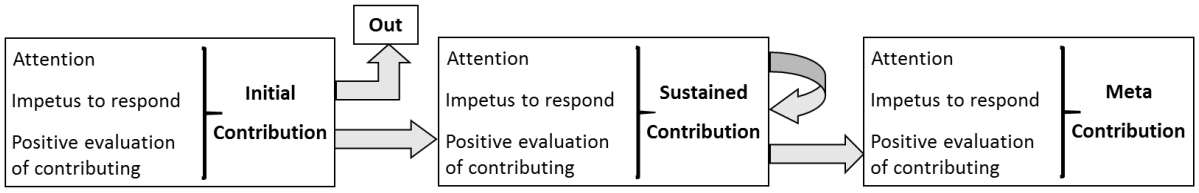
According to Weinstein et al. (1998) the requirements for a stage theory are as follows: 1) a classification system to define the stages; 2) an ordering of stages; 3) identification of common

barriers to change impacting people in the same stage (i.e., barrier to making contributions); and 4) identification of different barriers to change impacting people in different stages.

Applying these requirements to contribution to UGC, the first step was to define the stages. We define the stages in UGC by the quantity and type of contributions made and use the quantity and type of contribution as evidence that participants have changed stages. Initial contributors are the contributor's first contributions to the project; sustained are the subsequent contributions; and meta-contribution are those that structure and enable further contributions rather than directly adding to the project's output.

The quantity and type of contributions also defines the ordering of the stages: from initial to sustained to meta. Figure 1 presents our stage model and shows the progression between the different stages from initial to meta-contribution. The model shows that one set of motives prompt an initial contribution; many volunteers leave the project after that initial contribution but a few become sustained contributors, continuing their engagement, prompted by different motives; and a few become meta-contributors with still other motives.

Figure 1. Theoretical Stage Model



To develop hypotheses regarding barriers to contribution, we discuss motives for contribution in each stage, and then identify expected differences between motives for subsequent stages (i.e., hypothesizing about the differences between the 1st and 2nd stages in Figure 1, then the 2nd and

3rd stages). One could similarly develop hypotheses for the differences between initial contributors and non-contributors, but we omit these because we have no data about non-contributors with which to test such hypotheses.

2.2 Initial contributor

The first stage we consider is *initial contributions*. Contributors begin their involvement with a project with an initial contribution ("clicking the edit button" in Wikipedia, as Gina Trapani put it). It is important to consider motives for initial contributions for two reasons: first, all contributors must pass through this stage, and second, in most UGC projects, only a small fraction of users of a system actually contribute. For example, in the month of November 2016, Wikipedia was reportedly the 6th most popular site on the Web, according to Alexa[4], visited by an estimated 125 million distinct users in October 2016[5], but with only about 30 million registered accounts, of which only 126,216 had performed an action in the previous 30 days[6]. (An account is not required to edit, but anonymous edits are only a small fraction of the total in the English Wikipedia.) Comparable ratios are reported for other UGC projects. Tancer (2007) reports that fewer than 1% of visits to most user-contributed sites are contributions, with the exception of Wikipedia, where the rate at the time was reported to be 4.6%. Dahlander and McKelvey (2005) found that only 7 of 50 users of Linux surveyed had ever sent comments to the author of an application, with even lower ratios for operating system itself and for more substantive contributions; the rest were passive users. In a study of Internet mailing lists, Stegbauer and Rausch (2001) reported that the "proportion of lurkers… observed ranges between

4 http://www.alexa.com/siteinfo/wikipedia.org
5 http://siteanalytics.compete.com/wikipedia.org/?metric=uv
6 http://en.wikipedia.org/wiki/Special:Statistics

56% and 81%". As Aigrain (2003) points out, free riders are not really a problem in an information commons where the cost of reproduction is close to zero and where there may even be positive externalities of usage (Stegbauer and Rausch, 2001), but projects do need visitors to become contributors to sustain and grow the collaboration, making it important to understand the motives leading to this initial step.

Attention. According to the helping theory, the first step in volunteering is becoming aware of the project's need for help. The most basic factor for an initial contribution is simply having heard of the project at all and knowing that contribution is possible.

Impetus to respond. Once the prospective contributors become aware of the possibility of contribution, the helping theory suggests that there must be some impetus for the response based on a perceived capacity to contribute, coupled with a perceived obligation to contribute. Factors that would increase the perception of capability to contribute include particular knowledge about the domain of the UGC. For example, Bryant et al. (2005) suggest that new Wikipedia users start by correcting mistakes on topics they know (even fixing typos, as in Gina Trapani's case) or adding topics that are not covered, rather than by making big additions or corrections.

Positive evaluation of contributing. Finally, the helping theory suggests that potential contributors make an evaluation of the costs and benefits of contributing. We suggest that projects that reduce the needed effort or increase the likelihood of effort leading to a desired performance will increase motivation to contribution. For example, Bryant et al. (2005) note that the ease of editing a Wikipedia page is important in facilitating a reader's transition to being an editor. No login or registration is required and additional features are available but do not get in the way of that first step (though the increased complexity of Wikipedia has made editing more

difficult). Similarly, many blogs and other sites that aggregate user contributions make it easy for individuals to post comments.

We now consider possible benefits to participation. In the case of UGC, outcomes rarely include direct monetary or material benefit. Instead, prior research has suggested a number of non-monetary benefits. We review such extrinsic motivations (i.e., separable outcomes of doing the task, Ryan and Deci, 2000) below when we discuss sustained contribution, but note that few of these seem likely to apply to an initial contributor who is not familiar with the workings of a project or with other contributors.

Bryant et al. (2005) noted that initial users were often curious about the claim that they could just edit a page, so we suggest instead that the benefit to initial participation is simply satisfaction of curiosity about the project. Curiosity has been identified as an important part of intrinsic motivation to use a computer system (Malone, 1980), where intrinsic motivation is defined as "the doing of an activity for its inherent satisfactions rather than for some separable consequence" (Ryan and Deci, 2000). Malone (1980) separated curiosity into two components: sensory curiosity (the attention-attracting value of changes) and cognitive curiosity (the prospect of modifying higher-level cognitive structures). He argued that cognitive curiosity can be incited by indicating discrepancies or paradoxes in a learner's knowledge, which motivate the learner to learn more. The psychology of players described by Malone (1980) can be excited also by UGC projects in which contributors' curiosity might be aroused, e.g., when they notice a missing element in an article or a missing topic that deserves an article and wonder if they can edit it.

In summary, we view the decision to make an initial contribution as largely curiosity-driven ("testing the waters") driven by project visibility and the project's being easy to use with low barriers to entry.

2.3 Sustained contributor

The second stage we consider is *sustained contributors*, i.e., those who continue to contribute beyond an initial tryout of the project. Of course, there are differing degrees of sustained contribution. For example, Wikipedia authors range from occasional contributors to "high end" authors who explicitly try to improve "their" articles (Thom-Santelli et al., 2009) with the goal of having them appear as a featured article (Riehle, 2006) or those who take on responsibility for multiple articles. Nevertheless, it is striking that the majority of contributors to UGC projects do not participate past an initial trial ("One and done" in the words of McInnis et al., 2016). For example, numerous studies of FLOSS teams have found large numbers of contributors, but most contributors provide only a single contribution, such as a single bug report or modification request (Howison et al., 2006). Similarly, there were more than 1 million Wikipedia contributors in October 2016[7], but the median number of edits was only 1, meaning that most members drop out of contributing at the moment they start. The large gap between initial and sustained contributors is difficult to explain with a continuum theory that assumes that sustained and initial contributors have the same motivations, just at different intensities. We argue instead that motivations for initial and sustained contribution need to be conceptualized separately. Understanding the distinct motivations of sustained contributors is also practically important

[7] http://stats.wikimedia.org/EN/TablesWikipediansContributors.htm

because sustained contributors account for the bulk of contributions to most projects, as discussed above.

As noted above, we are building a stage model, which requires developing hypotheses regarding barriers to movement from stage to stage (the "a" hypotheses below). We do so by identifying expected differences between motives for initial and sustained contributors. In addition, we hypothesize about which motives will prompt increased contribution from sustained contributors (the "b" hypotheses).

Attention. Again, the first step in our theory is attention. We can assume that sustained contributors are aware of the project from their initial encounter. However, we suggest that to continue contributing, a second factor is whether the contributor perceives that further contributions are needed by the project. Dahlander and McKelvey (2005) note that the second most cited reason for not contributing to a FLOSS project is that there did not seem to be a need, e.g., the software worked well enough. Therefore, we hypothesize that:

H1a: Participants who report a higher perceived project need for contributions are more likely to be sustained contributors than initial contributors.

H1b: Sustained contributors who report a higher perceived project need for contributions will contribute more.

Impetus to respond. The second step in the theory is the impetus to respond, based on a perceived capacity to respond and a feeling of obligation. Considering the first, perceived capacity, we suggest that feelings of domain knowledge that prompted an initial contribution remain important. Therefore, we hypothesize:

H2a: The level of reported domain expertise will not distinguish initial and sustained contributors.

H2b: Sustained contributors who report higher domain expertise will contribute more.

Turning to obligation, in contrast to initial contributors, we suggest that feelings of social obligation are likely key in decisions to be a sustained contributor (Schroer and Hertel, 2009). To explore motivations for these feelings of obligation, we draw on the literature on social movements, defined by Marshall (1998) as an organized effort by a group of people to effect societal change (Eyerman and Jamison, 1991). To the extent that UGC can develop the characteristics of a social movement, we suggest that they will be better able to retain and motivate participants.

Extending our overall helping theory, we draw in particular on theories suggesting that participation in a social movement happens when that individual weighs the costs and benefits of taking part in a social movement (Klandermans, 1984). Klandermans's (1997) theory of motivations (as augmented by Simon et al., 1998) suggests four distinct areas of motivation for participation in a social movement: collective motives, identification with the group or a subgroup, reward motives and social motives. We consider the first two of these motives in this section and the others in the final two sections.

Collective motivations come from the individual's evaluation of the group's goals or ideology, relevant because many or most social movements coalesce around a shared ideology. Kavanagh (2004) notes that part of the motivation for some to contribute to FLOSS was identification with a narrative of resistance to proprietary software, which may explain the finding that license choice, often a reflection of ideology, seems to affect the amount of output per developer

(Fershtman and Gandal, 2007). More specifically, Xu et al. (2009) found that stated agreement with a project's values, norms and beliefs was a strong predictor of FLOSS developer involvement. In Wikipedia, sustained contributors express feelings of agreement with the project's goals, contributing to the greater good (Bryant et al., 2005). Therefore, we hypothesize:

H3a: Participants who report agreement with the project's philosophy are more likely to be sustained contributors than initial contributors.

H3b: Sustained contributors who report agreement with the project's philosophy will contribute more.

Group or community identification means that individuals join a movement because of their feelings of being part of or wanting to contribute to a valued group. The feeling of community identity is essential to the transformation of interests (group or individual) into collective action (Gotham, 1999). Group identification differs from social motives in that social motives arise directly from interactions with other people—whether group members or not—while the group identification is a preferred state of mind based on a sense of belonging. This sense is part of the explanation for the feeling of obligation to the group that provides a motivation for sustained contribution. Ellemers et al. (2004) stated that when individuals self-identify as part of a collective, they are more inclined to work towards improving the collective and its identity and Johnson et al. (2010) suggested that commitment to the group is an important motivation for work more generally. Such identification has been found across a range of UGC projects. Bryant et al. (2005) noted that active Wikipedia contributors develop an identity in the project, e.g., by having a Wikipedia home page and use a talk page to interact with others. Ren et al. (2010) claimed that online community members who identify with the group and/or with particular

members of the group have a higher commitment and hence continued participation. Balestra et al. (2016) found that high-volume Wikipedia editors were more motivated by "reputation and social motives". Therefore, we hypothesize:

H4a: Participants who report a higher level of identification with the group are more likely to be sustained contributors than initial contributors.

H4b: Sustained contributors who report a higher level of identification with the group will contribute more.

Positive evaluation of contributing. The third step of the theory is the comparison of costs and benefits of contributing. We considered costs of contributing above, and those propositions hold for sustained contributors as well. However, we expect sustained contributors to derive benefits beyond mere satisfaction of curiosity. Curiosity can be satisfied quickly, perhaps explaining why many initial participants drop out so quickly. Therefore, we hypothesize:

H5a: Participants who report curiosity about being able to contribute as a reason to start contributing are less likely to be sustained contributors than initial contributors.

H5b: Sustained contributors who report curiosity about being able to contribute as a reason to contribute will contribute less.

To develop an alternative set of hypotheses regarding evaluation of contribution, we consider in what ways the project might be rewarding to sustained contributors. We start by considering individual extrinsic rewards for contribution, which aligns with reward motivations identified by Klandermans's (1997) theory of motivations for social movements. A commonly cited personal benefit of contribution is learning (Ghosh, 2002, 2005) as working on a UGC project provides an

opportunity for contributors to learn new skills (Ye and Kishida, 2003). Therefore, we hypothesize:

H6a: Participants who report that participation provides opportunities to learn are more likely to be sustained contributors than initial contributors.

H6b: Sustained contributors who report that participation provides opportunities to learn will contribute more.

A sometimes-overlooked motivation is the intrinsic personal satisfaction of contribution. For example, researchers studying FLOSS projects have noted factors such as personal interest (Freeman, 2007) and the enjoyment of programming. Raddick et al. (2009) reported that contributors to the Galaxy Zoo project describe an interest in astronomy and in science. Lakhani and Wolf (2005) identified as a motive for contribution to FLOSS projects the chance to feel creative, i.e., using a set of skills that may not otherwise be regularly exercised. As a result, contribution may simply be viewed as fun, providing sufficient motivation for participation in a project (Bitzer et al., 2007; Freeman, 2007; Nov, 2007). Therefore, we hypothesize:

H7a: Participants who report that participation is fun are more likely to be sustained contributors than initial contributors.

H7a: Sustained contributors who report that participation is fun will contribute more.

A final factor we consider is feedback, a factor that has been consistently noted in prior research. For instance, Bandura and Schunk (1981) claim that "consistent positive feedback should encourage high collective efficacy". In turn, feedback is essential in developing efficacy perceptions that influence goal setting (Gist, 1987). Feedback can come from the task itself, such

as the positive feedback of seeing a modified program run (Chin and Cooke, 2004) or a contribution to Wikipedia accepted.

Feedback can also come from other participants. Klandermans's (1997) social motives are based on the direct social reinforcement provided by others (e.g., praise). Feedback is typically sought by individuals (Klein, 1989); by contributors in the case of UGC projects who wish to be recognized by the user community (Jeppesen and Frederiksen, 2006). Forte and Bruckman (2005) suggest that Wikipedia authors are also rewarded by recognition in the group for their work, in informal responses or through explicit mechanisms such as a featured article or "barnstars" and other awards for contribution (Kriplean et al., 2008). We note the possibility of a virtuous cycle here: as an individual contributes, they become more visible, which increases the likelihood of feedback and thus further contributions. On the other hand, negative feedback, e.g., having a contribution rejected, is expected to be demotivating. For example, editors' feelings of territoriality towards articles could result in their rejecting others' contributions, which could have a negative effect and demotivate a contributor from editing or contributing again (Thom-Santelli et al., 2009). Therefore, we hypothesize:

H8a: Participants who report receiving positive feedback about their contributions are more likely to be sustained contributors than initial contributors.

H8b: Sustained contributors who report receiving positive feedback about their contributions will contribute more.

H9a: Participants who report receiving negative feedback about their contributions are less likely to be sustained contributors than initial contributors.

H9b: Sustained contributors who report receiving negative feedback about their contributions will contribute less.

In summary, we view the decision to continue contributing as driven by the contributor's feelings of obligation to the project, the intrinsic motivation of the task and feedback from the task and other participants. These hypotheses are summarized in the third and fourth columns of Table 1.

2.4 Meta-contributor

Finally, we turn to consideration of motivations for meta-contributors. We note that in successful UGC projects, a few contributors shift their focus from substantive contributions to what we label "*meta-contributions*," that is, contributions that structure and enable further contributions (Bryant et al., 2005). For example, on Wikipedia, a few meta-contributors structure large sections of the encyclopedia, check that the style of articles is consistent, fight vandalism or administer the Wikipedia rules, rather than writing text for specific articles. In citizen science projects, especially active volunteers may be asked to moderate discussion fora. Indeed, the presence of such structuring and the resulting coordination amongst contributors is what makes UGC projects collaborations. We discuss the motivations for these contributions, though we note that many aspects of motivations overlap the motives of active sustained contributors. Again, we develop hypotheses about the barriers to moving between stages by comparing motives at this and the previous stage (the "c" hypotheses). However, we do not offer hypotheses regarding the volume of meta-contribution, as we lack data with which to test such hypotheses.

Attention. As with sustained contributors, we note that becoming a meta-contributor starts with awareness of the project's need for this kind of work. The distinguishing characteristic of meta-contributors is that they are concerned with structure of the whole project, not just a few pieces, and with the state of the community, not just its output (Bryant et al., 2005). For example, a

meta-contributor might focus on recruiting or encouraging members with necessary skills for a project, rather than on making those contributions personally.

Projects can make the needs for meta-contributors more visible to potential meta-contributors, e.g., by making these roles explicit and having those in them providing role models to others. As a result of this change in focus, awareness of a need for regular contributions is not hypothesized as a motive for meta-contributors. Therefore, we hypothesize that:

H1c: Participants who report a higher perceived need for contributions are less likely to be meta-contributors than sustained contributors.

Impetus to respond. Regarding the second step in the helping process, we suggest that meta-contributors go through much the same evaluation as sustained contributors in determining their capacity to respond. However, rather than domain knowledge, meta-contributors must have a good knowledge of the community and its norms and rules. Therefore, we hypothesize:

H2c: Participants who report a higher level of domain expertise are less likely to be meta-contributors than sustained contributors.

As with sustained contributors, we believe that meta-contributors continue to feel a social obligation to respond based on their adoption of the project's shared ideology, though in their role of meta-contributor, they also help shape this ideology. Therefore, we hypothesize:

H3c: The level of reported agreement with the project's goals will not distinguish sustained and meta-contributors.

H4c: The level of reported identification with the group will not distinguish sustained and meta-contributors.

Positive evaluation of contributing. The third step of the theory is the comparison of costs and benefits of contributing. Considering benefits, above we hypothesized a set of individual benefits that motivate sustained contribution. While meta-contribution may still be intrinsically motivating, we suggest that individuals receive little direct personal benefit from meta-contribution, relying instead on social rewards, as discussed above. Therefore, we do not expect that the impact of these benefits will distinguish meta-contributors and so hypothesize:

H5c: The level of reported curiosity about being able to contribute will not distinguish sustained and meta-contributors.

H6c: The level of reported opportunities to learn will not distinguish sustained and meta-contributors.

H7c: The level of reported fun will not distinguish sustained and meta-contributors.

H8c: The level of reported positive feedback received will not distinguish sustained and meta-contributors.

H9c: The level of reported negative feedback received will not distinguish sustained and meta-contributors.

In summary then, we view the decision to contribute as a meta-contributor as driven by a sense of group membership, leading to feelings of obligation to the group, as well as by the intrinsic motivation of the task. These hypotheses are summarized in the final column of Table 1.

Table 1. Summary of hypotheses.

Hyp	Construct	Sustained vs. intial (a)	Amount of contribution (b)	Meta vs. sustained (c)
H1	Perceived need for contributions	>	+	<
H2	Domain expertise	=	+	<
H3	Agreement with project philosophy	>	+	=
H4	Identification with the project	>	+	=
H5	Curiosity	<	–	=
H6	Opportunities to learn	>	+	=
H7	Fun	>	+	=
H8	Positive feedback	>	+	=
H9	Negative feedback	<	–	=

3 METHODS

To test the hypotheses presented above, we analyzed data from the Wikimedia Editor Survey that was run by the Wikimedia Foundation in April 2011 (Wikimedia Foundation, 2011). This survey was the first survey done by the Wikimedia Foundation on Wikipedia editors (i.e., only logged in Wikipedia users). There were 5,073 complete responses of which 4,930 reported having edited at least once. As the study was based on a publicly-available anonymized data set collected before the start of the research, it was exempt from United State human subjects research regulations.

3.1 Data

The study from which we derived our data was a cross-sectional survey. As we were relying on previously-collected data, the initial step in the analysis was to connect the available data to the concepts included in the hypotheses developed above. This approach is common in fields like economics that rely on published data but less common in research on online systems. Had we developed our own survey, we could have developed survey items designed to precisely measure the theoretical constructs of interest, but we would likely have only a fraction of the respondents,

with increased concerns about the representativeness of the sample. Therefore, for this paper, we are making a best effort to test our theory with data that are already available.

Independent variables. The two outcome variables we are trying to predict with our theory are the stage of the contributor (initial, sustained or meta) and the quantity of contributions. The quantity of contributions was measured by the self-reported count of edits done (question D1b).

The stage of the contributor was assessed in two ways. First, among survey respondents, we considered as *initial contributors* those who had reported making fewer than 10 edits since starting with Wikipedia (question D1b), regardless of user access level. As we are arguing that the motivations for initial contributors are those that get them to contribute for the first time, conceptually the threshold should be 1 edit, but we considered that a one-time editor might make a few changes in a single editing session (consistent with Arazy et al. (2017)'s finding of a large number of editors active in only a single article). We therefore picked a slightly higher threshold, but still below what a sustained editor could be expected to contribute.

We considered as *sustained contributors* respondents reporting 10 or more edits and without special access on Wikipedia (i.e., regular registered users). Note that this category includes editors with a wide range of levels of activity.

Finally, we considered as *meta-contributors* users with a user access level higher than a regular user (e.g., administrators or other roles, question Q3). A user access system is defined as "an access control procedure for computer systems, which allows a system administrator to set up a hierarchy of users" (System access level, n.d.). Wikipedia has many user access levels with different rights and permissions. Initial and sustained contributors who add or edit contributions

to Wikipedia have the same standard user access level, while those who are granted higher levels of access can use that access to meta-contribute to the project.

We acknowledge that these operationalizations of user stage involve some arbitrary decisions, and discuss possible threats to validity caused by these decisions below.

Motives. We selected survey questions that seemed to measure factors hypothesized as motives in our theory and included in the hypotheses above. We included all the survey items in Question Q5, which asked respondents which of a set of reasons to start or to continue editing Wikipedia applied to them (e.g., to see whether anyone could edit, or for fun). These items were binary, yes/no questions. Items used to assess the hypotheses are shown in Table 2, labelled by the number of the item on the questionnaire. Note that some of the items do not map to constructs of our theory.

To measure feedback (H8 and H9), we grouped together selected responses to question Q18, regarding interactions with other editors (a number of binary items) into two variables, one for positive feedback and one for negative feedback, as shown in Table 2. The resulting variable was a count of the number of kinds of positive or negative feedback received.

Finally, we included as controls two demographic variables that have been suggested to affect participation in Wikipedia, namely age (question D2) and gender (question D10). Including gender led us to drop respondents who reported genders other than male or female, as there were too few of these to analyze. Control variables are also shown in Table 2. The questions about reasons to start or continue editing that did not map to theory constructs are listed as additional control variables.

Table 2. Mapping of constructs from hypotheses to Wikipedia Editor survey questions.

Hyp	Construct	Questions
H1	Perceived need for contributions	Reasons to start editing (Q5a): I saw an error and wanted to fix it (Q5a2) I saw a red link or noticed an article was missing, so I wrote it (Q5a3) Reasons to continue editing (Q5b): I keep finding or looking for mistakes (Q5b2) I find articles that are incomplete or biased (Q5b3)
H2	Domain expertise	Reasons to start editing (Q5a): I knew a lot about a subject that was poorly covered (Q5a4) Reasons to continue editing (Q5b): I like to contribute to subject matters in which I have expertise (Q5b4)
H3	Agreement with project philosophy	Reasons to continue editing (Q5b): I believe that information should be freely available to everyone (Q5b8) I like Wikipedia's philosophy of openness and collaboration (Q5b7)
H4	Identification with the project	Reasons to continue editing (Q5b): I want to gain a reputation within the Wikipedia community (Q5b9)
H5	Curiosity	Reasons to start editing (Q5a): I wanted to see whether anyone could edit (Q5a1)
H6	Opportunities to learn	Reasons to start editing (Q5a): I wanted to learn new skills (Q5a8)
H7	Fun	Reasons to continue editing (Q5b): It's fun (Q5b11)
H8	Positive feedback	Interactions with others (Q18): Having others compliment you on your edits/articles Having your article(s) selected as featured article(s) Article(s) making it to the front page Having your picture(s) used in articles Getting a barnstar or similar award from another editor Another editor adding content/photos to an article you are working on Having other editors add content to article(s) you started
H9	Negative feedback	Interactions with others (Q18): Other editors pushing their point of view Being looked down on by more experienced editors Having your edits reverted without any explanation Having an article that you were working on deleted

Control variables	Age Gender (female vs. male) Reasons to start editing (Q5a): My friends, family or colleagues contribute to Wikipedia (Q5a5) I wanted to demonstrate my knowledge to a wider public or community (Q5a6) I liked the idea of volunteering to share knowledge (Q5a7) I wanted to participate in a discussion on Wikipedia (Q5a9) I was assigned to edit for a school project or work (Q5a10) Reasons to continue editing (Q5b): I do it for professional reasons (Q5b1) I like to contribute to subject matters in which I have expertise (Q5b4) I want to popularize topics I care about (Q5b6) I like the idea of volunteering to share knowledge (Q5b10)

3.2 Analysis approach

To test our hypotheses, we analyzed the data set by running two regression models using Stata. The models contained the above-mentioned variables as predictors. We chose regression as the hypotheses included the number of edits as well as classification of users into stages. We note though that classification problems are often addressed with logistic regression, the analysis approach we used to predict sustained vs. initial and meta vs. sustained contributors.

The first model, a zero-inflated negative binomial regression predicted whether the respondent was an initial or a sustained contributor (Table 5), to test the "a" hypotheses, and the reported number of edits made, to test the "b" hypotheses. Zero-inflated regression is used when the data seem to have more zeros than expected, suggesting that the data come from a two-step process. The test is two regressions bundled together: first, a logistic regression, to predict whether the value is one of the excess zeros, and second, a regression to predict the count for the non-excess-zero cases. For this analysis, edits fewer than ten were recoded to zero, meaning that a prediction of zero edits (the first regression) is a prediction of an initial contributor. The outcome variable, count of edits, is a count, so the appropriate regression is a choice between Poisson or negative binomial regression. Negative binomial was chosen as the data seemed over-dispersed: as shown

in Table 3, the variance for count of edits was much greater than the mean, counter to the expectation of equality for a Poisson distribution.

The second model was also a logistic regression, as the regression was for a binary outcome variable, whether a respondent was a sustained or meta-contributor (Table 6) to test the "c" hypotheses. As noted above, we did not attempt to predict the number of edits made by the meta-contributors, as meta-contributions can take on many forms, not just editing.

4 FINDINGS

As our focus was on motivations to contribute, we dropped respondents who reported not having accounts or making no contributions. The coding for stage described above resulted in 404 initial contributors, 4060 sustained contributors and 666 meta-contributors.

In Table 3 we present descriptive statistics for the continuous variables in our study, broken down by the stage of the user. For these variables, we report mean and standard deviations. Note that age was somewhat skewed, and so square root transformed for analysis. For the yes/no questions, we report in Table 4 the number of "yes" answers to the question by user type.

Tables 5 and 6 report the results of the two regressions. As mentioned above, the first model reported in Table 5 was a zero-inflated model, which is two regressions in one. The first, the logistic to predict excess zeros, is shown in columns a, and the second, the negative binomial to predict the count of edits, is shown in columns b. The fit of the models was assessed using the Stata fitstat macro. For the first regression, the Vuong test statistic was significant, indicating that the zero-inflated model (i.e., our stage model) provides a better fit to the data than a regular negative binomial (i.e., a regular continuum model).

Table 3. Descriptive statistics for variables in study.

Variable	Initial		Sustained		Meta-		Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Count of edits	2.5	2.4	5,813	14,235	30,872	39,445	8,363	20,688
Positive feedback	0.61	1.00	2.71	1.77	4.41	1.65	2.70	1.92
Negative feedback	0.41	0.73	1.47	1.24	2.24	1.20	1.45	1.27
Age	30.7	15.3	32.2	14.2	31.9	11.99	32.0	14.0
N	404		4060		666		5130	

Table 4. Number of "yes" responses to yes/no questions by user type.

Question	Initial	Sustained	Meta-	Total yes
female	65	328	56	449
Q5a1: start, see whether anyone could edit	75	478	77	630
Q5a2: start, saw an error	149	1335	355	1839
Q5a3: start, article was missing	321	2863	480	3664
Q5a4: start, knew a lot	246	1969	348	2563
Q5a5: start, friends contribute	382	3915	639	4936
Q5a6: start, demonstrate my knowledge	286	2896	477	3659
Q5a7: start, liked to share knowledge	151	1271	204	1626
Q5a8: start, learn new skills	271	2991	520	3782
Q5a9: start, participate in a discussion	337	3644	623	4604
Q5a10: start, assigned to edit	373	3992	656	5021
Q5b1: continue, professional reasons	361	3790	619	4770
Q5b2: continue, looking for mistakes	245	1649	351	2245
Q5b3: continue, incomplete articles	257	1806	352	2415
Q5b4: continue, have expertise	197	1445	283	1925
Q5b5: continue, demonstrate my knowledge	295	2890	473	3658
Q5b6: continue, popularize topics	255	2213	366	2834
Q5b7: continue, Wikipedia's philosophy	180	1756	232	2168
Q5b8: continue, information should be free	144	1332	181	1657
Q5b9: continue, gain reputation	354	3388	502	4244
Q5b10: continue, liked to share knowledge	158	1223	169	1550
Q5b11: continue, fun	230	1602	200	2032
N (total responses)	404	4060	666	5130

Table 5. Results of zero-inflated negative binomial regression, a logistic regression predicting initial vs. sustained contribution (a hypotheses) and a negative binomial regression predicting number of edits (b hypotheses). Count of edits < 10 were recoded to zero for this analysis.

		Initial vs. sustained (a)		**Count of edits (b)**	
Hyp	**Variable**	**IRR**	**p**	**IRR**	**p**
H1	start, saw an error (Q5a2)	120.7%	0.253	70.2%***	0.000
H1	start, article was missing (Q5a3)	140.2%+	0.090	114.1%*	0.013
H1	continue, looking for mistakes (Q5b2)	289.7%***	0.000	114.2%*	0.013
H1	continue, incomplete articles (Q5b3)	147.4%*	0.020	88.6%*	0.017
H2	start, knew a lot (Q5a4)	104.0%	0.816	71.7%***	0.000
H2	continue, have expertise (Q5b4)	147.2%*	0.016	79.7%***	0.000
H3	continue, Wikipedia's philosophy (Q5b7)	83.1%	0.285	95.2%	0.360
H3	continue, information should be free (Q5b8)	98.1%	0.914	81.3%***	0.000
H4	continue, gain reputation (Q5b9)	108.7%	0.743	112.4%+	0.087
H5	start, see whether anyone could edit Q5a1)	66.5%*	0.042	98.4%	0.832
H6	start, learn new skills (Q5a8)	69.7%*	0.038	107.3%	0.206
H7	continue, fun (Q5b11)	169.6%**	0.001	136.3%***	0.000
H8	Positive feedback (Q18)	318.2%***	0.000	153.2%***	0.000
H9	Negative feedback (Q18)	150.4%***	0.000	133.4%***	0.000
	Age (sqrt transform)	126.4%	0.553	467.4%***	0.000
	Female	49.9%**	0.001	60.8%***	0.000
	start, friends contribute (Q5a5)	91.2%	0.789	191.6%***	0.000
	start, demonstrate my knowledge (Q5a6)	71.0%+	0.099	99.2%	0.905
	start, liked to share knowledge (Q5a7)	110.7%	0.562	89.0%+	0.054
	start, participate in a discussion (Q5a9)	55.0%**	0.006	61.5%***	0.000
	start, assigned to edit (Q5a10)	44.4%*	0.021	32.8%***	0.000
	continue, professional reasons (Q5b1)	76.3%	0.330	72.0%**	0.001
	continue, demonstrate my knowledge (Q5a6)	87.8%	0.546	105.4%	0.446
	continue, popularize topics (Q5b6)	101.7%	0.923	103.8%	0.477
	continue, liked to share knowledge (Q5b10)	124.2%	0.224	110.9%+	0.099

N = 4433 (4,037 non-zero and 396 zero). Maximum Likelihood Pseudo-R^2 = 0.403.
*** p < 0.001 ** p < 0.01 * p < 0.05 + p < 0.10

Table 6. Results of logistic regression predicting sustained vs. meta-contributors (c hypotheses).

Hyp	Variable	IRR	p
H1	start, saw an error (Q5a2)	65.1%***	0.000
H1	start, article was missing (Q5a3)	98.8%	0.903
H1	continue, looking for mistakes (Q5b2)	77.9%*	0.016
H1	continue, incomplete articles (Q5b3)	73.0%**	0.002
H2	start, knew a lot (Q5a4)	83.2%+	0.066
H2	continue, have expertise (Q5b4)	73.3%**	0.003
H3	continue, Wikipedia's philosophy (Q5b7)	147.8%***	0.000
H3	continue, information should be free Q5b8)	121.1%	0.104
H4	continue, gain reputation (Q5b9)	153.8%***	0.000
H5	start, see whether anyone could edit (Q5a1)	97.0%	0.837
H6	start, learn new skills (Q5a8)	77.7%*	0.027
H7	continue, fun (Q5b11)	132.8%**	0.006
H8	Positive feedback (Q18)	162.2%***	0.000
H9	Negative feedback (Q18)	117.9%***	0.000
	Age (log transform)	68.4%	0.169
	Female	111.0%	0.539
	start, friends contribute (Q5a5)	116.1%	0.536
	start, demonstrate my knowledge (Q5a6)	100.3%	0.983
	start, liked to share knowledge (Q5a7)	101.3%	0.910
	start, participate in a discussion (Q5a9)	69.1%*	0.046
	start, assigned to edit (Q5a10)	105.9%	0.886
	continue, professional reasons (Q5b1)	108.9%	0.647
	continue, demonstrate my knowledge Q5b5)	81.9%	0.122
	continue, popularize topics (Q5b6)	85.2%	0.116
	continue, liked to share knowledge (Q5b10)	99.3%	0.954

N = 4693. Pseudo-R^2 = 0.175.

Table 7. Hypotheses and support for hypotheses.

Hyp.	Motive	Initial vs. sustained (a)	Count of edits (b)	Sustained vs. meta-contributor (c)
H1	Perceived need for contributions	Supported: Two of four items predict sustained contributor; others are in predicted direction	Two of four items predict increased contribution; but two predict decreased contribution	Supported: three of four items predict sustained contributor
H2	Domain expertise	Supported: domain expertise does not predict sustained contributor	Counter to hypothesis: domain expertise predicts less contribution	Counter to hypothesis: domain expertise does not predict sustained contributor
H3	Agreement with the project's philosophy	Counter to hypothesis: agreement with philosophy does not predict sustained contributor	Counter to hypothesis: one of two items predicts less contribution	Counter to hypothesis: one of two items predicts meta-contributor
H4	Identification with the project	Counter to hypothesis: seeking to gain reputation does not predict sustained contributor	Marginally supported: seeking to gain reputation weakly predicts increased contribution	Counter to hypothesis: seeking to gain reputation predicts meta-contributor
H5	Curiosity	Supported: curiosity predicts initial contributor	Not supported: no effect of curiosity on contribution	Supported: item does not predict meta-contributor
H6	Opportunities to learn	Counter to hypothesis: learning predicts initial contributor	Not supported: no effect of learning on contribution	Supported: learning does not predict meta-contributor
H7	Fun	Supported: fun predicts sustained contributor	Supported: fun predicts increased contribution	Counter to hypothesis: fun predicts meta-contributor
H8	Positive feedback	Supported: positive feedback predicts sustained contributor	Supported: positive feedback predicts increased contribution	Counter to hypothesis: positive feedback predicts meta-contributor
H9	Negative feedback	Counter to hypothesis: negative feedback predicts sustained contributor	Counter to hypothesis: negative feedback predicts increased contribution	Counter to hypothesis: negative feedback predicts meta-contributor

The coefficients are reported as incidence rate ratios (IRR), which indicate the increased probability of a user being a sustained vs. initial user (Table 5, column a), the change in number of edits done (Table 5, column b) or the probability of a user being a meta vs. sustained contributor (Table 6) for a unit change in the input variable. Note that many of the variables are binary yes/no questions, so for these, the IRR is the predicted impact of the user answering yes to that question. For example, a user agreeing that they started editing because they saw an error is not predictive of being a sustained editor but predicts making only 70% as many edits and being only 65% as likely to be a meta-contributor. A user answering yes to the question that they continue editing because they keep finding or looking for mistakes is 3 times more likely to be a sustained editor, less likely to be a meta-editor and is predicted to make 15% more edits.

5 DISCUSSION

Table 7 summarizes the results of the regressions, showing the support or lack of support for each of the hypotheses. The results of our analysis are somewhat at odds with the details of the theory we proposed. As a result, there are several points on which the original theory or our test of the theory should be reconsidered in light of the data.

First, H1 was mostly supported: perceived need for contributions predicted that a volunteer would be a sustained contributor rather than initial or meta. This finding confirms the role of attention in motivating helping behaviour.

Second, for H2b, contrary to our prediction, we found that self-reported domain expertise was associated with fewer edits, not more. This result suggests that perceptions of ability to respond are driven by factors other than knowledge about a domain, e.g., by more general knowledge of the Wikipedia community and of how to make an effective contribution.

Third, we expected that feelings of social obligation would be an important motivation for sustained contributors. However, agreement with project philosophy (H3) in fact distinguishes meta-contributors from sustained contributors. Similarly, identification with the project (H4) was expected to predict sustained contributors, but again predicted meta-contributors. These results are surprising and suggest that ideology and group identification plays a lesser role for ordinary contributors than they do for meta-contributors, that is, that Wikipedia acts like a social movement only for contributors at higher levels.

Fourth, with regards to factors leading to a positive assessment of contribution, we expected opportunities to learn to predict sustained contributors (H6), but in fact it did not distinguish levels of contribution. It appears that learning is not as important a motive for Wikipedia as it has been reported to be for other forms of UGC, open source software development in particular. Rather, the positive evaluation seems to derive other motivations, such as a sense that contributing is fun (H7).

Finally, positive feedback was hypothesized to be equally important for sustained and meta-contributors (H8), but in fact seemed to predict meta-contributors as well as distinguishing sustained from initial contributors. Also surprising was the positive impact of negative feedback, which was assumed to be demotivating (H9). A possible explanation is the way the questions were worded on the survey, asking if these events had ever happened, not about frequency or recency. It might be that users who have been active for longer simply have accumulated more experiences, positive or negative, as well as being more advanced in the stage of participation, creating a spurious correlation. Another interpretation is that some of the negative interactions

reflect conflicts that may in fact be productive for writing better articles and so actually appreciated by experienced editors.

In summary, the empirical results paint a somewhat different picture of sustained contribution than originally hypothesized. Specifically, sustained contributors appear to be motivated by a perception that the project needs their contributions (H1); by abilities other than domain expertise (H2); by personal rather than social motives (H3 & 4); and by intrinsic enjoyment of the process of contributing (H7) rather than extrinsic factors such as learning (H6). In contrast, meta-contributors seem to be motivated by social factors (H3 & 4), as well as by intrinsic enjoyment (H7).

The demographic variables also provide some interesting results. First, the regression results show that being female reduces the likelihood of being a sustained contributor as well as the number of edits. The results show no gender difference for sustained vs. meta-contributors, but by that stage of contributions the project has already lost women participants. Understanding the root cause of this gender effect is a pressing issue for future research. Age was not a predictor of sustained vs. initial contribution but being older predicts a higher level of contributions as well as a lower likelihood of being a meta-contributor.

Finally, the regression included questions from the survey about motives for starting and continuing that did not map directly to one of the hypotheses, but which do reflect on the theory. Being assigned to edit and editing for professional reasons were both predictive of being an initial rather than sustained contributor and of reduced contribution. These results reinforce the importance of intrinsic rather than extrinsic motivations for contributing.

Despite the differences between our initial hypotheses and our findings, the main message of the paper is supported: that different motives are important for participants in different stages of contribution. The data paint a picture of initial contributions motivated largely by curiosity; sustained contributions largely by intrinsic interest; and meta-contributions increasingly by social motives. These data show that it is a mistake to assume that the motives for sustained participants are just more of whatever got them to initially participate.

5.1 Threats to Validity

As with any study, there are possible threats to the validity of our conclusions. We cover in turn threats to construct validity, to internal validity and to external validity.

5.1.1 Construct validity

Construct validity concerns the ability of the measured data to represent the construct of interest. In the current study, we had to pick variables from an already conducted survey of Wikipedia editors to test the theory. This approach has the advantage of being able to take advantage of an already administered survey with many responses. However, a disadvantage is that the survey was not designed to test the theory, so the available items are not ideally suited as measures of the constructs. In some cases, the chosen survey items may only partially cover the construct of interest. For example, we included domain expertise as an indication of an editor's ability to respond to a perceived need for an edit, but ability to respond also reflects factors such as knowledge of editing practices and available time, which the survey does not address. A further problem is that many of the survey items were yes/no questions, which do not reflect possible gradations in the intensity of the belief. As discussed above, it may be that some of the unexpected findings of our survey are due to such problems with the constructs.

A second issues is that we had to assess a respondent's stage—initial, sustained or meta-contributor—from their answers to the survey. Doing so involved some arbitrary decisions. First, we set the cut off between an initial and sustained contributor at 10 edits. To test the whether this choice affected our results, we reran the first model using a cutoff of 5 edits (the threshold used by the Wikimedia Foundation in reporting levels of participation). We obtained essentially the same results, suggesting that our findings are not overly sensitive to the choice of threshold.

Second, we identified meta-contributors as those with higher access levels than ordinary users. However, while having a higher access level may be sufficient to be a meta-contributor, it is not necessary, meaning that some of the respondents we classified as sustained contributors are actually meta-contributors. Such misclassification would reduce the observed differences between these groups, which might explain some of the negative results obtained and the low R^2 of the regression.

5.1.2 Threats to internal validity

Threats to internal validity are those that affect the conclusions drawn from the study by offering explanations for the outcome beside the independent variables. Many well-known threats to internal validity do not apply to a non-experimental study, e.g., history, maturation, instrumentation change, resentful demoralization or interaction of treatment and construct. However, the data for our study came from a self-administered survey, so there are threats to internal validity that arise from the study itself influencing the behaviours of the participants. Hypothesis guessing seems unlikely for the questions analyzed, but there could be an effect of testing or demand. For example, respondents might have inflated their answer to the question of the number of edits done or selected motives they thought would be seen as appropriate. The

former would have resulted in our treating some initial contributors as sustained contributors, reducing the observed difference, which could explain some of our negative findings. Overall though, we do not expect this threat to have a major impact. The survey was anonymous and on line, and it seems unlikely that respondents would feel a need to impress a web page.

Finally, there are threats to internal validity that arise from selection bias. Unsurprisingly, many more sustained editors answered the survey than did initial editors. It could be that the initial editors who did answer were unrepresentative of initial editors more generally, perhaps again reducing the observed difference between initial and sustained contributors.

5.1.3 Threats to external validity

Finally, external validity concerns the possibility to generalize from the study findings to other settings. The survey has a large response rate and administered by the Wikimedia Foundation, so there is reason to believe that the results apply to Wikipedia more generally. However, the survey was only of Wikipedia editors, so the findings may not apply to other kinds of UGC as hypothesized.

5.2 Implications for Future Research

Our theories have implications for both academic and practitioner communities. First, a key point is that future studies of UGC should consider different kinds of contributors separately rather than treating them as all the same. For example, surveys of motives for contribution should be careful to measure the stage of participation and to separate motives for different stages of contribution.

Having distinguished different stages of contributions, research should examine empirically how contributors move from one stage to the other. Based on our theory, we suggest that contributors are motivated to move from one stage to the other through sustained interaction with the project. However, future research should examine additional factors that explain the dynamics described above. As well, we have grouped contributors very roughly into only three categories. It may be useful to distinguish different levels of sustained contribution or to assess different roles adopted by participants (e.g., the roles identified by Arazy et al., 2017; Détienne et al., 2016). Finally, looking at teams in varying stages of progression can help us understand the phases of development, growth and maturity of a UGC project.

The Wikipedia Editor survey data offers a few opportunities for refining the analysis. First, the survey asks how often respondents engage in different kinds of activities. These reported activities might be suitable to assess which users are meta-contributors, rather than relying on user access level as a proxy. They might also be used to refine the definition of a sustained contributor. Second, there may be other survey items that could be used to further explore some of the unexpected findings discussed above, e.g., the role of ideology or community as motives for contribution.

Finally, research should consider the generalizability of the proposed theory. As noted above, skewed levels of and different stages of participation are found in many voluntary settings, not just UGC projects (Pearce, 1993). More broadly, future research can examine the interaction between UGC projects and other organizations (e.g., contributions to company-sponsored projects, Jeppesen and Frederiksen, 2006). We also expect participation to be affected by project-level variables that cannot be tested with data from a single project, but which could be explored

in a cross-project comparison. For example, it has been suggested that a low barrier to entry is hypothesized to increase initial contributions (Kraut and Resnick, 2011); but to test this hypothesis requires comparing projects with different barriers.

5.3 Implications for Practice

To the practitioner community, the framework provides an explanation of the motivations behind those who join UGC projects and their existing efforts. Knowing the importance of different motives can be of use to organizations as they work through development and implementation of virtual teams or collaborative media such as wikis in their work practices. A specific take-away is that the motives for initial and sustained contribution can be quite different, so it is not enough just to get new contributors to start using the system. Rather, different motives should be provided to attract contributors and then to convert initial to sustained contributors and increase their participation. Further, organizations (and indeed, any developer of a UGC system) should consider the need for meta-contributions and meta-contributors, e.g., by providing opportunities for meta-contribution and explicitly recognizing and rewarding those who take on such roles.

6 CONCLUSIONS

The purpose of this paper was to develop and test a theory of motives for contribution to UGC projects that distinguished different motives for three stages of contribution. Using helping behaviours as a framework and incorporating different theoretical perspectives, including stage theories, work satisfaction and social movements, we can understand the phenomenon at multiple stages. This paper proposed and tested a set of theories for understanding the motivations behind those who join and sustain such efforts. By looking at these efforts in a

broader context, we can understand how to make these efforts more fruitful for contributors and for those who benefit from their voluntary efforts.

7 ACKNOWLEDGEMENTS

This research was partially supported by the United States National Science Foundation, Grant 05-27457.

8 REFERENCES

Aigrain, P., 2003. The individual and the collective in open information communities, 16th Bled Electronic Commerce Conference, Bled, Slovenia.

Ajzen, I., Madden, T.J., 1986. Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. Journal of Experimental Social Psychology 22, 453–474. doi: 10.1016/0022-1031(86)90045-4

Arazy, O., Lifshitz-Assaf, H., Nov, O., Daxenberger, J., Balestra, M., Cheshire, C., 2017. On the" How" and" Why" of Emergent Role Behaviors in Wikipedia, Proceedings of the ACM Conference on Computer-Supported Work and Social Media, pp. 2039-2051.

Balestra, M., Arazy, O., Cheshire, C., Nov, O., 2016. Motivational determinants of participation trajectories in wikipedia, Proceedings of the International Conference on Weblogs and Social Media, pp. 535-538.

Bandura, A., Schunk, D., 1981. Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. Journal of Personality and Social Psychology 41, 586–598

Bitzer, J., Schrettl, W., Schröder, P.J.H., 2007. Intrinsic motivation in open source software development. Journal of Comparative Economics 35, 160–169

Bryant, S.L., Forte, A., Bruckman, A.S., 2005. Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia, GROUP Conference. ACM, Sanibel Island, FL.

Chin, P.O., Cooke, D., 2004. Satisfaction and coordination in virtual communities, Tenth Americas Conference on Information Systems (AMCIS), New York, New York, pp. 2699–2708.

Crowston, K., Fagnot, I., 2008. The motivational arc of massive virtual collaboration, IFIP WG 9.5 Working Conference on Virtuality and Society: Massive Virtual Communities, Lüneberg, Germany.

Dahlander, L., McKelvey, M., 2005. Who is not developing open source software? Non-users, users, and developers. Economics of Innovation and New Technology 14, 617–635

Détienne, F., Baker, M., Fréard, D., Barcellini, F., Denis, A., Quignard, M., 2016. The Descent of Pluto: Interactive dynamics, specialisation and reciprocity of roles in a Wikipedia debate. International Journal of Human-Computer Studies 86, 11–31. doi: 10.1016/j.ijhcs.2015.09.002

Ellemers, N., DeGilder, D., Haslam, S.A., 2004. Motivating individuals and groups at work: A social identity perspective on leadership and group performance. Academy of Management Review 29, 459–478

Eyerman, R., Jamison, A., 1991. Social movements: A cognitive approach. Pennsylvania State University, University Park, PA.

Fershtman, C., Gandal, N., 2007. Open source software: Motivation and restrictive licensing. International Economics and Economic Policy 4, 209–225

Fishbein, M., Ajzen, I., 1975. Belief, attitude, intention, and behavior: An introduction to theory and research. Addison-Wesley, Reading, MA.

Forte, A., Bruckman, A., 2005. Why do people write for Wikipedia? Incentives to contribute to open-content publishing, GROUP Workshop: Sustaining community: The role and design of incentive mechanisms in online systems, Sanibel Island, FL.

Freeman, S., 2007. The material and social dynamics of motivation: Contributions to Open Source language technology development. Science Studies 20, 55–77

Ghosh, R.A., 2002. Free/Libre and Open Source Software: Survey and Study. Report of the FLOSS Workshop on Advancing the Research Agenda on Free / Open Source Software. European Commission, Brussels, Belgium.

Ghosh, R.A., 2005. Understanding Free Software Developers: Findings from the FLOSS Study, in: Feller, J., Fitzgerald, B., Hissam, S., Lakhani, K. (Eds.), Making Sense of the Bazaar: Perspectives on Open Source and Free Software. MIT Press, Cambridge, MA, pp. 23–45.

Gist, M.E., 1987. Self-efficacy: Implications for organizational behavior and human resource management. Academy of Management Review 12, 472–485

Gotham, K.F., 1999. Political opportunity, community identity and the emergence of a local anti-expressway movement. Social Problems 46, 332–354

Howison, J., Inoue, K., Crowston, K., 2006. Social dynamics of free and open source team communications, in: Damiani, E., Fitzgerald, B., Scacchi, W., Scotto, M. (Eds.), IFIP WG 2.13 Working Conference on Open Source Systems. Springer, Lake Como, Italy, pp. 319–330.

Jeppesen, L.B., Frederiksen, L., 2006. Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. Organization Science 17, 45–63. doi: 10.1287/orsc.1050.0156

Johnson, R.E., Chang, C.-H.D., Yang, L.-Q., 2010. Commitment and motivation at work: The relevance of employee identity and regulatory focus. Academy of Management Review 35, 226–245

Kavanagh, J.F., 2004. Resistance as motivation for innovation: Open source software. Commun. Assoc. Inf. Syst. 13, 615–628

Klandermans, B., 1984. Mobilization and participation: Socio-psychological expansions of resource mobilization theory. Amercian Sociological Review 49, 583–600

Klandermans, B., 1997. The Social Psychology of Protest. Basil Blackwell, Oxford.

Klein, H.J., 1989. An integrated control theory model of work motivation. Academy of Management Review 14, 150–172

Kraut, R.E., Resnick, P., 2011. Building successful online communities: Evidence-based social design. MIT Press, Cambridge, MA, pp. xi, 309 p.

Kriplean, T., Beschastnikh, I., McDonald, D.W., 2008. Articulations of WikiWork: Uncovering valued work in Wikipedia through barnstars, ACM Conference on Computer supported Cooperative Work (CSCW), pp. 47–56.

Lakhani, K.R., Wolf, R.G., 2005. Why hackers do what they do: Understanding motivation and effort in free/open source software projects, in: Feller, J., Fitzgerald, B., Hissam, S., Lakhani, K.R. (Eds.), Perspectives on Free and Open Source Software. MIT Press, Cambridge, MA.

Malone, T.W., 1980. What makes things fun to learn? Heuristics for designing instructional computer games, Proceedings of the 3rd ACM SIGSMALL Symposium and the First SIGPC Symposium on Small Systems. ACM, New York, NY, USA, pp. 162–169.

Marshall, G., 1998. A Dictionary of Sociology. Oxford, Oxford.

McInnis, B.J., Murnane, E.L., Epstein, D., Cosley, D., Leshed, G., 2016. One and Done: Factors affecting one-time contributors to ad-hoc online communities, Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. ACM, pp. 609-623.

Mockus, A., Fielding, R.T., Herbsleb, J.D., 2000. A case study of Open Source Software development: The Apache server, International Conference on Software Engineering (ICSE), p. 11 pages.

Nov, O., 2007. What motivates Wikipedians? Communications of the ACM 50, 60–64

Pearce, J.L., 1993. Volunteers: The organizational behavior of unpaid workers. Routledge.

Preece, J., Shneiderman, B., 2009. The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. AIS Transactions on Human-Computer Interaction 1, 13–32

Raddick, M.J., Bracey, G., Gay, P.L., Lintott, C.J., Murray, P., Schawinski, K., Szalay, A.S., Vandenberg, J., 2009. Galaxy Zoo: Exploring the motivations of citizen science volunteers. arXiv.

Reed, P.B., Selbee, L.K., 2001. The civic core in Canada: Disproportionality in charitable giving, volunteering, and civic participation. Nonprofit and voluntary sector quarterly 30, 761–780

Ren, Y., Harper, F.M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., Kraut, R.E., 2010. Increasing Commitment to Online Communities: Designing from Theory. University of Minnesota, Minneapolis, MN.

Riehle, D., 2006. How and why Wikipedia Works: An interview with Angela Beesley, Elisabeth Bauer and Kizu Naoko, International Symposium on Wikis (WikiSym). ACM Press, pp. 3–8.

Ryan, R.M., Deci, E.L., 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary educational psychology 25, 54-67

Schroer, J., Hertel, G., 2009. Voluntary engagement in an open web-based encyclopedia: Wikipedians and why they do it. Media Psychology 12, 96–120. doi: 10.1080/15213260802669466

Schwartz, S.H., Howard, J.A., 1980. Explanations of the moderating effect of responsibility denial on the personal norm-behavior relationship. Social Psychology Quarterly 43, 441–446

Schwartz, S.H., Howard, J.A., 1982. Helping and cooperation: A self-based motivational model, in: Derlega, V.J., Grzelak, J. (Eds.), Cooperation and Helping Behavior: Theories and Research. Academic, New York, pp. 327–353.

Simon, B., Loewy, M., Stürmer, S., Weber, U., Freytag, P., Habig, C., Kampmeier, C., Spahlinger, P., 1998. Collective identification and social movement participation. Journal of Personality and Social Psychology 74, 646–658

Stegbauer, C., Rausch, A., 2001. The silent majority: "Lurkers" on mailing lists. Zeitschrift Fur Soziologie 30, 48–64

System access level, n.d., BusinessDictionary.com.

Tancer, B., 2007. Who's Really Participating in Web 2.0?, Time.

Thom-Santelli, J., Cosley, D.R., Gay, G., 2009. What's mine is mine: Territoriality in collaborative authoring, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Boston, MA, USA, pp. 1481–1484.

Tuckman, B.W., Jensen, M.A.C., 1977. Stages of small-group development revisited. Group and Organization Management 2, 419–427. doi: 10.1177/105960117700200404

Uslaner, E.M., 2003. Civic engagement in America: Why people participate in political and social life. Department of Government and Politics, University of Maryland, College Park, MD.

Weinstein, N.D., Rothman, A.J., Sutton, S.R., 1998. Stage theories of health behavior: conceptual and methodological issues. Health Psychology 17, 290

Wikimedia Foundation, 2011. The Wikipedia Editors Survey.

Wilson, J., 2000. Volunteering. Annual Review of Sociology 26, 215–240

Xu, B., Jones, D.R., Shao, B., 2009. Volunteers' involvement in online community based software development. Information & Management 46, 151 – 158. doi: 10.1016/j.im.2008.12.005

Ye, Y., Kishida, K., 2003. Toward an understanding of the motivation of Open Source Software developers, 2003 International Conference on Software Engineering (ICSE2003), Portland, OR.

9 AUTHOR BIOGRAPHIES

Kevin Crowston is a Distinguished Professor of Information Science and Associate Dean for Research in the School of Information Studies at Syracuse University. He received his Ph.D. (1991) in Information Technologies from the Sloan School of Management, Massachusetts Institute of Technology (MIT). His research examines new ways of organizing made possible by the extensive use of information and communications technology. Specific research topics include the development practices of Free/Libre Open Source Software teams and work practices and technology support for citizen science research projects, both with NSF support.

Isabelle Fagnot, Ph.D. (Syracuse University School of Information Studies, 2011) is Associate

Professor of Management of Information Systems and Director of Quality, Accreditations, Rankings at Audencia Business School. Her research interests lie in behavioural issues in Information Systems, particularly related to organizational security and to the culture of the information profession. Her research has been presented at leading scholarly conferences, such as The Americas Conference on Information Systems, The Hawaii International Conference on System Sciences and the Security Conference and published in journals such as Systèmes d'Information et Management (French Journal of Management Information Systems).