The Effects of Diversity in Global, Distributed Collectives: A Study of User Participation in Open Source Projects

Sherae Daniel
University of Pittsburgh

Ritu Agarwal
University of Maryland

Katherine Stewart
University of Maryland

December 2009
Abstract
Diversity is a defining characteristic of global collectives on the Internet. Although there is substantial evidence to suggest that diversity can have profound implications for a variety of outcomes including performance, member engagement, and withdrawal behavior, prior research has examined the effects of diversity predominantly in the context of organizational workgroups or virtual teams. In this paper we use a diversity lens to investigate the success of non-traditional virtual work groups exemplified by open source software (OSS) projects. We build on the diversity literature to propose that three types of diversity: separation, variety and disparity diversity, influence user participation in OSS projects both directly, and through an effect on contributions to the software project. To the degree that OSS projects are critically dependent on human capital, user participation is a key outcome. We instantiate the operational definitions of the three forms of diversity to the specific and unique context of open source projects. Using archival data from 337 projects hosted on SourceForge.net, we find that variety diversity, as measured by diversity in both project roles and participant tenure, positively affects user participation. These effects are partially mediated by project contributions. The impact of separation diversity, operationalized as diversity in the languages spoken by participants, is not significant. Disparity diversity, reflecting variations in participant activity, has a negative influence on user participation. We discuss how understanding the nuanced effects of different kinds of diversity in OSS development contexts both advances the literature on diversity and OSS and provides practical implications for OSS participants.

Key words: Open source software, diversity, global collectives, user participation

1. Introduction

As the Internet continues to spawn virtual communities and facilitate collaborations that extend across the globe, groups that are diverse in many aspects are ubiquitous. Escalating diversity in groups is
not confined to virtual collectives alone; it is increasingly acknowledged that organizations must be able to leverage the best talent from around the world in order to remain competitive (Gibson and Gibbs 2006; One-Ki Lee et al. 2006). Evidence of leveraging global talent is abundant in organizational practices such as outsourcing software development to offshore destinations (Levina and Vaast 2008), using virtual teams for the execution of complex projects that require diverse skills (Majchrak and Malhotra 2005), global volunteer labor forces responsible for creating strikingly successful products such as Wikipedia, and the increasing prevalence of OSS.

It is now widely accepted that the OSS development model, in which software is generally constructed by a virtual community that includes volunteers, is available for modification by anyone, and is obtainable by users for free (Scacchi et al. 2006), has fundamentally transformed traditional software development practices. OSS competes with closed and proprietary software and alters the actions of closed and propriety software organizations (Jaisingh et al. 2008). The growing maturity of OSS is evident in strong market signals from both industry and government organizations that are beginning to rely on OSS for critical infrastructures (Festa 2001; Mauri 2004). Flagship OSS products such as Linux and Apache have gained rapid acceptance; Red Hat, a company that supports Linux, was listed in Forbes magazine as one of the 25 fastest growing technology companies in America in 2008 (Ray and Murdock 2008). The accelerating and widespread adoption of OSS and the practices it embodies has also fueled the interest of software vendors; Google is actively experimenting with open source development practices, and IBM has opened some of its code for public development.

It is estimated that over 168,000 OSS projects are currently underway (Madey and Christley 2008). However, in contrast to the celebrated and highly visible instances of products such as Linux and Apache, a large number of OSS projects perform poorly (Chengalur-Smith and Sidorova 2003). Poor performance may be manifest in several ways: the project is unable to attract developers and users to participate in software construction, the product does not gain traction with users, development activity ceases before the product offers any useful functionality, or the released software has low quality. Given this variability in OSS projects outcomes, coupled with the importance of the OSS development model,
an important question arises as to why some projects are more successful than others.

We focus on the diversity of workgroups to shed light on why some projects are more successful than others. A key characteristic of settings such as OSS projects is that participants are not formally recruited to be part of the effort; rather, individuals volitionally choose to participate, driven by a variety of motivations (Roberts et al. 2006; Shah 2006). Volitional participation by volunteers unconstrained by geography suggests that OSS project participants are likely to be diverse on multiple dimensions, where diversity represents “the distribution of differences among the members of a unit with respect to a common attribute” (Harrison and Klein 2007). Significant conceptual and empirical work has shown that workgroup diversity can have profound impacts, both positive and negative, on a variety of outcomes such as performance, member engagement, withdrawal behaviors, and turnover (e.g., Van Knippenberg and Schippers 2007). Indeed, the existence and challenge of diversity is acknowledged in OSS projects: the Python open source project website notes, “The most important skill Python can teach is the delicate skill of working in a diverse group” (quoted in Ducheneaut 2005: 339). However the notion of diversity has yet to be incorporated into studies of OSS projects. In this paper we use a diversity-focused theoretical lens to understand the success of OSS projects.

Prior research has identified several factors that drive success in OSS projects. Projects that adhere to OSS ideology, use a popular programming language, or build applications for use by developers have been shown to achieve better outcomes than those that do not (Crowston and Scozzi 2002; Manenti et al. 2007; Stewart and Gosain 2006). In each of these studies the success of the project is attributed to the fact that the presence of these factors attracts the interest of the key input for OSS projects – human capital, embodied in the participants in the project. As observed by Butler et al. in the context of online communities in general, “Perhaps the most basic type of investment in an online group is active participation, in the form of creating content and consuming it” (2002; p. 7).

Following prior work that has argued that outcomes of an OSS project are contingent on its ability to attract users (e.g. Stewart et al. 2006, Grewal et al. 2006), we conceptualize the success of an OSS project in terms of user participation. We define user participation as the number of active users...
associated with the project during a limited period of time. Active users are those who participate mainly by providing user assistance, finding and reporting bugs and requesting features (Lakhani and von Hippel 2003). User participation reflects the degree to which the OSS project is able to garner user attention to apply to these project tasks, and is consequential for OSS projects for a variety of reasons. Not only do active users assist with projects tasks, they also represent an audience that motivates developers to exert effort (Lakhani and von Hippel 2003). Indeed, proprietary software development companies that interact with OSS communities believe that developing a community of users who can be active domain experts is one of the core obligations of the OSS community (Agerfalk and Fitzgerald 2008). Crowston and Howison (2006) argue that active users represent an important buffer between developers and peripheral users, insulating the former from a barrage of trivial and mundane questions. Further, a larger number of users implies a more robust potential market for consulting services or the sale of complimentary products including hardware or manuals, creating economic payoffs for developers. Finally, inasmuch as other OSS performance outcomes, including usefulness and popularity, depend on users submitting requests that generate ideas for developers about how to make the application more useful and popular, user participation is a crucial mediator in the causal process of OSS project success.

We build upon the diversity literature to theorize the distinctive effects of three forms of diversity: variety, separation, and disparity, on user participation. We further suggest that effects of variety disparity on user participation are partially mediated by information sharing as reflected in all members’ contributions to the project. Archival data from 337 open source projects hosted on SourceForge are used to test the research hypotheses. This study makes several contributions to theory and practice. Although there is an extensive literature on the impacts of diversity in workgroups (see Joshi and Roh 2009 for a recent review), and while diversity is a core defining characteristic of distributed volunteer organizations, diversity research has not conceptualized the effects of differences among participants in such settings. We extend prior research to develop a theoretical understanding of the implications of multiple types of diversity among OSS project participants, thereby illuminating the influence of participant diversity in a virtual and volunteer context. In our theoretical development we
pay special attention to the fact that the variables that are salient for participants in an OSS project are arguably distinct from those that are salient for individuals in collocated non-volunteer contexts. By doing so, we are able to offer OSS managers practical advice about managing diversity in their projects. Such guidance is increasingly useful as a greater proportion of organizational work is conducted in work structures that require managers to construct effective groups from a pool of workers that are geographically distributed and fundamentally diverse.

In the next section we provide a brief review of research on OSS project success, followed by an overview of the literature on workgroup diversity. We suggest that OSS projects are different from the traditional foci of diversity research, and discuss how diversity may be conceptualized in this context to introduce our key constructs. Following this we develop hypotheses relating various forms of diversity in OSS projects to user participation. The methods section describes the operationalization of measures based on archival data from 337 OSS projects, followed by results. We find that variety diversity has a positive direct and mediated impact on user participation. Our results further show that disparity diversity exhibits a negative impact on user participation, while separation diversity in the OSS project is not significantly related to user participation. The final section of the paper discusses the implications of the results and suggests directions for future research.

2. Theoretical Background and Prior Research

Given the growing significance of OSS, there is a robust body of research literature in this domain. We summarize key findings from prior research that are directly related to our conceptualization, beginning with a brief discussion of theoretical and empirical work that has identified factors associated with user participation. This is followed by an overview of the diversity literature, a discussion of the unique nature of the OSS context, and associated implications for the study of diversity.

2.1 OSS Projects and User Participation

Following prior work (e.g., Stewart et al. 2006), we identify OSS projects as those that develop software in a public forum and release it under the restrictions of an Open Source Initiative (OSI)-
approved license\(^1\). The key requirement of OSI licenses is that source code be available at little or no charge. The license does not require the developers to be volunteers, though many are. In a typical OSS project, software development is done by a dispersed group coordinating its efforts through computer-mediated channels with limited or no face-to-face interaction (Ducheneaut 2005). These groups include different kinds of participants, each playing distinct and important roles (Ducheneaut 2005; Ye and Kishida 2003). We focus on the difference between two key types of participants: active users and developers. We do not consider passive users (those who only use the software and do not participate in the project in other ways). Active users include individuals who participate in mailing lists, request features, report bugs, fix bugs, or perform other peripheral development tasks that do not involve making substantive code contributions (Ye and Kishida 2003). We include among developers those whom Ye and Kishida classify as project leaders, core members and active developers. This delineation separates those whose main contributions focus on developing the software code (developers) and those who do not have a large direct impact on the code, but contribute in other ways by providing feedback and engaging in discussions.

The unique nature of the OSS development process, coupled with the unprecedented success of a handful of OSS projects, has naturally spurred research focused on understanding the determinants of OSS project performance. Prior research has utilized a variety of operational definitions of performance. For instance, as noted, due to the voluntary nature of much OSS work, several authors have argued that attracting developers and motivating them to make contributions is a critical performance outcome in this context (Stewart and Gosain 2006b, Stewart et al. 2006, Crowston et al., 2003). Others have focused on the popularity of OSS software as an outcome, similar to market share or user satisfaction in studies of traditional software development contexts (Delone and McLean 1992, 2003). As in a traditional software development context, users represent the market for the sale of related products. For instance, OSS developers may sell consulting services, manuals, specialized hardware or more advanced versions of the applications while still giving the code away for free. Thus, attracting users to OSS projects can yield

\(^1\) See [www.opensource.org](http://www.opensource.org) for a detailed description of the licensing requirements.
financial opportunities for the developers. Studies focusing on success as measured by “use” have assessed outcomes such as the number of times a project’s website has been viewed (Grewal et al. 2006) or how many people subscribe to user-focused mailing lists (Stewart et al. 2006).

In addition to simply attracting more users, several authors have also highlighted the importance of users becoming engaged in an OSS project. Active user participants make important contributions such as reporting bugs, suggesting new features, or providing user-to-user assistance (Lakhani and von Hippel 2003), thereby helping to improve the software. However, beyond treating user participation as a possible means of converting users into developers (von Krogh et al., 2003; Fang and Neufeld 2009; Ducheneaut 2005), prior empirical work has not generally focused on user participation as an outcome. Further, while the importance of user engagement has been implicitly recognized, consideration of the factors driving it has been limited. This work has mostly focused on the use-based motivations for joining an OSS project – i.e., a user contributes to the project to satisfy his or her own utilitarian needs. Though not addressed in academic research, the impact of diversity on software usefulness, and thereby on use-based motives of participants, has been recognized by OSS community leaders. For example, Bruce Perens, a key figure in creating the Open Source Definition, states:

We need to acknowledge that good developers come from many different walks of life, that their motivations differ widely, and that the overall strength of an Open Source project comes from the diversity of its users and developers rather than their homogeneity. Diversity of interests leads to generality of the code, to development of useful pathways that might otherwise never be trod, and continued support of the project when business falters.  
Perens 2009

Drawing on the research that has differentiated the key roles in OSS projects and highlighted the importance of active user participants, we focus on user participation as the outcome of interest in this study. We build on the prior literature by considering how diversity may influence two major types of motivations driving user participation: use-based motives and identity-based motives (Shah 2006; Roberts et al., 2006; Bagozzi and Dholakia 2006; Fang and Neufeld 2009). Based on the large body of research demonstrating the effects of diversity on participation and outcomes in work settings, we develop
arguments that OSS user participation is likely to be impacted by the diversity within an OSS project via both theoretical mechanisms.

2.2 Diversity in Collocated and Virtual Groups

Diversity has been examined in multiple disciplines including sociology, psychology, and organizational behavior, and there is a substantial body of theoretical and empirical literature associated with it (Van Knippenberg and Schippers 2007.) Although a complete review of this work is beyond the scope of this paper, we provide a broad overview. While diversity has been defined and conceptualized in a variety of ways across studies, there is agreement that at its essence, it represents “differences” among members of an organizational unit, be it a workgroup, the top management team, or a specific department within a larger organization (Harrison and Klein 2007.) The presence of diversity is expected to influence a range of workgroup outcomes including group processes, performance, cohesiveness, creativity, innovation, social integration, withdrawal, and turnover. Van Knippenberg and Schippers (2007) note that despite the fact that diversity can exist in many different types of collectives, much of the empirical literature has focused on understanding workgroup diversity – i.e., differences within a clearly identified and bounded collection of individuals with articulated goals and tasks. It is also important to point out that a majority of the studies on diversity have been conducted in the context of collocated or face-to-face teams, with a handful of exceptions examining virtual teams (e.g., Peters and Karren 2009; Staples and Zhao 2006).

In the past three decades a number of empirical studies have theorized and tested the effects of diversity across a range of settings, workgroups, and tasks (see Horwitz and Horwitz 2007 and Joshi and Roh 2009 for recent meta-analyses of these studies). Three broad conclusions emerge from this literature. One, diversity may be characterized as a “double-edged sword” (Horwitz and Horwitz 2007) in that heterogeneity among group members can yield both positive and negative outcomes (Williams and O’Reilly 1998). On the one hand, differences are purported to result in a broader range of perspectives and knowledge becoming available to the group, thereby leading to greater creativity and innovation. On the other, differences can be a source of conflict among members, yielding dysfunctional outcomes.
A second conclusion that emerges from reviews of the diversity literature is that there is wide variation in the observed effects of diversity. The same diversity dimension may have positive effects in one study, negative effects in another, and no effect in a third (Harrison and Klein 2007; Kirkman et al., 2004). One dimension of diversity associated with equivocal findings is heterogeneity in status distinctions as indicated by pay dispersion. For example, Bloom and Michel (2002) observed that some studies found positive performance outcomes from pay dispersion, and others found greater pay dispersion to have a negative effect on performance. Similarly, the same outcome may be affected differently by different dimensions of diversity. For example, in a meta-analyses of studies related to the effects of diversity in workgroups Horwitz and Horwtiz (2007) found that although task-related diversity (acquired characteristics such as expertise and education) was related to team performance, bio-demographic diversity (innate attributes of individuals such as age, gender, and race/ethnicity) was not. This lack of consistency in findings has led researchers to suggest that a more nuanced approach to studying the impacts of diversity, one that accounts more carefully for context, is needed (Joshi and Roh 2009).

A final conclusion that can be drawn from the diversity literature is related to the conceptualization of the core construct. Scholars have noted that there are many attributes and dimensions along which a group can be diverse such as cognitive abilities, personality, age, gender, ethnicity, prior experience, attitudes, etc. (e.g. Milliken and Martins 1996, Harrison and Klein 2007). Attempts to abstract these attributes have resulted in the development of diversity typologies such as the task-oriented and relations-oriented categories proposed by Joshi and Roh (2009). The former category includes the attributes of education, function, and tenure that are typically associated with the individual’s stock of knowledge and skills, while the latter represents relatively stable individual characteristics such as gender, race, and age.

In recent work Harrison and Klein (2007) suggest that one explanation for the equivocal findings in research on diversity is that the construct has not been precisely defined at a conceptual level. They propose a theoretical typology for the study of diversity, arguing that diversity encompasses three distinct
elements: variety, separation, and disparity, with varied underlying theoretical logics responsible for observed effects. Variety diversity reflects the breadth of information that members bring to the group that is relevant to the task at hand. As proposed in theories of information processing, greater information breadth leads to higher creativity, more effective idea generation, and broader exploration of diverse perspectives on the task at hand, thereby improving performance (e.g. Ashby 1956, Jackson et al. 1995).

A second form of diversity, separation diversity indicates differences in position or opinion among unit members, likely arising from heterogeneity in values, beliefs, and attitudes. As suggested by theories of similarity and attraction, to the degree that individuals like to be with “similar others,” separation diversity should hinder performance (Schneider 1987; 1995, Locke and Horowitz 1990). Social categorization (Williams and O’Reilly 1998) likewise suggests that homogeneity among group members engenders feelings of an “in-group”, thereby facilitating the accomplishment of team tasks. Finally, disparity diversity reflects situations where some members are associated with more power and resources than others. Following from theories of relative deprivation, the unequal distribution of resources has detrimental effects on group outcomes by creating dysfunctional situations such as increased competition among team members, domination and suppression of “voice” by powerful members, and poor communication quality (Pfeffer 1988, Pfeffer and Davis-Blake 1992). Harrison and Klein (2007) argue that to the degree that each aspect of diversity is associated with distinct underlying processes and represents a fundamentally unique dimension of differences, they should not be conflated in studies of disparity. In particular, they urge researchers to carefully consider what type of diversity a particular attribute represents in the context of a specific study.

In summary, prior research has underscored the significance of diversity as a key antecedent of group processes and outcomes. Recent theorizing in diversity emphasizes the importance of clear conceptualizations of the construct (Harrison and Klein 2007), the need to incorporate an understanding of context (Joshi and Roh 2009), and more granular theorizing that pays attention to mediators in the relationship between differences among group members and outcomes. We build upon the insights from
this literature and Harrison and Klein’s (2007) typology to examine the effects of diversity in the
distributed and volunteer context of OSS projects.

2.3 OSS Projects and Diversity

As noted, the concept of diversity has been largely studied in the context of face-to-face, collocated
teams, or virtual teams. There are at least four ways in which distributed, volunteer groups such as OSS
projects are simultaneously distinct from and similar to such settings, and indeed to other virtual
communities such as those represented by UseNet groups (Butler 2001). First, unlike most organizational
workgroups or teams, individuals in OSS projects are unlikely to have a prior history of interaction and
limited knowledge about other participants (except in rare circumstances when a high-profile contributor
participates in a project). Second, unlike other virtual communities but similar to organizational
workgroups or virtual teams, OSS projects are task-focused in that the collective has the specific goal of
developing software rather than simply information exchange or social interaction (Butler et al. 2002).
Third, whereas members of organizational workgroups or virtual teams are typically assigned to the team
(or recruited into it), participation in virtual communities and OSS projects is volitional. Finally, by
virtue of their method of interaction, OSS projects are similar to virtual teams and virtual communities,
but distinct from collocated workgroups.

Collectively, these similarities and differences point to the need for a distinct conceptualization of
the effects of diversity that pays particular attention to the nuances of the context. For example, prior
research suggests that diversity on attributes such as race and gender has implications for group outcomes
because they are observable (Stangor et al. 1992, Riordan and Shore 1997), but for technology-mediated
distributed groups, such attributes are not immediately obvious. Other characteristics, such as a
participant’s status in the group, may become more salient as an indicator of differences. However,
attributes such as job title that are often related to status in traditional settings, may not be relevant in a
volunteer context, where status may be indicated in different ways, for example by the history of activity
as catalogued in the communication system. Further, the manner through which individuals garner
resources (social and technical) can be distinct in volunteer groups compared to settings where members
are paid and there are formal, hierarchical processes that govern the organization. In the latter setting, an individual’s wage may signal reputation or the possession of resources while wages are not relevant as a differentiator in a volunteer organization. In such settings the role played in the organization could be a better way to infer individuals’ relevant resources. Based on these unique characteristics of distributed and volunteer groups, next we propose hypotheses related to the effects of diversity in the OSS context.

3. Research Hypotheses

Our conceptualization of the impacts of diversity on OSS project user participation is depicted in Figure 1. We use Harrison and Klein’s (2007) typology to theorize the effects of the three distinct types of diversity: variety, separation, and disparity. Variety diversity is expected to have direct effects on user participation, in addition to indirect effects via an impact on project contributions. We describe specific instantiations of the three types of diversity and the theoretical logic underlying the proposed relationships below.

3.1 Variety Diversity

Variety diversity is defined as the variation that exists in the knowledge and experience of the group: “the number and spread of ‘batches’ of information content, experience, or unique network ties available across unit members” (Harrison and Klein 2007, p. 1204). Following this conceptualization, we consider two sources of variety diversity in an OSS project: role-based, which is the relative proportion of developers to active users in the project, and tenure diversity, which refers to variation in platform tenure among participants. In general, following the logic of theories of information processing and the principles of requisite variety (Van Knippenberg and Schippers 2007; Harrison and Klein 2007), which suggest that heterogeneity in informational resources is valuable for problem solving and performance, variety diversity is posited to exert a positive influence on group outcomes. This may be especially true in high technology settings, where combining knowledge sources to create innovative products is the key to success (Joshi and Roh 2009).

It has been argued that knowledge, representing the individual and collective stock possessed by a
group, is a key input for innovation (Leonard and Senipser 2000). Software development has often been characterized as a knowledge-intensive activity (Robillard 1999) that requires two major types of knowledge, how-to knowledge and awareness knowledge (Rogers 1995; Tornatzky and Fleischer 1990). For OSS projects how-to knowledge concerns an understanding of the processes and practices involved in developing software, and includes knowledge about software engineering and programming languages, while awareness knowledge relates to the use of the software in a particular context.

Participants may have more or less access to the different types of knowledge depending on their role in the project. Developers are best positioned to have the how-to knowledge needed to further the project goals. The designation of the developer role is predicated on demonstrated mastery of relevant how-to knowledge, including programming and the use of the tools of the development platform (von Krogh et al. 2003; Ducheaneaut 2005). However, developers may possess a limited stock of awareness knowledge based only on their own local contexts of use. Users are positioned to bring greater awareness knowledge to the project based on their more varied usage settings (i.e., use of the software in the different organizational environments in which the users reside, in the different software environments present in those organizations, and for a potentially wider variety of tasks). Indeed, users possess critical knowledge to guide the development of the software in the direction where it will create the most value in its uses, and the importance of obtaining user input for ensuring success is widely recognized in the literature (Hartwick and Barki 1994).

A broad repertoire of knowledge available in a project, reflected in a balance between users and developers, should enable the OSS project to construct a product that meets a variety of user needs. This will encourage users to return to the project over time to acquire new releases and updates and to recommend the application to their friends and colleagues - potentially expanding the user base. Indeed, prior qualitative studies have demonstrated that user assessments of the project drive initial participation in project discussions (e.g., a freenet user says, “I just joined the freenet-dev list today…I think the concept is absolutely brilliant!” quoted in von Krogh et al 2003: 1227). Thus one way that role based diversity amplifies user participation is by enhancing the usefulness of the software, thereby fulfilling the
use-based motivation of users.

A second mechanism by which role based diversity may have an effect is via users’ development of a social identity associated with the project. Users develop an identity associated with an OSS group via self-categorization based on similarities with other group members (Bagozzi and Dholakia 2006). Thus if a group contains a balance of both active users and developers there will be greater opportunity for current and potential active users to see others within the group who are similar to themselves. Another important aspect of identity construction in OSS rests on how one is received by the other group members and whether one's identity within the group is confirmed by them (Fang and Neufeld 2009). A balance of roles in the project may make the group appear more likely to be receptive to user input and open to incorporating user ideas, thus enhancing current and potential active users’ group identity. Thus based on the expectation of positive effects via identity and use-based motivations, we propose:

**H1: Role based variety diversity is positively associated with user participation in OSS projects.**

A second form of variety diversity is reflected by differences in tenure in an OSS development platform. This form of diversity signals heterogeneity in resources related to experiences using the platform, and social resources related to knowing others using the platform (Moreland and Levine 1989, Morrison 2002). Those with longer tenure on the platform are likely to be aware of a larger set of features that the platform offers, which will give them more tools to achieve software development goals. Developers with longer tenure also may be aware of many other developers who use the platform. This can give them access to the skill sets of these other developers, potentially expanding the range of knowledge resources available to solve problems in the project (Kuk 2006). As argued by Monge et al. (1985), when the “eyes and ears” of the project participants are in different environments they have access to a greater variety of relevant information. A range of external social ties allows group members to draw from larger and potentially non-redundant sets of knowledge (Argote and Ingram 2003).

Further, as in other organizations, OSS projects frequently depend on new participants to be the “bearers of technological innovations and new values” (Reed 1978). New comers may display ingenuity in using the features on the platform, creating prior untapped opportunities. As suggested in prior work,
innovation frequently occurs at the intersection of non-overlapping knowledge sets (Cohen and Levinthal 1990). In OSS projects, participants with longer tenure in the setting have greater platform knowledge and experience, while those who are relatively new to the environment likely infuse new perspectives. Greater tenure diversity may enable the combination of experience with new ideas to enhance the usefulness of the software and thereby attract more user participation. This type of interaction is exemplified in the following exchange between long and short term members of the Azureus project on SourceForge:

**Short term Participant** “Nobody is talking about the one important missing on Azureus!?...This is strange…Why this professional soft is not including on load (torrent) menu, a mathematical calculation of drive space needed for all files or for selected files only, when not entire content will be downloaded....What do you say about this extremely important missing?"

**Long term Participant** “I’d say, I agree with you. I’m so used to do that mentally that I forgot it is a obvious feature to ask. You have my vote. Trying to be specific and offer some ground for discussion, the place where I think that could be, are in the torrents detail/Files, probably in the file pieces, a bottom indicator with one pair of values, the "reserved space" and the "needed space to finish"

In addition to providing means to enhance the usefulness of the software, tenure diversity may also have positive effects on user participation by encouraging identity construction. Our reasoning here is similar to the arguments proposed for role based diversity. Diversity in platform experience in the OSS group is expected to make the group receptive to a wider range of user input. Newer developers may be more willing to respond to the, possibly naïve, questions and ideas of new users since they have likely been in the user role recently themselves (von Krogh et al. 2003) and because providing assistance to others may be a way for new members to enhance their own learning and standing in the group (Lakhani and von Hipple 2003). We therefore test:

*H2*: Tenure-based variety diversity is positively associated with user participation in OSS projects.

In reviewing three decades of literature on diversity, Van Knippenberg and Schippers (2007) urge researchers to consider the microprocesses mediating the effects of various diversity outcomes on workgroup outcomes. As shown in Figure 1, we propose that participant contributions represents an important intervening variable in the relationship between the two forms of variety diversity and user
participation. Contributions are distinct from user participation: while the latter reflects the number of individuals who participate in the project as users, the former is an indicator of the extent or volume of engagement among all project participants (users and developers), and reflects the degree to which the community is active in terms of information sharing. Information sharing is an important outcome in online communities (Marett and Joshi 2009) and mediates the relationship between diversity and performance (Bunderson and Sutcliffe 2002), and in the OSS context, contributions are the means by which project members share information and discuss ideas.

The association between both forms of variety diversity and contributions is predicated on the fact that a robust knowledge stock, diversity in opinions and perspectives, and varied levels of cognitive resources made available by the dispersion present in the group creates a vibrant interaction setting. Heterogeneity in points of view encourages discussion, it infuses new thinking and innovation among participants, and it yields both social and informational benefits for participants; motivations typically ascribed to higher levels of contribution to communities (Wasko and Faraj 2005). For example, Kuk’s (2006) case study of the KDE open source project found that more knowledge sharing occurred when discussions included diverse views and ideas. Following from this logic we propose:

H3: Role-based variety diversity is positively associated with contributions to an OSS project.

H4: Tenure-based variety diversity is positively associated with contributions to an OSS project.

When developers make knowledge contributions, the contributions can serve as a communication mechanism with other developers on the project so that they can better understand the vision that their fellow developers have for the project. Likewise, as users make contributions, they reveal their preferences to developers; these may then be used to improve the application and attract more users. Thus a larger volume of contributions is likely to enhance the usefulness of the software. Even in the absence of actual improvements, activity is a signal that the software will continue to be evolved, updated, and supported. Therefore, when the OSS project community is active and vibrant with many contributions from participants, the project is more appealing to potential users (Golden 2004).

In addition to their influence on user participation as a result of satisfying use-based motivation, a
A large number of contributions may have a positive effect on the extent to which existing users and potential users identify with the project. Because contributions represent the diverse ideas of individuals who span different roles, are embedded in different and possibly non-overlapping networks, and encompass a range of prior experiences, they may provide a large pool of content in which users can find perspectives that match their own. A user may be more likely to participate if s/he perceives that an idea will be well-received by someone else in the group. Thus, we suggest:

\[ H5: \text{The number of contributions is positively associated with user participation in OSS projects.} \]

3.2 Separation Diversity

Separation diversity “indicates differences in position or opinion among unit members” (Harrison and Klein 2007), especially as they relate to team processes and goals. Such differences have been implicated in a range of negative outcomes, including reduction in group cohesiveness, heightened conflict, and poor task performance. A recent meta-analysis (Joshi and Roh 2009) has shown negative impacts associated with some separation variables to be especially prevalent in high technology settings, and in interdependent teams, both of which characterize OSS. A key form of separation diversity that is pervasive in OSS projects by virtue of the global and distributed nature of OSS communities (Agerfalk and Fitzgerald 2008) is national culture. We focus on the number of different languages spoken by the participants of an OSS project as an indicator of the level of heterogeneity in national culture.

To the degree that cultural diversity reflects different perspectives, expectations, norms, and work styles, it can lead to conflict (Gibson and Gibbs 2006). Cross-national teams frequently face significant communication challenges (Gibson and Vermeulen 2003; Shachaf 2008) that can reduce member identification with the team (Fiol 1991). Team identification is an important motivational factor for inducing individuals to engage in interactions and contribute to the team discourse (Bergami and Bagozzi, 2000). In the context of software development teams in particular, He et al. (2007) find that cultural diversity hampers the development of team cognition – an awareness of “who knows what” and a shared

---

2 Although we do not propose a mediation hypothesis explicitly, partial mediation of the relationship between role-based and platform tenure-based variety diversity and user participation by contributions is implied by hypotheses H3-H5.
task understanding reflected in common mental models.

Cultural diversity may create conflict in OSS projects in at least four ways. One source of conflict is varied expectations of how the software should function and with which other applications it should integrate. For example, multiple applications have been created to run on the major OSS operating system, Linux. However, Linux has many different distributions including Turbolinux, popular in Asia, and Debian, which is more popular in North America. While Turbolinux is available in English, Japanese, Traditional Chinese and Simplified Chinese, Debian is largely available only in English. There are slight idiosyncrasies associated with the different distributions, which could impact how complimentary applications are developed. Thus members of a project who use Turbolinux will prefer that the application integrates with it, while users of Debian may focus on a different set of integration issues specific to that distribution. When users have competing desires and developers cannot meet all of the user’s requests some users may cease their participation in the project.

In addition to differences that stem from application use, cultural differences can also dictate how and when participants expect group work to progress. Western cultures are known to be more focused on the individual, while eastern cultures are more focused on the group (Earley and Mosakowski 2000). Distinct ideas about the value of the individual compared to the group will likely affect how participants approach the activity of open source development. For instance, a western participant may treat frequent questions from a novice as inappropriate, while an eastern developer might expect and welcome such interaction. If a user is met with a response pattern that is not expected they may cease to participate.

Third, the larger the number of languages represented in the OSS project, the more likely that team members are geographically diverse. Geographic location can lead to distinct and often conflicting opinions about when is a good time to work together (Espinosa et al. 2006). In other words, cultural diversity is likely to lead to conflicting opinions about both the goals for the application development as well as the appropriate process the project should follow to achieve those goals. Finally, cultural diversity can trigger social categorization processes (Van Knippenberg and Schippers 2007; Gibson and Gibbs 2006) that evoke stereotyping, distrust, suspicion, and other negative attitudes and behaviors, further
reducing the quality of interactions within the community and slowing down the progress of the project.

Because OSS communities are fundamentally dependent on the computer mediated communication provided by the platform, such dependence may limit the group’s ability to resolve conflicts about work styles or goals for the project (Dreu and Weingart 2003; Gibson and Gibbs 2006). A group that is culturally diverse may find it difficult to establish common ground in communication (Gibson and Gibbs 2006). Limited ability to resolve conflict and multiple diverse ideas about the options for programming styles can easily lead to attrition, as diversity based conflict has been shown to lead to turnover (Jackson 1992; McCain et al. 1983). In an OSS environment where participation is voluntary, this may result in a phenomenon labeled ‘forking’ where developers start a different version of the project (Raymond 1998). When this occurs, development on the focal project slows as project resources are depleted. If development slows, users are likely to be frustrated because errors, or bugs, are left in the application and feature requests they desire remain ideas without implementation. Thus we predict:

\[ H6: \text{Language-based separation diversity is negatively associated with user participation in OSS projects.} \]

3.3 Disparity Diversity

The final form of diversity in Harrison and Klein’s (2007) typology, disparity, reflects differences in concentration of valued social assets or resources such as pay and status among unit members. The importance of status earned through reputation in the OSS community is well documented (Fitzgerald 2006). Raymond deemed the OSS community a meritocracy in which reputation is gained by the knowledge that is given to the community (Raymond 2001) and Roberts et al. (2006) find evidence that Apache, a large OSS project, is based on a meritocracy structure. Reputation within an OSS community may be monetizable in that status as a skilled developer could result in an individual being hired as a paid developer, and indeed, reputational gains have been identified as a major motivator for individuals’ contributions to virtual communities (Butler et al., 2002; Wasko and Faraj 2005).

Reputation in the OSS community is commonly established through sharing quality contributions of knowledge or solutions on a consistent basis (Sharma et al. 2002; Ljungberg 2000). Lakhani and von
Hippel (2003) note that the path to becoming a decision maker in OSS projects is through the type and number of contributions made. Contributions of knowledge may take the form of code contributions, bug reports, feature requests, or more general participation in project discussions. A person who is active, i.e., makes many contributions, is likely to have a higher status (Ducheaneaut 2005, Fang and Neufeld 2009, von Krogh et al. 2003).

When a group is diverse in that some group members have a higher status and associated valued assets, these group members may limit the degree to which lower status members feel comfortable contributing, and the overall quality of the relationships among group members will decline (Phillips et al. 2009). In some instances, the highest status members of an OSS project may purposely silence other participants, eventually leading to project stagnation, as occurred in the Linux on 8086 project (Cox 1998, cited in Crowston and Howison 2006). Even without active silencing of peripheral members, a failure of high status members to respond to or encourage other members may cause decreasing participation in a project due to the negative impact on members’ identity construction (Fang and Neufeld 2009).

In short, status differences may reduce the “visibility” of new contributions to the OSS project (Lerner and Tirole 2000) thus impeding the accumulation of knowledge to create useful software and further diminishing the expectation of reputational payoffs and identity formation (Fang and Neufeld 2009). The result will be withdrawal of user participants. Thus we propose:

\[H7: \text{Activity-based disparity diversity is negatively associated with user participation in OSS projects.}\]

4. Methodology

4.1 Data and Sample Construction

Data for this study are drawn from the SourceForge Research Data Archive, an archival database of OSS projects provided by SourceForge to the University of Notre Dame (Van Antwerp and Madey 2008). The archival data set includes information about project developers and users, such as participant identification numbers, when they registered on SourceForge, and how and when they interact with each
project. The archive also includes information about the project including the stage that it is in and the intended users of the application.

SourceForge is one of the most popular OSS development platforms; it allows developers and users to observe others’ project activities, send bug reports, feature requests and patches, post to forums, launch new OSS projects and join existing ones, coordinate and work jointly in specific OSS projects, and integrate the software produced into a larger software application. While there are more than 168,000 projects registered on SourceForge there is considerable heterogeneity in project start time, project goals, and the access projects have to resources that can be leveraged to meet those goals. We constructed a sample of OSS projects so as to minimize the influence of factors outside the focus of this study, viz., OSS development platform, time since the OSS project registered on SourceForge, project activity and development group size. The choice of the SourceForge platform as the single source of projects for sample construction controls for any effects due to variation in development platform.

Development activity could be affected by events such as the release of an updated programming language package or a previously proprietary product as occurred when Netscape opened up some of its source code. Projects that register on SourceForge around the time of such an event or after it could have inflated activity that is not related to the factors observed in this study. To help ensure that this type of event does not confound our results we selected a sample of projects registered on SourceForge between June, 2004 and June, 2005. There were 31,591 projects that registered during that year.

Because we are interested in group level constructs, and many projects on SourceForge are individual efforts with no associated community, we exclude projects that have less than 5 developers. Of the projects that registered between June, 2004 and June, 2005, 1,351 of them had at least 5 developers. Finally, because many projects remain listed on SourceForge even after they have effectively been abandoned, we only include projects that have some activity by developers during our observation period. There were 337 projects out of the 1,351 that had developer activity during the observation period. These 337 projects are the focus of our analysis and they collectively represent a broad range of application types. Included in the sample are projects that develop applications to aid in software development,
management, statistical analysis, education, online sales and entertainment.

We observe each project for two years. Given that prior work has shown typical SourceForge projects to be active for approximately one year (Stewart et al. 2006), this period should capture a significant portion of the life of most projects. Because OSS projects in the sample started on different days, the observation period does not necessarily correspond to the same calendar days for each project.

### 4.2 Operationalization of Variables

**Dependent Variable.** User participation is operationalized as the number of users participating during the second year of the project’s life. We distinguished members of the development group from users using SourceForge userids. Userids listed as developers or administrators during December, 2005, i.e., between six to eighteen months after project initiation, are identified as developers. The users for each project are isolated using userids that reported bugs, or requested features or support during the second year, but were not identified as developers. These users may or may not have participated during the first year. If the same userid makes multiple contributions during the second year it is only counted once.

**Independent Variables.** To allow for a time lag between the presence of diversity and its effects, we use data from the first year of the project for the measurement of independent variables. The degree to which participants are dispersed across roles (developer versus user), i.e., role based variety diversity, is operationalized using Blau’s index (1977). A second form of variety diversity, participant platform tenure diversity, is measured by first calculating each participant’s tenure as the amount of time since the day he or she registered on SourceForge, and then computing the variance across participants. This is similar to Peretti and Negro (2006)’s measure of “newness” to industry compared to “newness” to

---

3 Blau’s index is calculated as \(1 - \sum p_k^2\) where \(p\) is the proportion of unit members in the Kth category. We have 2 categories, developers and active users. Values can range from zero to \((K-1)/K\). Higher numbers indicate that members are more equally spread across the groups.
the team. Separation diversity is measured by participant cultural diversity. We use the sum of spoken languages by all participants in a project as a proxy for cultural diversity. Disparity diversity is operationalized using an activity based disparity measure. For each participant, we compute activity based on the number of bug reports or features requests they submit to any project on the Sourceforge platform. We then use the coefficient of variation associated with the activity of all group members to create a group level measure of activity based disparity diversity. Contributions are measured based on the number of bug reports, feature requests and other knowledge contributions that are submitted to the project by developers and users. We use the number of contributions in year two as the mediator and the number of contributions in year one as a control variable.

**Control Variables.** Together with contributions in year 1, we controlled for four additional potential sources of variation in user participation. First, we include the total number of participants as a control, as a larger number of participants may be associated with more resources that can be leveraged to develop software and attract users (Butler 2001). Second, we include an indicator for the stage of the project, i.e. we control for whether it is in production or not, because Stewart and Gosain (2006a) have shown that stage impacts outcomes for OSS projects. The number of intended audiences could also impact user interest because if the software is intended for use by more audiences then it might be expected that more users would be attracted to the project. Thus, the number of intended audiences is our third control. Finally, we control for whether or not the project uses the Gnu GPL, as prior studies have shown that license choice affects user interest (Stewart et al. 2006).

5. Analysis and Results

5.1 Analysis

Table 1 displays descriptive statistics and summarizes variable intercorrelations for all research variables. Comparison with descriptive statistics reported in prior studies indicated that our sample appears similar to past work in terms of the number of developers and distribution of licenses (Stewart et al. 2006, Lerner and Tirole 2002, Stewart and Gosain 2006a).
We used mediated regression analysis to test the research hypotheses. To ensure that the data satisfy the underlying assumptions of the statistical technique, we performed a series of analyses. The dependent variable, user participation was left skewed, and so we performed a natural logarithmic transformation. The model variance inflation factors (all below 4) and the correlations between the independent variables (see Table 1) suggest that multicollinearity is unlikely to be a concern in model estimation. The Durbin Watson value (<3) showed no evidence of serial correlation. Observation of the residuals and Cook and White’s test suggest that the errors are normally distributed.

5.2 Hypothesis Tests

Mediated regression results are presented in Table 2. All models are statistically significant at p<.01. Model 1 includes only control variables and explains 23% of the variance in user participation. Model 2 includes main effects and controls and explains 40% of the variance in user participation. In Model 3 we add contributions in year two. Model 3 explains 44% of the variance in user participation, suggesting that contributions in year two is associated with user participation. Model 4 includes the controls and the variety diversity variables, with contributions in year two as the outcome. It explains 66% of the variance in the number of contributions in year 2. Model 5 includes the controls, the diversity independent variables and the contributions mediator. It explains 55% of the variance in user participation. The increase in $R^2$ from Model 3 to Model 5 is significant (p<.01). We use the results of model 5 to test the research hypotheses related to the effects of diversity on user participation, and the effects of contributions on user participation, i.e., H1, H2, H5-H7. The coefficients of role-based variety diversity and platform tenure based variety diversity are significant, supporting H1 and H2.

Contributions have a significant impact on user participation, supporting H5. Contrary to predictions, separation diversity as measured by number of spoken languages in a project is not significant, thus, H6 is not supported. The significant coefficient for disparity diversity in Model 5 provides support for H7.

Hypotheses 3 and 4 predicted that, in addition to their effects on user participation, the different forms of variety diversity would influence contributions. We use the results of Model 4 to test H3 and H4. The coefficients for both forms of variety diversity are significant, supporting H3 and H4. Finally,
we note that the implied expectation of partial mediation by contributions is supported using the results from Models 3, 4 and 5 (Baron and Kenny 1986). The Sobel test provides further evidence that the impact of variety diversity on user participation is mediated by contributions (Sobel 1982). Results of the hypothesis tests are summarized in Table 3.

5.3. Discussion

Despite the emergence of the OSS development paradigm as a viable alternative to traditional software development models and the significance of human capital in this activity, there is limited theoretical understanding of how the characteristics of OSS project participants influence project outcomes. Acknowledging that diversity is a defining characteristic of open source projects, in this study we leveraged the robust diversity literature to conceptualize drivers of user participation. We drew on the typology developed by Harrison and Klein (2007) and instantiated it in the context of a distributed, volunteer collective. Results from the empirical analysis of data on 337 projects from SourceForge generally supported the hypotheses, reinforcing the importance of diversity in driving outcomes for non-traditional work settings. We find that various types of diversity have both positive and negative effects on the ability of open source development projects to attract user participation.

The one theoretical prediction that was not supported by the analysis is our expectation of a negative relationship between separation diversity and user participation. We argued that a variety of spoken languages, reflecting cultural diversity, is likely to be a source of conflict in the OSS project and impede its effective functioning. There are two plausible explanations for the non-significant finding. One is that language is not an adequate reflection of cultural diversity and arguably, in a collective whose raison d’être is software construction, commonality in the language used for software development is more important than the natural language of the participants. A second explanation may be that many of the developers are multi-lingual and so the barriers across some languages are not significant.

5.4 Limitations

Prior to discussing the contributions and implications of the findings, we acknowledge important limitations of the study. The first limitation relates to generalizability. Although SourceForge is among
the largest open source development platforms, flagship projects such as Linux and Apache are not hosted on it. The sample of OSS projects observed in this study is taken from SourceForge, and therefore may not be representative of all OSS projects, such as those that use other platforms or those that do not use a platform that hosts multiple projects. OSS projects hosted on other platforms such as GForge (http://gforge.org) may possess different characteristics from those on SourceForge. If a project does not use a platform that hosts multiple projects, the developers on that project may have limited opportunity to attract developers from other projects, thereby limiting the level of “variety” among participants. Further, it may be more difficult to attract users if the project does not share a platform that draws a large amount of users. Thus, results from this study generalize only to projects that have a profile and infrastructure similar to those on SourceForge.

Second, the variables in this model are measured using archival data. While this is a strength of the study in that the data are objective and free of respondent bias, with archival data it is challenging to map the data perfectly onto the theoretical constructs. One instance of this limitation is that we identify individuals by their userid. It is possible that multiple people use the same userid or that one person uses multiple ids, and therefore a userid does not correspond to a unique individual. However, we note that this method of identifying participants is commonly used in other studies (e.g. Hahn et al. 2008). Future research can use surveys to more tightly tie the archival data to constructs. Finally, we note that although we use lagged variables to support a causal interpretation of the findings, the mediator and outcome are measured contemporaneously. Thus, causality is only implied theoretically and cannot be confirmed by the empirical findings.

6. Contributions, Implications, and Future Research

6.1 Contributions to the Open Source Literature

Although other work has sought to understand the factors that lead to user-related outcomes in OSS projects, this study extends extant literature in three ways: by considering the nuanced effects of diversity in OSS groups, by focusing on user participation as an outcome, and by highlighting the
importance of differentiating roles in OSS projects. While OSS leaders have recognized the important impacts of diversity in their projects (Perens 2009; Ducheanaut 2005), this is the first study we are aware of that directly addresses the influence of diversity on outcomes in OSS projects. We highlight specific kinds of diversity that are relevant in the unique context of OSS and demonstrate that they have both positive and negative effects on outcomes. The current work supports the notion that different kinds of participants in OSS projects have dissimilar kinds of knowledge and resources, as reflected by their roles and tenure, and that the degree to which the participants bring a variety of knowledge to the group has a positive impact on project success. However, when members have more varied status in the group, as reflected by their activity, this results in a negative impact on success. Thus diversity is not simply all good or all bad in OSS projects; what is important is the nature of differences and their implications for group knowledge sharing and project outcomes.

The second major contribution to the OSS literature relates to our focus on user participation as an indicator of project success. As has been extensively noted in traditional development contexts, users are valuable contributors to software construction activities and are vital to ensuring the final success of the product (e.g. Harwick and Barki 1994). Likewise, the importance of user participation in OSS has been recognized by both practitioners and researchers (Agerfalk and Fitzgerald 2008). However prior studies have mainly focused on attracting developer participation to projects rather than understanding what is important to attract user participation (e.g., Stewart et al. 2006). Those studies that have focused on user-related outcomes have tended to highlight measures of market success rather than actual engagement of users in the project (e.g., Grewal et al. 2006).

Our indicator is a stronger measure of the amount of effort that the user community is investing in the development of the application. User effort is critical to efficient and effective software development, and is a key resource that companies often attempt to access when leveraging OSS communities (Agerfalk and Fitzgerald 2008). Collectively, prior work and the current study explain the antecedents of many success factors associated with users that are of interest to proprietary software development companies.
A final contribution to the OSS literature is the evidence we provide regarding the importance of distinguishing among the roles participants occupy in an OSS project to understanding how participant differences impact outcomes. While many developers are also users, there is evidence that there are participants who are only users. Ye and Kishida (2003) distinguish participants according to their role in the project and also acknowledge that projects are different from each other in terms of the number of participants in each role. However, this line of research stops short of indicating how different balances of participants of various types can influence the project. This study is among the first to theorize and provide empirical evidence for the differential impact of users versus developers. Results highlight that this distinction is useful in understanding the antecedents of success; thus, it is imperative that users be treated as distinct from developers to gain a finer-grained understanding of the dynamics of OSS development and outcomes in OSS projects.

6.2 Contributions to the Diversity Literature

Diversity research has limited its attention to the effects of differences among unit members in traditional organizations, characterized by paid employees, formal hierarchies, departments, and shared geographical space, with only a select few studies examining virtual teams (e.g., Peters and Karren 2009; Staples and Zhao 2006). We applied the theoretical concepts related to diversity developed in this body of work to a setting exemplified by a volunteer labor force, informal hierarchies, geographical dispersion, swift changes in user needs and high levels of user participation. Specifically, we leveraged the typology developed by Harrison and Kline (2007) and responded to the call to identify mediators that are critical in unraveling the mechanisms underlying the impact of diversity.

Consistent with prior research, we find that the impacts of diversity can be positive or negative. We find support for the underlying theoretical arguments developed by Harrison and Kline (2007) in that variety diversity positively influences success and disparity diversity negatively impacts success. The literature on diversity has long sought to understand what attributes represent differences that are consequential in a group setting. Our work contributes to the search for relevant attributes that create diversity by conceptualizing and testing the relationship between measures that are distinct from those
commonly used in diversity research and chosen specifically based on the characteristics of OSS projects. We follow Harrison and Kline (2007)’s suggestion to consider the specific context chosen in selecting measures that fit well with the underlying interaction dynamics of the context. Thus, rather than relying on typical socio-demographic attributes such as race and gender that may not be salient in a virtual context, and are also not expected to directly impact contribution to the application’s development, we focus on characteristics that are related to interacting with the other OSS project participants about the application development. Specifically, we argue that activity is a useful proxy for status in OSS projects based on the notion that OSS projects operate as a “meritocracy.” This measure of status is in contrast to typical measures of status such as income or formal organizational title. Such measures may be useful to consider for future researchers that seek to understand the implications of diversity in virtual, globally distributed and volunteer contexts.

We also find evidence that the impact of variety diversity on user participation is mediated by contributions, a form of information sharing among participants. The diversity literature has not explicitly isolated and directly measured mediators, rather, the mediation process has been proposed theoretically and presumed to exist based on the significant relationship between diversity and outcomes (Van Knippenberg and Schippers 2007.) Finding empirical support for mediation may encourage diversity researchers to employ longitudinal research designs such as ours that allow for the measurement of the intervening variables and, as a result, development of a more robust understanding of the effects of diversity.

6.3 Implications for Practice

As the distinction between OSS development and proprietary software development begins to blur, managers need a deeper understanding of the dynamics underlying the OSS development process. Our findings shed light on the importance of managing the balance among different types of participants, thereby yielding implications regarding the types of individuals that managers should target for participation in an OSS project. For instance, while many companies have begun paying employees to work on OSS projects, our findings imply that managers should incentivize participation from both
developers and users, as the latter play a key role in creating the vibrant setting that leads to success. Managers should also seek to achieve a balance between participants who have longer experience with OSS tools as well as new comers to the platform. This might imply actively recruiting users to participate on a development platform. Heterogeneity in participant roles and platform tenure will have beneficial effects on expanding the market for the software.

A broader implication for managers in traditional work settings relates to a redefinition of the notion of “status.” The integration of the diversity literature into the OSS setting required the development of granular insights such as understanding the unique process through which reputation is developed in a context where volunteers are prevalent. Understanding that contribution level can be associated with status and that disparity of that status across participants can have negative impacts is important as organizational structures become more flexible, and the Internet becomes even more significant as an essential foundation for work practices. As more groups collaborate through this foundation, status based on socio-demographic factors or by virtue of occupying the corner office could lose significance. Our results point to the need for careful consideration of how status is visibly represented through system design features. For example, it may not always be desirable to make the level of participants’ past contributions transparent.

Finally, though their contributions are valuable, managers should be wary of highly active participants. While it is well known that it is difficult to attract voluntary effort, this research suggests that participants who monopolize the work can have negative implications for expanding user participation. Managers should make sure that highly active participants do not limit the contributions of those with less activity. Those who contribute less may also have something valuable to offer and highly active participants could squelch their ability to share their skill with the group. One way to manage this could be to create transparency so that highly active participants do not possess private information that others cannot access. If the logic behind contributions is not transparent the more active participants could have more information about the status of the code and make it difficult for less active participants to offer suggestions about or make contributions to the application. When the code and the logic behind
its evolution are available for inspection by all, it would enable those who contribute less frequently to provide timely and useful comments. A second way to disallow highly active participants from crowding out the effort of less active participants is to have procedures that ensure that the ideas of less active participants are valued and addressed. For instance, before a feature is added, there could be a rule that says the project will accept ideas about its implementation for one week and all ideas received during that time will be addressed before a decision is made as to how to implement the feature.

6.4 Future Research

OSS projects are at the vanguard of leveraging the Internet for innovation, however there are many other examples of Internet use in innovative and collaboration processes. For example, the Internet is used extensively in the creation and evolution of Wikipedia. Future research could begin to explore whether the findings in this study extend to contexts such as Wikipedia and academia (Kane and Fichman 2009). The process driving participation on a single Wikipedia project or page may be distinct from the process observed here because contributing to a Wikipedia project is likely to be more challenging for people speaking different natural languages than it is for people speaking different languages to contribute to an OSS project, as shown by our findings.

In addition to exploring the impacts of diversity in other online innovation communities, future research can attempt to identify factors that will ameliorate the negative effects of diversity. Often researchers suggest that the impacts of diversity can be mitigated or amplified by alternative recruiting techniques, but this may be challenging in a volunteer context. Future consideration should be given to how implementing procedures could alter the negative influence of observable disparity. If, as we argued, the more active developers have special knowledge based on their activities that they use as a reason not to accept less active member’s contributions or to intimidate them, transparency of information could minimize this problem. Projects could implement procedures that require developers to explain what kinds of contributions they make and how they impact the evolution of the project. This would enable less active developers to contribute.

Joshi and Roh (2009) suggest that the context in which the team is embedded alters the effects of
diversity on team outcomes. While we considered unique aspect of the OSS context generally, individual projects within that context operate within their own specific sub-context. For example, if a project is associated with a for profit software development company, then there may be an expectation that some participants will be more active than others and so a disparity measure based on activity may be associated with a less negative impact. Similarly, the platform used (i.e. Sourceforge or a company sponsored platform) could affect the relationship between disparity diversity and success.

In addition to exploring factors that may alter the impact of diversity on user participation, it will be fruitful to explore how diversity influences other outcomes. Given that diversity affects many outcomes including turnover, creativity and innovation there is room to explore the impact of diversity on these outcomes in the context of OSS projects. Identifying the antecedents of developer turnover is likely especially important in OSS projects since developers are often volunteers. Fang and Neufeld (2009) leverage the legitimate peripheral participation perspective to understand the factors that limit developer turnover in OSS projects, but a diversity lens could be used to complement this perspective. While they identify behaviors of participants that are associated with core developers remaining active, a diversity perspective suggests that characteristics of the developers who contribute less frequently and active users in the group could also help to explain sustained contributions.

The current study begins to bridge the gap between the well developed literature on diversity and one of the most technically advanced organizational structures. We considered the unique aspects of the OSS development project and the diversity literature to suggest OSS community characteristics that lead to user participation. In doing so we identified antecedents of user participation. While the organizational forms that will emerge as a result of the Internet have many things in common with traditional forms, we hope this research provides a foundation for other work that explores the subtle differences between how constructs influence outcomes in traditional and online environments.
Table 1: Variable Descriptive Statistics and Intercorrelations (Notes: N=337, *p<.1, **p<.05, ***p<.01)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. User Participation (Year 2)</td>
<td>13.44</td>
<td>35.37</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Variety Diversity (User vs. Developer)</td>
<td>0.21</td>
<td>0.20</td>
<td>16*</td>
<td>.14</td>
<td>.24*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Variety Diversity (Platform Tenure)</td>
<td>2.01</td>
<td>1.42</td>
<td>.14*</td>
<td>.24*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Separation Diversity (Language)</td>
<td>3.26</td>
<td>1.94</td>
<td>.34**</td>
<td>.33**</td>
<td>.13*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Disparity Diversity (Activity)</td>
<td>0.99</td>
<td>0.79</td>
<td>.12*</td>
<td>.25**</td>
<td>.030</td>
<td>.23**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Contributions (Year 2)</td>
<td>48.07</td>
<td>84.29</td>
<td>.35**</td>
<td>.24**</td>
<td>.15**</td>
<td>.33**</td>
<td>.46**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Contributions (Year 1)</td>
<td>55.14</td>
<td>109.40</td>
<td>.21**</td>
<td>.20**</td>
<td>.14**</td>
<td>.33**</td>
<td>.52**</td>
<td>.66**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Number of Participants</td>
<td>15.11</td>
<td>21.72</td>
<td>.42**</td>
<td>.10</td>
<td>.07</td>
<td>.62**</td>
<td>.23**</td>
<td>.47**</td>
<td>.55**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Production Stage</td>
<td>0.45</td>
<td>0.50</td>
<td>.15**</td>
<td>.16**</td>
<td>.04</td>
<td>.10</td>
<td>.12*</td>
<td>.16**</td>
<td>.13*</td>
<td>.16**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Gnu Public License</td>
<td>0.57</td>
<td>0.50</td>
<td>.11*</td>
<td>.054</td>
<td>-.06</td>
<td>.20**</td>
<td>-.03</td>
<td>.06</td>
<td>-.01</td>
<td>.12*</td>
<td>.11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11. Number of Intended Audiences</td>
<td>1.72</td>
<td>1.37</td>
<td>.08</td>
<td>-.04</td>
<td>-.01</td>
<td>.01</td>
<td>.06</td>
<td>.07</td>
<td>.05</td>
<td>.08</td>
<td>.18**</td>
<td>.06</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Mediated Regression Results (Notes: N=337, *p<.1, **p<.05, ***p<.01)

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>User Participation (Year 2)</th>
<th>Contributions (Year 2)</th>
<th>User Participation (Year 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
</tr>
<tr>
<td>Variety Diversity (User vs. Developer)</td>
<td>.32***</td>
<td>.08**</td>
<td>.27***</td>
</tr>
<tr>
<td>Variety Diversity (Platform Tenure)</td>
<td>.21***</td>
<td>.09***</td>
<td>.15***</td>
</tr>
<tr>
<td>Separation Diversity (Language)</td>
<td>.04</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>Disparity Diversity (Activity)</td>
<td>-.06</td>
<td></td>
<td>-.12**</td>
</tr>
<tr>
<td>Contributions (Year 2)</td>
<td>.24***</td>
<td>.11</td>
<td>-.36***</td>
</tr>
<tr>
<td>Contributions (Year 1)</td>
<td>.22***</td>
<td>.2***</td>
<td>.19***</td>
</tr>
<tr>
<td>Number of Participants</td>
<td>.03</td>
<td>.02</td>
<td>.06</td>
</tr>
<tr>
<td>Gnu Public License</td>
<td>.05</td>
<td>.07*</td>
<td>.05</td>
</tr>
<tr>
<td>R²</td>
<td>.23</td>
<td>.4</td>
<td>.44</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>.22</td>
<td>.38</td>
<td>.43</td>
</tr>
<tr>
<td>F</td>
<td>20.24***</td>
<td>23.91***</td>
<td>37.59***</td>
</tr>
</tbody>
</table>
### Table 3: Results of Hypotheses Tests

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1(+): Role based variety diversity is positively associated with user participation in OSS projects.</td>
<td>Supported</td>
</tr>
<tr>
<td>H2(+): Tenure-based variety diversity is positively associated with user participation in OSS projects.</td>
<td>Supported</td>
</tr>
<tr>
<td>H3(+): Role-based variety diversity is positively associated with contributions to an OSS project.</td>
<td>Supported</td>
</tr>
<tr>
<td>H4(+): Tenure-based variety diversity is positively associated with contributions to an OSS project.</td>
<td>Supported</td>
</tr>
<tr>
<td>H5(+): The number of contributions is positively associated with user participation in OSS projects.</td>
<td>Supported</td>
</tr>
<tr>
<td>H6(-): Language-based separation diversity is negatively associated with user participation in OSS projects.</td>
<td>Not Supported</td>
</tr>
<tr>
<td>H7(-): Activity-based disparity diversity is negatively associated with user participation in OSS projects.</td>
<td>Supported</td>
</tr>
</tbody>
</table>

### Figure 1 Research Model

![Research Model Diagram](image)
References
Espinosa, J.A.; Lee, G.; and DeLone, 2006. W. Global boundaries, task processes and IS

Levina, N. and Vaast, E. 2008. Innovating or doing as told? Status differences and overlapping boundaries in offshoring collaboration, MIS Quart. 32(2) 307-332.


Majchrzak, A. and Malhotra, A. Virtual workspace technology use and knowledge sharing effectiveness in distributed teams: the influence of a team's transactive memory, Knowledge Management Knowledge Base, Online publication.


Marett, K. and Joshi, K.D. 2009. The Decision to Share Information and Rumors: Examining the Role of Motivation in an Online Discussion Forum. Communications of the Association for Information Systems 24(4) 47-68.


Pfeffer, Jeffrey and Nancy Langton. 1993. The Effect of Wage Dispersion on
Von Krogh, G., Spaeth S., Lakhani K. 2003. Community, joining and specialization in