OPEN SOURCE SOFTWARE PARTICIPATION: DEVELOPER MOTIVATIONS AND PERFORMANCE

Dissertation

Jeffrey Alan Roberts
Tepper School of Business
Carnegie Mellon University
Tech & Frew Streets
Pittsburgh, PA 15213
Phone: (412) 268-3681
Fax: (412) 268-7064
Email: jroberts@andrew.cmu.edu

April 24, 2006

Committee:

Prof. Christopher Forman

Prof. Il-Horn Hann (co-chair)

Prof. James D. Herbsleb

Prof. Sandra A. Slaughter (co-chair)

UMI Number: 3210014

Carnegie Mellon
Tepper
SCHOOL OF BUSINESS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
INDUSTRIAL ADMINISTRATION
(INFORMATION SYSTEMS)

Titled
"OPEN SOURCE SOFTWARE PARTICIPATION: DEVELOPER MOTIVATIONS AND PERFORMANCE"

Presented by
Jeffrey Alan Roberts

Accepted by

Sandra A. Slaughter
Co-Chair: Prof. Sandra A. Slaughter

April 24, 2006
Date

Il-Horn Hann
Co-Chair: Prof. Il-Horn Hann

April 25, 2006
Date

Approved by The Dean

Kenneth B. Dunn
Dean Kenneth B. Dunn

April 2, 2006
Date

Abstract

This dissertation examines questions related to the provisioning of software development expertise to globally distributed, freely available, public software development projects – i.e., open source software (OSS). Relying almost exclusively volunteer labor, OSS projects have played a fundamental role in the development of the Internet by contributing to such remarkable software such as Linux and the Apache-WEB server. This dissertation conducts a rigorous analysis into issues related to OSS developer motivations and performance. This work yields detailed data and models exploring the sources and consequences of OSS developer motivations. Understanding these motivations is critical for both software producers and consumers. Producers hope to exploit the competitive advantages of "costless" OSS development. Consumers seek to understand the risks associated with adoption of OSS products. A deeper understanding OSS development has implications for IT strategies that organizations should pursue over time. Understanding the incentive structure underlying OSS participation is a significant part of this process.

Part one of this work provides insight into the motivational profiles of OSS contributors. Adopting a functionalist view of motivation, I identify and empirically examine eight functional dimensions of motivation. This work contributes to the growing literature on OSS development by providing insight into the underlying motivational profiles of participation and by identifying the relative importance of motivations within those profiles. Why developers contribute freely without direct remuneration has been widely debated; part two of this work develops and tests the career incentives hypothesis of OSS participation. Analysis of panel-data covering a six year period reveals that participation may indeed, in part, be explained by existing theories in labor economics. Lastly, part three of this work explores the relationship between OSS developer motivational profiles and their project related outcomes. Understanding how OSS projects attract contributions in sufficient number is a central theme in OSS research. This study contributes to this research stream by revealing how develop motivations are interrelated, how these motivations influence subsequent participation and performance, and how past performance influences their subsequent motivations.

Acknowledgements

This dissertation would not have been possible without the hundreds of Apache Software Foundation developers who generously supported and participated in our data collection efforts. I would especially like to thank Dr. Roy Fielding, a founding member of the Apache Software Foundation, for his support. Roy's introduction and sponsorship of our research in the Apache community was instrumental to the success of this project.

I would also like to express my deepest gratitude to the faculty and staff at Carnegie Mellon's Tepper School of Business. The collective experience and knowledge housed within this institution are truly inspiring. I would especially like to thank my co-chairs Dr. Il-Horn Hann and Dr. Sandra Slaughter. Il-Horn is fearless in the pursuit of research ideas – no data is too difficult to obtain. The worst case scenario (no data) is, in the grand scheme, not the end of the world. Il-Horn also taught me that in empirical research sometimes the simplest "story" is the most interesting. Sandy exemplifies what is best about the academic community. The consummate academic, Sandy is a dedicated researcher who demands no more (or less) of her students than she demands of herself.

Lastly, and most importantly, I would like to express my appreciation and love to my family for their unwavering support during my doctoral studies. To my son Ty and my daughter Riley Anne, mere babies when this started, we are finally not *all* students – just you guys. I know at times this has been hard on you both – thank you and I love you both very much. To my beautiful wife, Sue Ann, I'm not sure where to begin. I doubt anyone could have had as much support as you have given me during this quest. More than your fair share of our family responsibilities fell to you and you always responded without complaint. Thank you for all of your love, support, and most of all friendship, without which, none of this would have been possible.

Table of Contents

List of Tables

List of Figures

CHAPTER 1.
INTRODUCTION

This dissertation examines questions related to the provisioning of software development expertise to open source software (OSS) projects. Open source development, i.e., public software development projects where participants can read, modify, and redistribute the software source code (von Krogh and von Hippel 2003), is arguably one of the most exciting phenomena in the software industry today. Relying largely on volunteer labor, open source projects such as the Apache WEB server and Linux have played, and continue to play, a pivotal role in daily operation of the Internet (e.g., Garvey 1998; Mears 2003; Hochmuth 2004). As important as open source technologies have been in the past, the recent trend toward deployment of open source technologies as critical components of corporate IT infrastructure may signal the beginning of an important new era in open source phenomenon (Hanrahan 2005). A recent market research study predicts that open source technologies will increase penetration rates across the entire core software stack by 2010. According to the authors, open source applications will enjoy significant penetration in both horizontal and vertical business applications such as database management systems, integration layer technologies, desktop productivity and business applications (Koening et al. 2005). Corporations are also becoming increasingly involved in the production side of open source. For example, according to a recent Business Week article companies such as Hewlett-Packard, IBM, Intel, Novell, Oracle, and Red Hat among others are all involved in the production of open source software (Cohen 2005). In 2002, IBM reportedly spent $1 billion, or 20% of its R&D budget, on projects related to Linux (Ante 2001). Interest in the production of open source is not limited to low-level system software, utilities, and tools however. For example, EBay recently began providing open source for some of its search and access applications in a bid to expand its external developer community and encourage developers to find new ways of using the online marketplace (Bostrom 2005).

Increased interest in open source technologies raises important questions about how such projects actually work. While there are many relevant issues associated with operating an open source project, one of the most significant (Koch 2005) and most widely debated questions (Strasser 2001) is the voluntary nature of contributions to open source projects.

Often quoted individual level motivations for participating in open source projects cover a broad spectrum including scratching a "personal itch" with respect to software functionality, enjoyment, and desire to be "part of a team" (Raymond 2001a). Others authors have used economic theories to explain participation in open source projects. For example, von Hippel and von Krogh (2003) discuss a setting that combines private incentives in a public good (collective action) context and term this the "private-collective" model of innovation. Lerner and Tirole (2002) argue that open source participation may be explained, in part, by theories from labor economics. Given the increasing strategic interest in both open source products and development, it is vital to seek a more formal approach of theory formulation and empirical validation to explain differences in open source participation. This dissertation conducts a rigorous analysis of several issues related to open source developer motivations and performance. This work is designed to make a significant contribution to the emerging literature surrounding the issues of motivation to participate in open source projects and the consequences for the project as a result of varying sources of motivation.

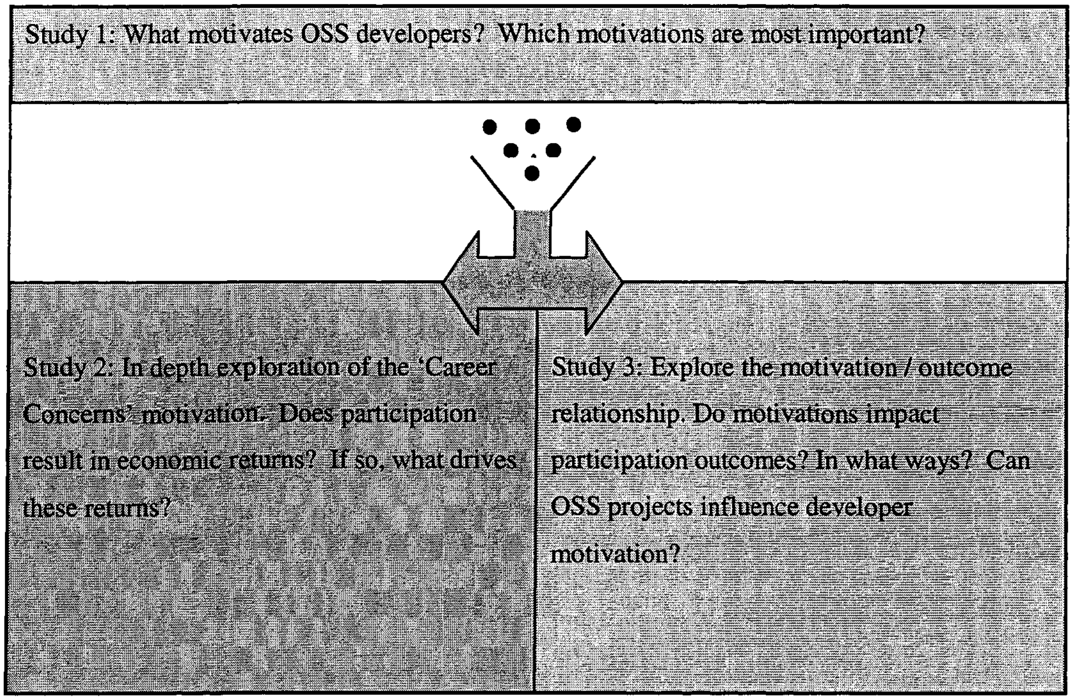
Although there has been considerable discussion concerning which explanations are best suited to describe open source participation, at the time this research was undertaken there were no rigorous empirical studies of motivations to participate in open source projects (See Hertel et al. 2003 for a recent exception). Study 1 examines two important unresolved questions regarding open source participation: (1) *what motivates developers to participate in open source projects* and (2) *which motivations are most salient in explaining participation*. A key finding of this study is the support for several distinct dimensions (factors), each varying in relative importance, underlying the motivation to participate in open source projects. These motivational profiles are found to vary across exhibited levels of commitment (measured as contributions) to the project. The dominant motivation among those exhibiting a low level of project commitment is the personal use value of the software (27%) while the dominant motivation among those exhibiting a higher level of commitment is the recreational aspects of participation (20%).

Study 1 revealed that developer motivations for participation in open source projects are multi-dimensional. Dimensions such as the use value of the software, recreational aspects of the software development task, and career related concerns were all shown to be salient in explicating open source participation. Unlike the other dimensions of developer motivation which operate at a private or personal level only, the motivational sway of career concerns requires a favorable reception by the IT labor markets. That is, the skills acquired in the service of one's open source project must actually have some market value. The value of this experience in the labor markets is an open empirical question and is the subject of Study 2. Study 2 answers two questions related to the career concerns motivation: (1) *does open source participation result in economic returns* and (2) if so, *what drives these returns*. A key finding of this study is that greater open source experience (measured as contributions) does not result in wage increases for contributors. This suggests that employers do not reward the gain in experience through open source participation as an increase in human capital. On the other hand, achieving a higher status in the merit-based ranking within the Apache open source community is associated with more than a 20% increase in wages, depending on the rank attained. Our results are consistent with the notion that a high rank within the Apache Software Foundation is a credible signal of the productive capacity of a programmer.

The third and final study comprising this dissertation explores the relationship *between* motivational orientation and various impacts on the open source project. As was demonstrated in Study 1, individual contributors have multiple salient reasons for participating in open source projects. An interesting and unanswered set of questions following from Study 1 relates motivational orientation to project outcomes. Thus, Study 3 addresses four questions related to motivational orientation and outcomes. (1) *How are the motivations of contributors related? I.e., are they independent, complementary or contradictory?* (2) *How do differences in contributors' motivations relate to* *differences in their participation?* (3) *How do levels of participation relate to changes in contributors' status or performance rankings within the project?* (4) *How does a change in performance ranking affect the subsequent motivations of open source participants?* Study 3 contains several important findings. First, developers' motivations are not independent but rather are related in complex and unexpected ways. For example, being paid to contribute to Apache projects is positively related to developers' status motivations but negatively related to their use value motivations. Second, different motivations do, in fact, differentially impact participation. While developers' extrinsic and status motivations lead to above average contribution levels, use value motivations lead to below average contribution levels, and intrinsic motivations have no significant impact on average contribution levels. Thirdly, developers' contribution levels positively impact their performance rankings. And lastly, the results suggest that past performance rankings enhance developers' subsequent status motivations.

In summary, this dissertation conducts a rigorous analysis of several issues related to open source developer motivations and performance. Figure 1.1 outlines and relates the three studies comprising this dissertation. Taken together, these studies make a significant contribution to the emerging literature surrounding the issues of motivation to participate in open source projects and the consequences for the project as a result of varying sources of motivation.

Figure 1.1 – Relationships Among Studies Comprising the Dissertation



THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2.
STUDY 1 - WHY DEVELOPERS PARTICIPATE IN OPEN SOURCE SOFTWARE PROJECTS: AN EMPIRICAL INVESTIGATION

Introduction

Open source software (OSS) development, i.e., public software development projects where participants can read, modify, and redistribute the software source code (O'Reilly 2000; OSI 2001), is arguably one of the most exciting phenomena in the software industry today. One widely debated question is why open source programmers voluntarily contribute significant amounts of time and effort to their projects, thereby foregoing any direct remuneration that they might accrue by spending that same time working on a commercial software project. Often quoted individual level motivations for participating in open source development projects cover a broad spectrum including scratching a "personal itch" with respect to software functionality, enjoyment, and a desire to be "part of a team" (Ghosh 1998; O'Reilly 2000; Raymond 2001a). Others liken the open source community to a gift culture where the status of a participant depends on "what he gives away" (Raymond 2001b). Alternatively, Lerner and Tirole suggest that open source participation may in part be explained by existing theories of labor economics (Lerner and Tirole 2002).

In this study we examine two unanswered but important questions regarding open source participation. What motivates developers to participate in open source projects? And, which motivations are most salient in explaining open source participation? With some exceptions, theorizing about the motives for open source participation has largely been left to the advocates inside the open source community. Although there has been significant discussion concerning which explanations are best suited to describe open source participation, there are few rigorous empirical studies of open source participation. Given the increasing importance of open source methods to software producers and open source products to software consumers, it is vital to seek a more formal approach of theory formulation and empirical validation to explain differences in open source participation. For producers of software this means understanding how to get things done in open source projects. For consumers of open source software this means understanding how things get done, thereby reducing some of the uncertainty associated with open source adoption. As is the case for software development in general, the viability of OSS projects ultimately depends on the individuals who staff the projects. As observed by Strasser (2001, p. 44), "the viability of open source software as a serious alternative to proprietary software hinges upon the resolution of the incentive issue".

Our study contributes to the growing body of research on OSS in several important ways. First, our work contributes to the research on OSS participation by drawing upon well-established theories of individual motivation to derive the motivational profiles of OSS participants. Applying factor analytic techniques to analyze survey data collected from open source participants, we find empirical support for five distinct factors underlying individual motivations to participate in open source projects. Our findings lend support to a functional explanation of open source participation. An additional and significant contribution of our work is establishing the relative importance of underlying motivations through the novel application of conjoint analysis in the OSS domain. Group analysis of the conjoint results reveals that motivational orientations are significantly different between contributors who exhibit low project commitment and those who exhibit high project commitment. The relative strength of motivations has significant practical implications for software producers contemplating the use of open source methods and software consumers seeking informative measures on which to evaluate OSS projects viability. Thus, our research provides the first clear picture of the underlying motivational structure of open source participation and the relative importance of those motivations within that structure.

In the following sections, we adopt a functionalist view of motivation, and identify five functional dimensions from the literature on volunteerism that are relevant to the open source context. We then consider the literature on open source software development and derive three additional functional explanations for OSS participation.

A Functional Explanation for Open Source Participation

The motivation for open source participation can be investigated from a variety of theoretical perspectives including social psychological, cultural and economic. Eric Raymond, an evangelist of the open source movement, popularized social psychological and cultural explanations of open source participation. In the cultural view, open source functions as a 'gift' culture, where the reputation of a programmer is primarily determined by his or her free contributions. As a second explanation, Raymond offers a 'craftsmanship' model where the artisan aspects of programming motivate developers to create works to be admired not only by themselves but also by others. In both cases, developers are motivated through the recognition of their contributions by their peers (Raymond 2001a, 200b). Such explanations find theoretical support in social psychology (Mauss 1967; Clary et al. 1998).

Social psychology is a branch of psychology that studies individuals in their social context. It is the study of how and why people think, feel, and do the things they do depending upon their situations. There are numerous theoretical bases of motivation and performance in social psychology including social exchange theory (Blau 1964), psychological contracts theory (Rousseau 1995), helping behavior and volunteerism (Benson et al. 1980; Clary and Orenstein 1991), work design (Hackman and Oldham 1980) and others. Both social exchange theory and psychological contracts theory are premised on the notions of an exchange relationship between two actors (typically an employee and an organization) where the relationship is based upon norms of reciprocity or rules governing the social exchange (Settoon et al. 1996) or upon promises made or implicit in the actors' interactions (Rousseau 1995). As Rousseau notes, in remote or distributed environments where the actors do not meet face-to-face (such as in open source software development), the formation of exchange relationships could be difficult and thus relatively unlikely. Indeed, recent empirical studies of field support for open source software projects have not found evidence of the operation of reciprocity in open source environments (Lakhami and von Hippel 2002). Thus, in this study, our focus is on the social psychological theories that closely relate to the motives most likely to operate in an open source context. Specifically, these include theories of helping behavior and volunteerism.

Functional Motivations for Open Source Participation

Helping behavior has enjoyed significant and long-lived attention as a field of psychological inquiry (Staub 1978). Research in this area distinguishes between spontaneous and planned helping behavior (Benson, Dohority et al. 1980). Volunteerism is often cited as an exemplar of planned helping (Clary and Snyder 1991) which "often calls for considerably more planning, sorting out of priorities, and matching of personal capabilities and interests with the type of intervention" (Benson et al., 1980, p. 89). As observed by Clary et al., volunteerism, as planned behavior, requires actively seeking volunteer opportunities, as well as deliberations about whether to volunteer and to what extent. A commitment to volunteer may require considerable personal costs such as time, effort and opportunity. That open source developers are, for the most part, volunteers is axiomatic. The question of interest is why do developers volunteer and what sustains this behavior? It is the voluntary, sustained, and ongoing nature of participation in open source projects that suggests such participation has motivational underpinnings. Theories of motivation seek to explain the processes that give behavior its energy and direction (Litwin and Stringer 1968). These concerns accurately reflect the nature of our inquiries regarding open source participation – that is, what motivates developers to participate.

Contributing to open source software involves voluntarily spending one's time in the coding of software. A functional approach to understanding motivation toward volunteer activities, one that emphasizes the social psychological purposes served by participation, is well established in the literature (Clary and Snyder 1991; Clary et al. 1998). One consistent theme of any functional approach is that people engage in the same activity in the service of different social psychological functions (Smith et al. 1956; Snyder and DeBono 1989). In the context of open source projects, developer contributions, while seemingly similar, may reflect distinctly different underlying motivational processes. As a consequence, the functions served by participation can significantly influence the outcome or products of participation.

Theorized social psychological functions served by volunteerism that are relevant to OSS participation include the following:

Normative: Motivation to volunteer is considered to serve a normative function when participation is initiated to better align oneself with the expectations or actions of peers or significant others. In the context of open source development, OSS participation may be esteemed within one's peer group such as on a college campus or within an online community. Thus, an individual motivated by a normative function will contribute to OSS projects to gain the respect of significant others.

Values: Motivation to volunteer is considered to serve a "values" function when participation is undertaken out of solidarity with the core beliefs or values of the organization. One example of this type of motivation in open source would be a common dislike for all things Microsoft that is prevalent in several OSS communities and projects (Stallman 2001). An individual motivated by such a values function would contribute to OSS projects because he or she shares the core value that software should be "free".

Understanding: Motivation to volunteer is considered to serve an understanding function when participation is initiated to facilitate general learning or the exercise of knowledge, skills and abilities. In the context of open source, an individual motivated by an understanding function may undertake OSS participation to become more familiar with the WEB in general or to become familiar with a technology out of intellectual curiosity.

Career Concerns: Motivation to volunteer is considered to serve a careers related function when participation is undertaken for the purpose of acquiring marketable career related skills or for the maintenance of existing skills. A careers based explanation of OSS participation is one of job training – where participation serves as on-the-job-training. For example, an individual motivated by career concerns could participate in OSS projects in order to gain or update programming skills.

Ego Enhancement: Unrelated to egotism, motivation to volunteer is considered to serve an ego enhancement function when participation involves a motivational process that centers on the ego's growth and development including self-esteem, self-improvement or personal growth. Similar to understanding, an ego enhancement explanation of OSS participation would involve the desire to feel better about oneself or one's abilities through accomplishment in an OSS project. Contributing to an open source project's code base and being acknowledged as a contributor could serve as a powerful motivation for an individual seeking to fulfill an ego enhancement function.

Other Functional Motivations for Open Source Participation

In addition to the various motivations to volunteer, there are several other functional motivations for open source participation in the OSS literature. In a recent article, Hars and Ou (2001) explore various theoretical bases for OSS participation and report the results of an exploratory survey. Two broad categories of motivation are identified – internal motivations and external rewards. Some argue that in work activities, an ideal internal motivator derives from the work content which should be satisfactory and fulfilling for workers (Hackman and Oldham 1980; Osterloh and Frey 2000). According to the job characteristics model, five core task dimensions (variety, identity, significance, autonomy, and feedback) influence perceptions of the motivating potential of tasks as well as work performance. In open source development both variety and autonomy are likely to be high as participants can select the different features they want to add or repair. Identity is also likely to be high as contributions are self-contained pieces of work. Task significance and feedback could be high, depending on the kind of code contributed and the reactions of others in the community to the contribution. Thus, contributing to OSS projects is likely to be internally motivating, i.e., developers may contribute because they enjoy it. This suggests that open source tasks, and hence participation, may serve a recreation function for OSS participants.

A source of motivation related to ego enhancement is reputation. Participating in open source projects may serve to build up one's reputation or status within the community. For example, Moon & Sproull (2002) cite a motivating "credit" policy as an important component of any successful open source project. A credit-where-credit-is-due philosophy may be critical in building and maintaining reputation within an open source community. Similarly, Bezroukov compares OSS projects to academic and scientific communities driven by competition for status and reputation (Bezroukov 1999). Thus, motivation to participate may to serve a reputation enhancement function for OSS participants.

Finally, we consider the prospect that the need for specific functionality from the open source software itself may play a role in motivating OSS participation (Hars and Ou 2000). The need for specific functionality is often cited by OSS developers as the reason they initially became involved in their projects (Bollinger 1999; DiBona et al. 1999). A study by Hertel et al. (2003) of Linux participation reveals pragmatic motives to improve the usefulness of the software as an important motivation for contributing. Indeed, increasing the use-value of the software has been cited as the primary motivation to initiate many of the projects that ultimately have become some of the most successful of open source projects to date. Lerner & Tirole (2002) report the results from 3 case studies into the origins of 3 popular open source projects – Apache, Perl, and SendMail. In all 3 cases, personal use-value of the software was cited as the primary motivation for the initial creation of the software. Thus, motivation to participate may derive the need to increase the use-value of the software for OSS participants.

In summary, drawing upon the literature surrounding volunteerism and job-design we have adopted a functionalist view of motivation to identify eight functionally based motives to participate in open source projects. These hypothesized functional motivations include normative pressures, shared values, understanding, career concerns, ego enhancement, recreation, reputation, and use-value.

Research Methods and Results

To identify and analyze the motivational profiles of open source contributors, we used a mixed methods approach (Tashakkori & Teddlie 1998) involving a survey and a conjoint analysis exercise. We surveyed the developers of three open source projects under the control of the Apache Software Foundation (ASF). The Apache HTTP (WEB) server and associated projects are some of the most successful open source products to date. The Apache server, the original ASF project, and its derivatives, have a dominant 63% share of the WEB server market (Netcraft 2003). Since its inception, the Apache WEB server has had over 7,000 source code contributions from more than 400 different open source developers (Mockus 2000). Although any of the Apache projects could provide an interesting vehicle to explore our research question, we have chosen to concentrate our data collection efforts on the HTTP, Jakarta and XML projects.[1] The Apache HTTP server project is a freely available source code implementation of an HTTP (WEB) server and is the project around which the Apache Group initially formed. The Jakarta project consists of all Apache related server side Java projects and includes more than 18 Java related subprojects. The Apache XML project is home for Apache XML related activities and has over 9 XML related subprojects.

The data for this study comes from a targeted, secure, WEB-based survey of Apache developers. The survey consists of two instruments designed to elicit the motivational profiles of the developers (instrument one – see Appendix II-A) and to reveal the relative importance of each motivation in influencing their participation (instrument two – see Appendix II-B). Using archival data on the Apache projects, we identified 237 Apache contributors.[2] The survey was introduced via e-mail, and data were collected in 2003. Eleven e-mail invitations were undeliverable. Of the remaining 226 contributors, 122 completed the first instrument yielding a response rate of 54%, and 86 completed the second instrument yielding a response rate of 38%. A detailed description of the methods, analysis and results for the survey and the conjoint analysis follows.

Identifying the Motivational Profiles of OSS Developers: Survey & Factor Analysis

To assess the social psychological functions served by volunteerism in an open source context, we adopted the Volunteer Functions Inventory assessment (VFI) instrument (Clary, Ridge et al. 1998). Because the instrument was originally conceived to measure the motivations of volunteers in non-profit organizations, we adapted the VFI questions to the OSS environment. In addition, we added measures for the three functional OSS motivations identified earlier – recreation, reputation, and use-value.

---

[1] We limit our attention to these 3 projects for practical reasons. First, these projects are by far the largest, both in terms of the number of developers and the number of contributions. Second, access to archival data for these projects has proved to be less problematic than for some of the smaller projects.

[2] Contributors' contact information (email) was extracted from publicly available Apache project archives using a series of tools developed for that purpose. Specifically, metadata from the source control software used by the Apache projects was 'mined' to associate Apache developers with their specific contributions.

We used exploratory factor analysis to identify the latent variables that account for the correlations among responses regarding our hypothesized functional dimensions of OSS participation (Cattell 1988). Although the VFI instrument has been validated in numerous settings, we had adapted the instrument to the OSS setting and had also added questions to the instrument to assess three other constructs. Thus, given our modification and enhancement of the VFI instrument, exploratory (not confirmatory) factor analysis is most appropriate. Several considerations are important to justify the use of factor analytic techniques. First is sample size. Cattell offers a heuristic of a 2:1 ratio of observations to items while (Gorsuch 1983) argues for an "ideal" ratio closer to 5:1. A total of 122 responses were available for this analysis resulting in an acceptable ratio of 5.1:1 of observations to items. Another issue deals with identifying the underlying factors or latent constructs. There are two considerations here – significance of factor loadings and simplicity of factor structure. While there is no precise lower bound where one should recognize factor loadings as significant, we used a cut-off of .45 to determine significant factor loadings. This lower bound is well above the published criteria of .30 deemed acceptable in exploratory studies such as ours (Sethi and King 1991).

Our common factor analysis procedure specified a principal axis initial extraction method using communality estimates set to the squared multiple correlations between the items. Preliminary factor analytic results of the scale items were obtained and analyzed from an unconstrained factor solution using orthogonal (varimax) rotation. Nine items failed to meet our minimum factor loading threshold of .45 and were dropped from analysis. One additional item cross-loaded on two or more factors at a level of .35 or greater and was also dropped from analysis. Thus, 14 items were retained for final analysis. The final analysis employed an oblique (promax, power 3) rotation as we did not want to presume orthogonality of the factors. Again we used a combination of several criteria to determine the appropriate number of factors to be retained (Loehlin 2004). First, given an average initial communality estimate of .54, an adjusted Eigen value criterion would recommend a 4 factor solution. Examination of the scree plot, however, recommends a 5 to 6 factor solution. Examining the proportion of the common variance accounted for by the factors in this case also recommends a 5 factor solution as the first 5 factors account for nearly 100% of the initial common variance estimate. We interpret the above results to indicate a 5 factor solution. As an additional check of the 5 factor solution, the analysis was repeated while constraining the results to a 5 factor solution. We then examined the resulting off-diagonal elements of the residual correlation matrix. In nearly all cases the residual values are less than .05 (all values below .07) with a root mean square off-diagonal residual of .032, indicating that the original correlations can be accurately reproduced from the 5 factor solution.

Table 2.1 - Survey Item Descriptive Statistics and Factor Scores

Question #	*Mean Response*	*Std Dev*	*Factor 1 use-value*	*Factor 2 Reputation*	*Factor 3 Career*	*Factor 4 Social*	*Factor 5 Recreation*
1	3.98	2.05	**76**	-8	-3	5	-10
2	3.80	1.91	**66**	-3	7	4	-2
3	3.43	1.97	**73**	4	-4	6	3
4	4.66	1.68	-16	**67**	22	1	14
5	5.43	1.41	-9	**62**	19	7	30
6	4.58	1.63	14	**65**	24	19	13
7	4.02	1.83	6	29	**60**	15	-1
8	4.74	1.67	-20	23	**68**	8	3
9	4.62	1.65	13	13	**70**	16	11
10	3.80	1.83	15	4	23	**60**	15
11	3.17	1.65	1	20	13	**97**	11
12	6.50	0.84	-11	10	7	13	**82**
13	5.50	1.73	1	20	-2	4	**56**
14	5.24	1.66	2	20	22	27	**48**

Table 2.1 Notes– 'Question #' column refers to questions in Table 2.2 below. Factor loadings are multiplied by 100 and rounded to the nearest integer

Cronbach's alphas were calculated to assess the internal reliability of each factor with all values exceeding the recommended cutoff of .70 (Nunnally and Bernstein 1978). Interpretation and discussion of the five factors that emerged from the exploratory factor analysis follow.

use-value motivations stem from a personal need for OSS product functionality that is currently not implemented or missing due to defects. Factor 1 consists of 3 items originally intended to capture motivations consistent with a use-value interpretation. This latent variable (factor) is labeled ***use-value***. The use-value factor accounts for 46.8% of the common variation of responses in our analysis. Reputation motivations are the result of needs to garner stature within the OSS community. Factor 2 consists of 3 items originally intended to capture motivations consistent with a reputation interpretation. This latent variable is labeled ***Reputation***. The Reputation factor accounts for 24.0% of the common variation of responses in our analysis. Career related motivations are driven by concerns about one's career - that OSS participation may serve as some form of job-training. Factor 3 consists of 3 items originally intended to capture motivations consistent with a career concerns interpretation. This latent variable is labeled ***Career Concerns***. The Career Concerns factor accounts for 13.3% of the common variation of responses in our analysis. Normative motivations for OSS participation stem a normative reaction to peers or significant others with whom respondents share similar interests. Factor 4 consists of 2 items originally intended to capture motivations consistent with the normative interpretation. This latent variable is labeled ***Normative***. The Normative factor accounts for 5.5% of the common variation of responses in our analysis. Recreation motivations are the result the motivational nature of the task (primarily programming) and that the task is internally motivating and thus serves as form of entertainment. Factor 5 consists of 3 items originally intended to capture motivations consistent with a recreation interpretation. This latent variable is labeled ***Recreation***. The Recreation factor accounts for 10.5% of the common variation of responses in our analysis. See Table 2.2 for a summary of factor scores, Cronbach's alpha scores, and questions that inform each factor.

Identifying the Relative Importance of Motivations: Conjoint Analysis

To assess the relative importance of the different motivational factors, we implemented a conjoint exercise. Conjoint analysis was first used by researchers in marketing and product development (Green and Rao 1971; Green and Wind 1975), but is now used in various disciplines ranging from agriculture to capital budgeting and healthcare (Wigton et al. 1986; Detsky et al. 1997). In applying this method, respondents make tradeoffs of perceived "benefits" against perceived "costs". The data from this exercise allow us to estimate a respondent's "utility" for various dimensions of hypothesized benefits and costs. To keep the conjoint tasks to a manageable size, Green and Srinivasan (1990) recommend that the number of attributes be limited to six or fewer where possible. Based on an earlier exploratory factor analysis, we selected the five unique motivational dimensions: Career Concerns, use-value, Reputation, Recreation, and Social / Normative.

Table 2.2 - Motivation Factors

Factor (Cronbach's α)		*Survey Questions*
USE-VALUE (.76)	1	My patches are for my or my company's benefit. If it helps others that's just a bonus.
	2	If a bug isn't affecting me or my work, I'm not likely to fix it.
	3	The main reason I contribute to this open source project is to make my paid job easier.
REPUTATION (.75)	4	Others in this project are aware of my technical skills.
	5	I am motivated when my technical achievements are recognized by others in this project.
	6	Open source projects give me a chance to showcase my technical skills.
CAREER CONCERNS (.74)	7	Contributing to an open source project can help me get my foot in the door at a place where I would like to work.
	8	By making contributions to an open source project, I can make new contacts that might help my business or my career.
	9	Open source experience will look good on my resume.
SOCIAL/NORMATIVE (.78)	10	Others with whom I am close place a high value on making contributions to open source projects.
	11	Making contributions to an open source project is an important activity to the people I know best.
RECREATION (.70)	12	Developing software is fun. I enjoy it.
	13	Writing software is a hobby of mine.
	14	People who know me, know that I like to write software.

Table 2.2 Notes - Respondents were asked "How accurate is this statement in describing your motivations for open source participation?" Responses were scored on a Likert scale – values 1 (very inaccurate) to 7 (very accurate).

Each of these five dimensions was operationalized with three levels in the conjoint analysis (i.e., with low, medium and high values to be used by individuals for making tradeoff decisions). The Career Concerns dimension is predicated on the belief that participation in OSS projects is sometimes rewarded by current and future employers. Thus, this dimension is operationalized as the anticipated annual dollar impact of OSS participation and has the following levels – \$0 (no impact), \$740, and \$2,100.[3] The use-

[3] Values for the Career Concern dimension derive from correlational results reported in Hann et al. (2002) and from Computerworld 2002 Salary/Skills Survey. Values for the Use-Value dimension derive from

value dimension is operationalized as the time participants might expect to save as a result of their contribution. Such time savings might take the form of increased efficiency in the performance of their jobs or less downtime as a result of their contributions. The anticipated time savings resulting from increased use-value is operationalized as 17, 34, or 51 hours per year. To operationalize the Reputation dimension we take advantage of the Apache "career" structure.[4] Specifically, as a result of their contributions, participants may anticipate achieving either the rank of Developer, Committer, or ASF member. The Recreation and Normative dimensions are operationalized as simple ordinal scales of "very high", "moderate", and "very low".

In addition to the five factors discussed above, we add two control dimensions that are relevant for making trade-off evaluations: (1) the amount of effort required by participation and (2) contribution type and the objective of the effort dimension is to impose a "cost" of participation. Many OSS participants devote significant amounts of time to their projects (Torvalds and Diamond 2001). If OSS participation were costless, higher levels of the five previously mentioned factors would always dominate lower levels. Thus, the Effort dimension is operationalized as the time a participant might expect devote to their OSS contributions and has values of 15, 35 and 100 hours per year.

Not all programming tasks have the same level of desirability (Weinberg 1998). While OSS participants are free to self-select the projects and tasks on which they would like to work, all options may not be open to them. For example, it is now well established that within the Linux community only a select few "lieutenants" regularly make contributions included in the Linux kernel (Browne 1998; Moon and Sproull 2002). The objective of the contribution type dimension is to capture the scarce nature of the most desirable work on many popular OSS projects. By motivating this notion of scarcity, we capture an

---

Kemerer (1999), Slaughter (1998) and Banker, et al. (1998). Values for the Effort dimension are from Hann et al. (2002).

[4] There are five observable levels of recognition or "rank" within the ASF. In order of increasing status, these are developer, committer, ASF member, project management committee member, and ASF board member. In all cases advancement in rank is in recognition of an individual's commitment and contributions to an Apache project. Thus, rank symbolizes both status (reputation) and the degree of accomplishment within ASF projects.

essential tradeoff of OSS participation. As a result, this dimension is designed to capture a realistic set of tasks common to all OSS projects. Thus, this dimension is operationalized as "Adding New Functionality", "Enhancing Existing Functionality", and "Fixing Software Bugs" (defects).

Taken together, our conjoint analysis assessed trade-offs along the seven dimensions set out above. Based on these seven dimensions and their treatment levels, there were a maximum of $3^7 = 2187$ possible conjoint stimuli. To avoid asking subjects to rank too many alternatives and following advanced techniques in conjoint analysis, we apply individualized hybrid models, which combine compositional, decompositional and hybrid preference models (Green 1984; Green and Krieger, 1996).

Hybrid conjoint models combine a two stage compositional procedure and a decompositional conjoint exercise. Specifically, the contributor's preference for various dimensions is measured in three separate steps. Steps 1 and 2 comprise the compositional aspects of the model (referred to as the self-explicated stage) while step three comprises the decompositional. In step 1, the respondent is asked to assign a weight to the various cost and benefit dimensions. In step 2, the respondent rates the desirability of the levels of each dimension. And in step 3, the respondent rank orders a set of hypothetical open source participation scenarios (referred to as full profiles or conjoint stimuli) and assigns a probability of participation. This allows us to estimate the contributor's utility function, which assigns a part-worth to each benefit and cost dimension. Essentially, the part-worth is the marginal utility of the dimension in the contributor's ranking of the conjoint scenarios. Establishing a part-worth facilitates comparison of economic benefits such as salary increases with non-economic benefits such as a raise in status or reputation. Hence, one can assess the weight that open source contributors give to various factors underlying their participation decisions.

The hybrid approach has emerged as the method of choice in situations involving complex tradeoffs or a large number of dimensions or attributes. There are several models used to analyze the results of hybrid conjoint data. We have chosen a model that has shown superior empirical performance in a recent review of competing hybrid models – the Modified Importances / Desirabilities model (Green 1984). The following equation defines this model:

$$y_{ml} = u_{ml} A + \varepsilon ;\ m = 1..M ;\ l = 1..L_m$$

$$Where :\ y_{ml} = w_m d_{ml} \text{ and } (s_r - \upsilon) / \tau = \sum_{m=1}^{M} \sum_{l=1}^{L_m} u_{ml} I_{rml} + \delta_r .$$

The implementation of this equation requires that we estimate the final part worth (or utility) of attribute m at level l to an individual in two stages. In the first stage we estimate a multiple regression with 21 observations of self explicated likelihood-of-participation scores derived using data collected during the compositional stage of the conjoint exercise. In addition, to the derived likelihood scores we add 8 likelihood-of-participation scores resulting from the decompositional stage of the conjoint exercise for a total of 29 observations per respondent. In the second stage we perform a simple regression of the full profile scores (s_r) on the current predicted score for profile r based on the current estimates of the part-worths (u_{ml}). The intercept and slope from this simple regression become the updated the location parameters (υ and τ) for the next iteration of the multiple regression. This procedure continues until the sum of the squared errors from the multiple regression decreases by no more than some user supplied amount, epsilon.

The key outcome of conjoint analysis is the part-worths (marginal utilities) of the various dimensions that comprise the conjoint stimuli. To estimate the part-worths, we used the Modified Importances / Desirabilities model outlined in the previous section. Recall that the dependent variable in our model is a vector of likelihood-of-participation scores from two sources: (1) the full profile scoring and (2) those derived from the self-explicated data collected during compositional phase. The independent variables in this model are indicators of the various levels of the five hypothesized dimensions of OSS motivation and the two control dimensions. The coefficient of each independent variable, then, would be the part-worth corresponding to that level of the dimension. Further, we calculated the relative importance of each dimension. We express the relative importance as a percentage by dividing the part-worth corresponding to the maximum level of that dimension divided by the sum of the part-worths corresponding to the maximum levels of all other dimensions. Table 2.3 shows the estimated part-worths and relative importance.

Table 2.3 - Part-worth Values and Relative Dimension Importance

Dimension	*Level*	*Part-Worth*	*Estimated $ Value*	*Relative Importance of Dimension*
Reputation	Developer	0.0722***	0	10.50%
	Committer	0.0878***	728	
	ASF	0.0815***	433	
Contribution Type	Bug Fix	0.0906***	0	12.36%
	Enhancement	0.1065***	744	
	Add Features	0.096***	254	
use-value	17 Hrs	0.1763***	0	26.97%
	34 Hrs	0.2013***	1164	
	51 Hrs	0.2094***	1545	
Effort	100 Hrs	0.066***	-1074	11.46%
	35 Hrs	0.0856***	-157	
	15 Hrs	0.089***	0	
Recreation Value	None	0.1256***	0	18.86%
	Moderate	0.1545***	1347	
	High	0.1464***	970	
Social Value	None	0.0481*	0	7.94%
	Moderate	0.0360*	-628	
	High	0.0617**	634	
Career Impact	0 USD	0.0475**	0	11.91%
	700 USD	0.0496**	700	
	2,100 USD	0.0925***	2100	

Conjoint Dimensions

First, we examine whether the respondents' preferences over the stimuli (OSS participation dimensions) show significant variation with a dimension. If the part-worth for a particular level of a dimension differs significantly from zero, then the evidence suggests that respondents are willing to make trade-offs between that level and the other levels within that dimension holding the other levels of the remaining dimensions constant. With the exception of low levels of the Social dimension, respondents are willing to make trade-offs among all levels for all dimensions of OSS participation motivations. For example, the part-worth for a use-value of 51 hours is .2094 and is statistically significant. This means that, on average, participants would be willing to increase their likelihood-of-participation estimate by 3.31% as compared to an identical experience having the base use-value level 17 hours (.2094 - .1763 out of 100%). A similar calculus can be performed for all levels within each dimension.

Of particular interest, however, is the Career Concern dimension. Our results can be used to calculate the marginal utility for each 100 dollars of Career Concern. Referring to Table 2.3, the part-worth for a Career Concern of $0 (no impact) is .0475 and is statistically significant. Following the procedure above, participants would be willing to increase their likelihood-of participation estimate by only .21%, or 3.0E-4 per 100 dollars of impact. Alternatively, between the $700 and $2,100 levels of Career Concern, the $1,400 increase raises the likelihood-of participation estimate by 4.29%, or .003 per 100 dollars of impact. These two estimates provide a range of 3.0E-4 to .003 per $100 of impact.

Using the Career Concern marginal utilities and the part-worths for any of the other dimensions, we can estimate the dollar "value" of that dimension level on a per-respondent basis. As an example, consider the use-value dimension. From Table 2.3, the part-worth for 51 hours of use-value is .2094. Using the upper-bound for the Career Concern marginal utility (.003 per $100), the value is $1,030. Using the Career Concern marginal utility between $0 and $2,100 (2.14E-3 per $100), the value is $1,545.

As a check on the external validity of our results, notice that the per hour dollar value from the use-value calculation above is close to the hourly market value of quality software developers. Using the average senior systems analyst salary reported in the 2002 Computerworld survey (Anonymous 2002) the market value of one hour is $38.15 ($76,300/2000 hours). Accordingly, a use-value of 51 hours would have a market value of $1,945 - very near the $1,545 value estimated in the preceding paragraph using the Career Concern marginal utilities. We can use the same method to derive estimated dollar values for the remaining dimension levels. Using the Career Concern marginal utility between $0 and $2100, the estimated dollar value for the remaining dimension levels are reported in column 5 of Table 2.3.

Conjoint Dimension Importance

It is clear from column 5 of Table 2.3 that the use-value is a significant motivational dimension for open source participation. However there is qualitative (Shaw 2003) and quantitative (Roberts et al. forthcoming) demonstrating that the use-value motivation is correlated with lower participation measured as source-code contributions.

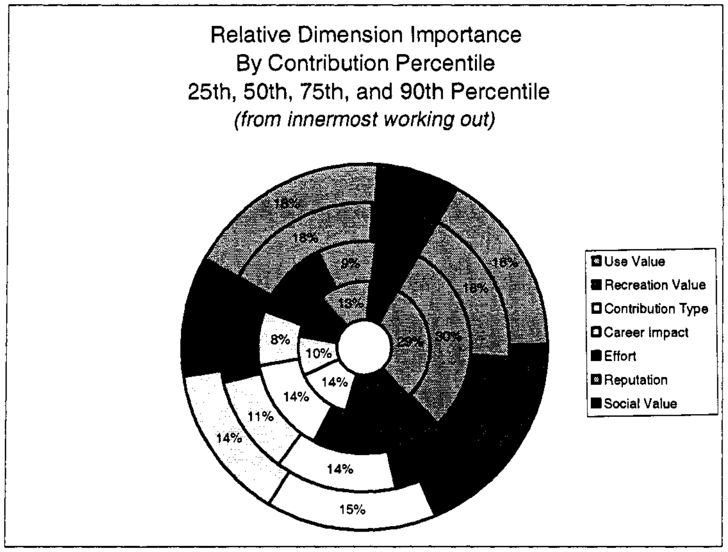
To explore participant motivations by level of contribution I assigned participants to a contribution class based on the sample contribution percentiles.[5] The 25th, 50th, 75th, and 90th contribution percentiles were each used to divide the sample. For each classification the part-worth scores were recomputed and compared across groups. Figure 2.2 shows that as participation increases there is indeed a change in motivational profile of the participants. Tests of mean dimension part-worths between levels of classification show that these differences become significant at the 75th percentile of contribution (44 contributions). At this level of contribution the 'trade-off' made by developers is between the use-value and Reputation dimensions – average use-value decreases by 16% ($p <.01$) while Reputation increases by 9% ($p < .01$).

In addition, not only is the difference between the use-value of participants in the 75th contribution percentile significantly different from those below the 75th percentile, use-value is no longer the dominant motivational dimension. At 18%, use-value is replaced by Recreation and Reputation (tied) as the dominant motivations at 20% and 18% respectively. As shown in Figure 2.2, this effect increases with the level of contribution. A sensitivity analysis of contribution levels between the 50th and 75th percentiles shows that the mean difference of the use-value dimension between participant classes actually becomes significant at 16 contributions to the project – far fewer than the 44 contributions of the 75th percentile.

Figure 2.2 - Dimension Importance by Contribution Percentile

---

[5] Measures of participant contribution data are taken from Studies 2 and 3 of this volume.



While the above analysis is compelling evidence that motivational profiles differ with respect to participant commitment to the project, use-value remains an influential dimension at all levels of contribution and, as discussed above, seems at odds with the evidence that use-value motivations are associated with below average levels of contribution. Perhaps part of the explanation lies in the interpretation of the conjoint results. The conjoint experiment is designed as a multiattribute preference measurement device for analyzing trade-offs among the attributes (or dimensions) of interest (Green et al. 2001). The dimension part-worth values shed light on the likelihood of participation verses non-participation – a binary decision. That use-value is an important dimension for the participation decision is not inconsistent with the inability of use-value motivations to sustain contributions. Indeed, this is the conclusion drawn by Shah (2003). Shah found that while use-value motivations were the most common reasons cited for initial project participation, use-value motivations were not associated with sustained project participation.

Thus, motivational profiles do vary according to demonstrated commitment to the project measured as contributions. In particular, higher levels of project commitment are associated with higher levels of Reputation and Recreation motivations and lower levels of use-value motivations.

Discussion and Conclusions

In this work we empirically investigated the motivations of open source contributors for three Apache projects. Based on theories of motivation and volunteerism, we identified functionally based motivations to participate in open source projects. A survey instrument and a subsequent factor analysis were used to explore the appropriateness of our hypothesized motivations. Of the motivations, five (normative, career impact, recreation, reputation, and use-value) were related to latent variables or factors that held significant explanatory power for the observed correlations among the observed variables. A hybrid conjoint exercise was introduced to investigate the relative importance of the motivations to contribute to open source projects. Of the relevant factors, use-value and recreational value were found to be of highest relative importance, followed by career concern and reputation. A subgroup analysis revealed that motivational profiles do vary according to demonstrated commitment to the project (measured as contributions). In particular, higher levels of project commitment are associated with higher levels of Reputation and Recreation motivations and lower levels of use-value motivations.

Gaining an understanding of the motivational structure of contributors is a critical first step in evaluating open source as a viable development model for commercial software engineering endeavors (Strasser 2001). Corporations such as IBM, SUN and AOL Time Warner have embraced open source development as a new paradigm for software engineering (Koch 2005). The information presented in this research provides important empirically grounded insights that can be used to inform the structure and operation of open source projects. This research has provided empirical evidence for the dimensions of open source motivations and that motivations to participate in open source projects are functional in nature. We have shown that participation motivations have different relative importance and that participation behavior is related to motivations in important ways. The functional nature of motivation should be considered during the design and operation of any open source project.

It is important to recognize that contributors' motivations are not mutually exclusive and that one or more motivations may be more or less important for any given contribution. As the data is cross-sectional, we can not say with confidence that motivations follow any predictable evolutionary path. That is, for example, whether use-value motivations tend to give way over time to reputation motivations or vise-versa. Studies of volunteerism, however, have demonstrated that organizations can make use of the function nature of motivation to attract and subsequently retain contributors by assuaging the psychological needs that motivated their volunteerism behavior (c.f. Clary et al 1998). Open source projects, as organizations that rely to varying degrees on volunteerism, could use a similar tactics to attract and retain open source developers. Projects should be aware of the functional nature of open source motivations and use this information to influence project structure. For example, to satisfy the reputation motivation the use of feedback and performance recognition mechanisms should be incorporated into the project's technical structure (presumably through project WEB site design) and organizational culture (through enactment).

The group analysis demonstrated that the use-value motivation is associated with below average levels of contribution. However, use-value motivations are commonly sited as the raison d'être of the majority of open source projects and thus play a critical role in the initiation and continued operation of the project (cf. von Hippel, 2001, Lerner and Tirole 2002, Raymond 2001a). The above discussion implies that projects should encourage use-value contributors where possible and conversely not discourage them by erecting barriers to entry in the form of "heavy" process and/or project management practices. Active consideration of contributor motivations as part of the open source project management equation should facilitate the "conversion" of contributors motivated by use-value to more durable motivations and should be a key concern or activity of project management.

Our results also have interesting implications for future research. First, our results would benefit from further validation in other open source contexts and methodologies. Our research approach is characterized by a deep sampling of the three major Apache projects. The results that we obtained are therefore strongest in internal and external validity with respect to the Apache projects sampled. While we have not encountered any systemic differences across the three projects that we have covered and have no indication why this should be the case with the other sub-projects, only a similar analysis in other open source projects can provide assurance. In addition, a validation through other methods would strengthen our results. For this work, we have relied on a survey instrument to provide us with data. Although the same individuals provided responses via the survey for independent variables (motivations) and dependent variables (utility), we see no reasonable alternative than to obtain the data from the same source. Motivation is a subjective construct. Only the individual involved can report on his or her own motivations (motivations can neither be observed nor inferred). Similarly, only the individual can express their personal utility for a particular motivation to contribute. However, future studies could relate self-reported motivations to objectively observed behaviors (such as documented source code contributions) to mitigate same-source bias.

CHAPTER 3.
STUDY 2: AN EMPIRICAL ANALYSIS OF ECONOMIC RETURNS TO OPEN SOURCE PARTICIPATION

Introduction

Open source software, i.e., public software development projects where participants can read, modify, and redistribute the software source code (OSI 2001), is arguably one of the most exciting phenomena in the IT industry today. Open source programmers have contributed industrial-strength software that has been recognized as a viable competitor in several product domains, including operating systems and WEB-server software. The open source movement has proliferated to other software domains, ranging from software development tools to office applications, and even computer games.

One widely debated question is why open source programmers contribute voluntarily, thereby foregoing any direct remuneration that they could accrue while working on a commercial system. Often quoted individual level motivations for participating in open source projects cover a broad spectrum including scratching a "personal itch" with respect to software functionality, enjoyment, and desire to be "part of a team" (Raymond 2001a). Others liken the open source community to a gift culture where the status of a participant depends on "what he gives away" (Raymond 2001b).

More recently, authors have used economic theories to explain participation in open source projects. von Hippel and von Krogh (2003) discuss a setting that combines private incentives in a public good (collective action) context and term this the "private-collective" model of innovation. Specifically, they argue that such an incentive model is possible because private returns such as enjoyment, learning, and superior reputation from open source involvement are not derived from the software code itself (which is a public good). Lerner and Tirole (2002) argue that open source participation yields two types of rewards: (1) immediate rewards that ensue from the increase in productivity (less the opportunity cost of time) and (2) delayed rewards relating to various career concerns such as one's future marketability. For the latter, a participant motivated by career concerns has incentive to signal his or her abilities to the labor market. This signaling incentive is likely to be stronger when performance is more visible to the relevant audience, and performance is informative about talent. Such an incentive is particularly relevant in the information technology industry for two reasons. First, programming is often viewed as more of an art than a skill (Weinberg 1998). Good programming is not confined to learning the syntax and specific features of a programming language and the practice of good documentation. Productive programmers are believed to have a certain aptitude that allows them to proceed logically from problem to solution, and in the process derive the most efficient and general software design possible. Subsequently, an effective programmer has to take the lead in propagating the software design to co-workers and in sharing sufficient insights such that the co-workers in turn can be productive. Hence, it has been documented in the software engineering literature that the productivity of a "star" programmer is an order of magnitude greater than that of an average programmer.[1] Second, the inability to formalize characteristics of highly productive programmers makes the programming process very difficult to evaluate. Not surprisingly, it often proves to be a challenge for employers to evaluate performance of the majority of programmers (Kirsch 1996).[2]

In addition to signaling imperfectly observable abilities, participating in open source projects has the potential to increase a contributor's human capital. In open source projects, contributors can select both the problem they want to attack and the implementation approach or solution. Once implementation is complete, other contributors provide timely feedback on the solution, ranging from identification of software defects to suggestions on how to improve the submitted software code (Raymond 2001a).[3] Hence, contributing to open source projects can be seen as learning experiences that increase the

---

[1] This view has been echoed by practitioners. For example, Jim Clark, founder of Silicon Graphics and Netscape, handpicked Pavan Nigam as the chief technology officer of WebMD (his third venture) because *"...the difference between a great software guy and an O.K. software guy is huge. A great software guy is worth 10 times an O.K. software guy"* (New York Times Magazine, Oct. 10, 1999).

[2] This difficulty with assessing software development performance has also been pointed out in the software contracting literature (Whang 1992; Ang and Beath 1993).

[3] Dubbed "Linus's Law", Raymond (2001a) notes the inverse relationship between the amount of time to identify/correct software defects and the number of independent developers simultaneously working on the problem.

programmer's knowledge. Inasmuch as this knowledge is transferable, open source participation increases the contributor's human capital.

In this paper we empirically investigate whether open source participation is consistent with theories in labor economics. As we have noted, contributing to open source projects can potentially be beneficial to contributors in two ways: (i) participation enhances existing skills or provides opportunities to gain new experience and hence makes contributors more valuable to employers and (ii) participation signals contributors' imperfectly observable productive characteristics to their employers. In order to measure the first benefit, we proxy the knowledge gained by contributors directly from the volume of their software source code submissions. To measure the second benefit, we exploit a unique setting of a specific open source project (the Apache Software Foundation) that ranks its members based on merit. From a signaling perspective, we maintain that certain abilities such as software design understanding and project leadership skills are endowments that are often difficult to evaluate. However, an open source community affords a venue in which such abilities can be discerned by highly skilled peers and rewarded with a higher rank. Using panel data collected on open source contributors and a fixed-effect specification of the standard wage equation to isolate contributors' time invariant qualities, we distinguish the learning effect and the signaling effect from time invariant characteristics such as intelligence in explaining participation.

We find that, in the context of the Apache open source projects, greater open source experience, as measured in contributions made, does not result in wage increases for contributors. This suggests that employers do not reward the gain in experience through open source participation as an increase in human capital. On the other hand, achieving a higher status in the merit-based ranking within the Apache open source community is associated with more than a 20% increase in wages, depending on the rank attained. Our results are consistent with the notion that a high rank within the Apache Software Foundation is a credible signal of the productive capacity of a programmer.

In the next section we present the rationale for open source participation. Section 3 describes our data setting and sources, the Apache projects, as well as the organization that governs the development process, the Apache Software Foundation. In Section 4, we develop the model used to estimate the delayed returns on open source participation and present the results. We conclude in Section 5.

Explaining Open Source Participation

The motivation for open source participation can be investigated from a variety of theoretical perspectives including social psychological, cultural and economic. Eric Raymond, an evangelist of the open source movement, popularized social psychological and cultural explanations of open source participation. In the cultural view, open source functions as a 'gift' culture, where the reputation of a programmer is primarily determined by his or her free contributions (Raymond 2001b). As a second explanation, Raymond (2001a) offers a 'craftsmanship' model where the artisan aspects of programming motivate developers to create works to be admired not only by themselves but also by others. In both cases, developers are motivated through the recognition of their contributions by their peers. Such explanations find theoretical support in social psychology (Mauss 1967; Clary et al. 1998).

Social psychology is a branch of psychology that studies individuals in their social context. It is the study of how and why people think, feel, and do the things they do depending upon their situations. There are numerous theoretical bases of motivation and performance in social psychology including social exchange theory (Blau 1964), psychological contracts theory (Rousseau 1995), helping behavior and volunteerism (Benson et al. 1980; Clary and Orenstein 1991), work design (Hackman and Oldham 1980) and others. Both social exchange theory and psychological contracts theory are premised on the notions of an exchange relationship between two actors (typically an employee and an organization) where the relationship is based upon norms of reciprocity or rules governing the social exchange (Settoon et al. 1996) or upon promises made or implicit in the actors' interactions (Rousseau 1995). As Rousseau notes, in remote or distributed environments where the actors do not meet face-to-face (such as in open source software development), the formation of exchange relationships could be difficult and thus relatively unlikely. Indeed, recent empirical studies of field support for open source software projects have not found evidence of the operation of reciprocity in open source environments (Lakhami and von Hippel 2000).

Economists have offered alternative explanations based on labor economics. Two possible views can be articulated: a human capital explanation and a signaling or sorting explanation. Human capital explanations for the value of open source participation are straightforward: participation allows developers to gain marketable technical skills (Becker 1962; Blaug 1976). An explanation for open source participation consistent with human capital theory would maintain that open source participation is an investment in training that leads to higher earnings in the future. As an investment, the choice to participate depends upon two considerations. First, the individual considers the opportunity cost associated with participation, and second, the individual considers the expected earnings in the job market after participation. Human capital theory predicts that the greater the investment, the greater the return. Therefore, higher earnings should be positively correlated with higher levels of open source participation. Thus, we hypothesize that:

H1: The greater the level of open source participation, the higher the wages.

While attainment of a skill may be an important result of participation, proponents of a signaling or sorting theory of labor markets argue that participation serves as a signal of imperfectly observable productive capacities to current and future employers (Weiss 1995). Given a distribution of inherent productivity among potential open source participants, the more productive developers would like to signal their superior productivity to employers (Spence 1973). As we noted in the introduction, this is especially applicable in the context of software development productivity. One study of "star" programmers, for instance, found that the top 1 percent were 1,272 percent more efficient than the average (Goleman 1998). At the same time, due to the nature of programming activities, it might be difficult for a programmer to convey fully his or her productive capacities. Programming, as a task, requires significant autonomy and creativity where the behaviors that transform inputs to outputs may not be well understood by management (Kirsch 1996). While it might be relatively easy to identify the "star" programmers, it is much more difficult to identify above average programmers who have a good understanding of the problem and often develop an efficient solution for the problem at hand. Further, the level of code contributions per se might not be the best indicator of productive capacity. Open source projects represent very large-scale, distributed development projects involving thousands of contributions from hundreds of developers (Mockus et al. 2000). High ability contributors typically make many submissions to the code base, but it could be the depth of their understanding, the efficient design of the solution, and their ability to persuade, to get people "on board" with their ideas and strategies that represent the true quality of their contribution (Moon and Sproull 2002; von Krogh et al. 2003). While possible, as a practical matter it is difficult for employers to efficiently evaluate these qualities based on individual code contributions. It seems reasonable then that employers seek a reliable proxy that is correlated with these desirable characteristics indicative of or obtained through successful open source participation. If potential employers can use open source participation as a signaling mechanism, then the existence of a "credential" or observable measure of successful participation would allow them to make inferences about a developer's productive capacity. In so far as open source participation is correlated with some desirable trait such as ability or motivation, it can be used by either employers to screen potential employees or by applicants to signal these desirable traits. Thus we expect:

H2: Successful open source project participation will be positively related to wages.

It is important to point out that some of these economic and "non-economic" explanations are overlapping. For example, a desire for a higher status among peers may be a strong incentive to contribute while serving as a signal to employers. However, a separation of these motives is, for our purposes, not necessary. As Spence observed, "A signal is a manipulable attribute or activity which conveys information ... in general it is not necessary to insist that the actor, in manipulating the attribute, think of himself as signaling or conveying information" (Spence 1974).

Data Setting and Sources

Our data set combines several primary data sources including archival data from large open source software projects, and three targeted surveys of open source participants. We briefly describe the setting of the data collection, each data source, and our measures of key variables.

Context: Apache Software Foundation Open Source Software Projects

We investigated three major open source projects under the control of the Apache Software Foundation (ASF). The ASF projects enjoy wide acceptance both in the open source software development community and in the marketplace where it maintains a dominant 64% share of the WEB-server market (Netcraft 2003). Similarly, the ASF projects consistently attract and retain the large number of participants vital for open source project success (von Krogh et al. 2003). For the years 1998 through 2004 the three ASF projects in this study incorporated over 173,000 changes from more than 1,300 different open source developers. The ASF is a not-for-profit corporation that provides the legal, organizational and financial infrastructure for the software projects gathered under the ASF open-source umbrella. Each of the ASF projects operates autonomously including all aspects of product development.

The ASF encompasses a number of subprojects related to the development of a full-featured WEB-server product offering. We studied the largest and most significant of these projects including the Apache server project which is a freely available source code implementation of an HTTP (WEB) server and is the project around which the Apache Group initially formed; the Jakarta project which consists of more than 18 related Java subprojects; and the XML project which includes over 9 related subprojects.

The Apache context is particularly well suited for an examination of economic returns to participation in open source development. Since its inception the Apache project has operated under a model of shared leadership and responsibility. This model of shared responsibility is reflected in the principles of the meritocracy that define advancement within the ASF (Fielding 1999). As a meritocracy, status, responsibility, and benefits are commensurate with contribution. There are five observable levels of recognition or rank within the ASF. In order of increasing status, these are developer, committer, project management committee member, ASF member, and ASF board member. As expected within a merit-based structure, the number of participants at any status level decreases with rank. In all cases, advancement is in recognition of an individual's commitment and contributions to an Apache project. This hierarchy within the ASF makes the Apache projects particularly appropriate for an evaluation of open source participation. As observed by Tyler, et al. (2000), data for identifying economic returns to a variable serving as a signal in labor markets should contain exogenous variation in the signal status among individuals with similar levels of human capital. Participants in ASF projects possess such a variable or credential – their rank within the ASF.

Anecdotal evidence suggests that both open source participants and IT labor markets value information regarding productive capacity revealed via open source participation. First, personal conversations between the authors and several participants of successful open source projects[4] revealed the participants' awareness of career related impacts of open source participation. Resumes of Apache contributors prominently mention declarations of contributions, specific technical accomplishments and, most interestingly, significant Apache project management responsibilities, as well as status within the ASF hierarchy. In addition, employers who seek software engineers or developers often require Apache specific experience.[5] It is important to note that software development involves skills that can often be transferred to other software projects (Basili and Caldiera 1995). If employers are aware that successful open source participation is correlated with some desirable, and somewhat fungible, trait(s) then a reliable "signal" of success such as rank within the ASF could serve as a sorting mechanism in the labor markets. Again, anecdotal evidence from technical recruiters suggests that this is the case.

Individual reasons for initial participation in any Apache project vary. Typical reasons cited include reporting a problem or "bug", or fixing a problem in the software that has become a nuisance or

---

[4] Sendmail and the Apache WEB-server.

[5] A representative sample of resumes of several Apache contributors and job listings requiring specific Apache skills are available upon request from the authors.

impairs usage. Another reason is to extend existing functionality or to add new features required by the user. For many contributors there is a single encounter with the project. Some developers, however, choose a deeper level of involvement and continue to make contributions. If developers' contributions are significant and consistent over a period of time they may be nominated for an increase in rank from developer to "committer." This promotion is an important advancement within the Apache community; it signifies that this contributor has obtained the privilege to submit code changes directly to the source code repository. All participants of rank developer must submit their code for review by a committer before it is accepted. Committers who continue their involvement in the project may be nominated for ASF membership by an existing ASF member. ASF membership is largely a matter of recognition and carries with it a certain prestige in the Apache community (Fielding 1999). ASF members are eligible to be nominated by the ASF Board of Directors or to serve on a project committee. Project committee members are responsible for all aspects of managing an Apache subproject including project plans and roadmaps, release schedules, etc. The ASF Board of Directors makes decisions regarding corporate governance as well as decisions regarding the addition of new projects under the ASF organizational umbrella.

Data Source: Archival Data on Apache Participants' Rank and Contributions

One of the basic tenets of open source software is that the development process and resulting products are "open" and freely available. As such, all open source work products are placed in the public domain under various "free software" licensing arrangements. Apart from the source and binary codes of the actual programs, Apache products include developer WEB sites, change logs, documentation, and developer communications in the form of email archives. From these products, we extracted two types of information: information pertaining to each individual's progression along the Apache career path, and information about each individual's source code contributions to the project. We discuss each in turn.

Our primary interest is the construction of an Apache career path for each contributor. Our objective is to capture upward progression within the five levels of the ASF meritocracy, i.e., developer, committer, project management committee member, ASF member, and ASF board member. Before actually joining an open source project, potential developers typically observe the "lay of the land" regarding both the form and substance of project membership (von Krogh et al. 2003). In our study, a participant's first source code contribution is considered a consummation of the joining decision, signaling entry into the meritocracy and the beginning of one's Apache career. During the time period prior to making an actual contribution, any latent contributors are considered to be outside the Apache meritocracy. To determine each individual's Apache career progression we used archival data from three sources: the Apache developer WEB site, contribution meta-data from the Concurrent Versioning System (CVS) revision control software,[6] and minutes from the Apache Board of Directors meetings. Each Apache subproject maintains a separate developer website that includes a list of contributors and project management committee members. By observing changes to the archives over time we are able to construct a time line for the promotion of individuals within each project. Progression up the Apache career ladder is captured as a series of discrete transitions. The first transition is from latent contributor to an Apache developer. The second transition occurs when a developer is granted commit access to the project source code archives and thus achieves the status of committer. Both of these changes in rank are derived from CVS meta-data. We observe initial contributions, and hence when developer status begins. Similarly, the transition from developer to committer occurs when contributors first exercise their newly acquired CVS commit privileges. Elevation from committer to the rank of project management committee member may occur either as a "field promotion" via a consensus among existing committee members or by appointment from the ASF Board. Attaining ASF membership requires nomination by an existing ASF member and election by secret ballot. New members are announced each year at the annual Board of Directors meeting. Thus, in our panel, the transition to either project management committee member or to ASF member occurs on the date of the announcement. Lastly, the rank of ASF board member is achieved via election by existing ASF members. Transition to this rank occurs on the date of recognition of new board members as recorded in board meeting minutes.

---

[6] Concurrent Versioning System is the de-facto standard source control system for OSS projects.

To extract information regarding individual contributions from the data, we developed tools to extract contribution information at the level of the individual developer. A submission to an open source project is known as a "patch." Patches are analogous to modification requests in traditional software development environments. Our research follows the method introduced by Mockus et al. (2000) to reconstruct patches from source code archives. For each patch we extracted and retained common software metrics including lines of code added and deleted, the date of submission, the names and number of source code files affected by the change, change log entries, and the list of patch authors.

We constructed a longitudinal data set of participant contributions by year. The longitudinal contribution data encompassed contributions to any of our three target Apache projects. Data collection was completed in Sept 2005 and included contributions from January 1, 1998 through December 31, 2004. See Appendix III for a more detailed description of the data extraction process.

Data Source: Survey Data on Apache Participants' Demography and Job History

To augment the longitudinal contribution data set outlined above, we collected demographic and job history data in three waves. Three secure WEB-based surveys of Apache contributors were conducted for this purpose. Of primary interest in each survey was the respondent's wages for the current and prior year. Dr. Roy Fielding, the then chairman of the ASF, introduced the first survey to 1,301 uniquely identified contributors via e-mail in November 2001 and was concluded in January 2002. Two hundred thirty-three e-mail invitations were undeliverable. Of the remaining 1,068 contributors, 325 completed the instrument, yielding a response rate of 30%. The second and third waves involved the 237 respondents from the first survey who agreed to participate in additional rounds of data collection. The second survey was concluded in January 2003. Eleven e-mail invitations were undeliverable. Of the remaining 226 contributors, 122 completed the instrument yielding a response rate of 54%. The third survey was concluded in January 2005. Thirteen e-mail invitations were undeliverable. Of the remaining 224 contributors, 96 completed the instrument yielding a response rate of 43%.

ASF projects are in essence globally distributed software development projects. As such, our sample includes wages reported in 25 different currencies. IT labor market characteristics can vary widely between countries (Schreyer and Pilat 2002). An econometric model that accounts for these cross-national differences would be difficult to implement requiring a number of currency and purchasing power related transformations. In addition, wage changes due to moves from one country to another pose another difficulty. Accordingly, we do not attempt to account for cross-national differences here. Rather, we limit our analysis to respondents who earned income in U.S. dollars. Lastly, a temporal relationship between open source participation and wages would prescribe a model involving lagged independent variables. Given this functional form, we retain only those cross-sections where both the dependent and independent variables result from a common labor market experience, viz. the U.S. Applying the above constraints yields a cross-sectional time series panel of 147 cross-sections (individual respondents) each having at least two years of reported wages for a total of 450 observations for any of the years 1999 through 2004. The instrument used in the first wave of data collection is shown in Appendix I. Subsequent data collection mirrors the data collected in the first wave.

Both participant ASF rank as well as the number of contributions plays a critical role in the analysis and interpretation of results that follow. To discern whether the distributions of rank and contribution in our sample are comparable to that of the population, we employed the Mann-Whitney U and Kolmogorov-Smirnov non-parametric tests of location, scale and empirical distribution. Specifically, we compare sample rank and contribution to that of the Apache contributor population. Both tests evaluate the hypothesis that, for the variables of interest, respondents and non-respondents are drawn from the same underlying population. The results of these tests indicate that, with respect to rank and contribution, these two groups are indeed drawn from the same underlying population, as we fail to reject the null hypothesis (For contribution, Mann-Whitney statistic=-.44, p= .66; Kolmogorov-Smirnov statistic=.80, p=.54. For rank, Mann-Whitney statistic=-.1.56, p= .13; Kolmogorov-Smirnov statistic=.86, p=.45). In addition, we also compared the Apache career path of participants in our sample with the Apache career paths of the overall Apache population. Table 3.4 shows the observed patterns of rank progression for our respondents over the period covered by the panel.

Table 3.4 – Patterns of Rank Progression in Sample

Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Freq	%
NORNK	NORNK	NORNK	DEV	DEV	DEV	43	29.25
NORNK	NORNK	NORNK	DEV	DEV	COM	2	1.36
NORNK	NORNK	NORNK	DEV	DEV	COM,PMC+	1	0.68
NORNK	NORNK	NORNK	DEV	COM	COM	1	0.68
NORNK	NORNK	NORNK	DEV	COM	PMC+	1	0.68
NORNK	NORNK	NORNK	DEV	COM,PMC+	PMC+	1	0.68
NORNK	NORNK	NORNK	DEV,COM	COM	COM	4	2.72
NORNK	NORNK	NORNK	DEV,COM	COM	PMC+	1	0.68
NORNK	NORNK	DEV	DEV	DEV	DEV	31	21.09
NORNK	NORNK	DEV	DEV	DEV	COM	1	0.68
NORNK	NORNK	DEV	DEV	COM	COM	1	0.68
NORNK	NORNK	DEV	COM	COM	COM	2	1.36
NORNK	NORNK	DEV	COM	COM	PMC+	2	1.36
NORNK	NORNK	DEV	COM,PMC+	PMC+	PMC+	1	0.68
NORNK	NORNK	DEV,COM	COM	COM	COM	5	3.4
NORNK	NORNK	DEV,COM	COM	PMC+	PMC+	1	0.68
NORNK	NORNK	DEV,COM	PMC+	PMC+	PMC+	1	0.68
NORNK	NORNK	DEV,COM,PMC+	PMC+	PMC+	PMC+	1	0.68
NORNK	DEV	DEV	DEV	DEV	DEV	8	5.44
NORNK	DEV	DEV	COM	COM	COM	1	0.68
NORNK	DEV	DEV	COM,PMC+	PMC+	PMC+	1	0.68
NORNK	DEV	COM,PMC+	PMC+	PMC+	PMC+	2	1.36
NORNK	DEV,COM	COM	COM	COM	COM	4	2.72
NORNK	DEV,COM	COM	PMC+	PMC+	PMC+	1	0.68
NORNK	DEV,COM	PMC+	PMC+	PMC+	PMC+	1	0.68
NORNK	DEV,COM,PMC+	PMC+	PMC+	PMC+	PMC+	1	0.68
DEV	DEV	DEV	DEV	DEV	DEV	25	17.01
DEV	COM,PMC+	PMC+	PMC+	PMC+	PMC+	2	1.36
PMC+	PMC+	PMC+	PMC+	PMC+	PMC+	1	0.68

Table 3.4 Notes – number of observations = 147.

Table 3.5 shows all observed patterns of ASF advancement for the overall Apache population. Of the 147 respondents (cross-sections) in our sample, a majority (119) reported wages for at least one time period prior to their involvement with the Apache project. Indeed, the most commonly occurring pattern (43 occurrences) is one where respondents begin their Apache careers in the fourth observation period. Further, it is quite common for an open source contributor to "plateau" at the rank of developer. Thus, the majority of respondents make very infrequent contributions to the project perhaps in response to issues that directly affect their work. The remaining roughly 27% of our respondents display contribution patterns demonstrating varying degrees of promotion within the meritocracy. Of these remaining respondents, 21 plateau at the rank of committer. The remaining 19 attain the rank of project committee member or higher. A similar distribution of advancement patterns can be observed from the 1,301 contributors in the overall Apache population.

Table 3.5 – Patterns of Rank Progression in Population

Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Freq	%
NORNK	NORNK	NORNK	NORNK	DEV,COM	COM	1	0.08
NORNK	NORNK	NORNK	DEV	DEV	DEV	342	26.27
NORNK	NORNK	NORNK	DEV	DEV	COM	8	0.61
NORNK	NORNK	NORNK	DEV	DEV	COM,PMC+	3	0.23
NORNK	NORNK	NORNK	DEV	COM	COM	2	0.15
NORNK	NORNK	NORNK	DEV	COM	PMC+	2	0.15
NORNK	NORNK	NORNK	DEV	COM,PMC+	PMC+	2	0.15
NORNK	NORNK	NORNK	DEV,COM	COM	COM	39	3
NORNK	NORNK	NORNK	DEV,COM	COM	PMC+	12	0.92
NORNK	NORNK	NORNK	DEV,COM,PMC+	PMC+	PMC+	2	0.15
NORNK	NORNK	DEV	DEV	DEV	DEV	330	25.35
NORNK	NORNK	DEV	DEV	DEV	COM	3	0.23
NORNK	NORNK	DEV	DEV	DEV	COM,PMC+	2	0.15
NORNK	NORNK	DEV	DEV	COM	COM	1	0.08
NORNK	NORNK	DEV	COM	COM	COM	17	1.31
NORNK	NORNK	DEV	COM	COM	PMC+	5	0.38
NORNK	NORNK	DEV	COM,PMC+	PMC+	PMC+	1	0.08
NORNK	NORNK	DEV,COM	COM	COM	COM	65	4.99
NORNK	NORNK	DEV,COM	COM	COM	PMC+	5	0.38
NORNK	NORNK	DEV,COM	COM	PMC+	PMC+	1	0.08
NORNK	NORNK	DEV,COM	PMC+	PMC+	PMC+	7	0.54
NORNK	NORNK	DEV,COM,PMC+	PMC+	PMC+	PMC+	8	0.61
NORNK	DEV	DEV	DEV	DEV	DEV	108	8.29
NORNK	DEV	DEV	COM	COM	COM	3	0.23
NORNK	DEV	DEV	COM	COM	PMC+	1	0.08
NORNK	DEV	DEV	COM,PMC+	PMC+	PMC+	3	0.23
NORNK	DEV	COM	COM	COM	COM	6	0.46
NORNK	DEV	COM	COM	PMC+	PMC+	1	0.08
NORNK	DEV	COM,PMC+	PMC+	PMC+	PMC+	3	0.23
NORNK	DEV,COM	COM	COM	COM	COM	21	1.61
NORNK	DEV,COM	COM	COM	COM	PMC+	3	0.23
NORNK	DEV,COM	COM	PMC+	PMC+	PMC+	1	0.08
NORNK	DEV,COM	PMC+	PMC+	PMC+	PMC+	8	0.61
NORNK	DEV,COM,PMC+	PMC+	PMC+	PMC+	PMC+	5	0.38
DEV	DEV	DEV	DEV	DEV	DEV	238	18.28
DEV	DEV	DEV	DEV	DEV	COM	1	0.08
DEV	DEV	DEV	DEV	DEV	COM,PMC+	1	0.08

DEV	DEV	COM,PMC+	PMC+	PMC+	PMC+	2	0.15
DEV	COM	COM	COM	COM	COM	1	0.08
DEV	COM	PMC+	PMC+	PMC+	PMC+	1	0.08
DEV	COM,PMC+	PMC+	PMC+	PMC+	PMC+	5	0.38
COM	COM	COM	COM	COM	COM	3	0.23
COM	COM	COM	COM	COM	PMC+	1	0.08
COM	PMC+	PMC+	PMC+	PMC+	PMC+	4	0.31
PMC+	PMC+	PMC+	PMC+	PMC+	PMC+	24	1.84

Table 3.5 Notes – Number of observations = 1301.

We compared the distribution of rank patterns over the period covered by our panel for our respondents and the population of Apache participants. We tested the equivalence of the location and scale of rank patterns across these two groups using the Mann-Whitney U test. In addition, we tested whether the distribution of rank patterns is the same across these groups using the Kolmogorov-Smirnov empirical distribution function statistic. In both cases, the results indicate that the rank patterns of our respondents and non-respondents are drawn from the same underlying population. (Mann-Whitney statistic=-.92, p= .27; Kolmogorov-Smirnov statistic=.74, p=.65)

Lastly, we examined whether there is selectivity bias in our sample. Inference based on balanced or unbalanced panel data may be subject to bias if the non-response within the panel is endogenously determined (Heckman 1979). We checked for selectivity bias using the variable addition test outlined in Verbeek and Nijman (1992). This test involves the introduction of three additional variables to the model, none of which should enter the model with significant coefficients under the null hypothesis of no selectivity bias. The first variable is the number of observation waves in which respondent i participates. The second is a dummy variable set to 1 when the respondent i has observations in all waves. The third is a time varying dummy variable indicating whether the respondent i is observed in the previous period. The results for these tests showed no indication of selectivity bias in our sample.

Empirical Methodology and Results

By combining the contribution data extracted from project archives with the job history data obtained from our surveys into a single panel, we can explore the relationship between open source participation and the change in wages over time. Our approach is to employ econometric models that take advantage of our repeated measures data to control for time-invariant participant endowments.

Model and Measures

Because we are interested in the nature of the relationship between open source participation and the market for IT labor, the human capital model provides a natural structure for assessing the returns to open source software participation. Accordingly, we formulate essentially Mincerian wage models commonly used to test the impact of education on log-earnings (Mincer 1974). The basic models specify the dependent variable as log wages, and the independent variables include observed demographic characteristics such as schooling, experience, and other variables of interest to the researcher (Weiss 1995). The general model is of the form:

(1) $\text{LWAGE}_i = \beta_0 + \beta_1 S_i + \beta_2 \text{EXP}_i + \beta_3 \text{EXPSQ}_i + \varepsilon_i.$

where LWAGEi is the natural logarithm of wages for individual i, si represents years of schooling, expi and expsqi represent labor market experience, and εi is the standard disturbance term. This general model has been extended to explore numerous other factors hypothesized to affect the wage relationship among individuals.[7] Following the literature in this regard, we extend the human capital model to include variables related to open source participation such as the participant's project contributions, Apache rank, job history, and other background information. In our setting, total wage is a function of accumulated Apache contributions, rank within the Apache Software Foundation, accumulated work experience, programming skills, education, firm size, firm type (publicly listed or private), firm industry, job type, and job switch. Columns one and two of Table 3.6 list our model variables and descriptions.

Table 3.6 – Model Variables and Descriptive Statistics

Time varying variables - Continuous	Variable	Grand Mean	Std. Dev.

[7] Such extensions include gender (Kay and Hagan 1995), race (Belman and Heywood 1991), union status (Jakubson 1991), use of technology (Benjamin et al. 2002), individual social network factors (Pfeffer and Konrad 1991), military service (Goldberg and Warner 1987), academic course work (Kang and Bishop 1986), GED attainment (Murnane et al. 1999), and volunteerism (Day and Devlin 1998).

Total wages in 1998 Dollars	TWAGE	83.76			32.48			
Average # hours worked per year	WKHRS	2107.44			449.51			
Average # hours contributed to Apache per year	APHRS	185.11			521.67			
Time varying lagged variables - Continuous	Variable	Grand Mean			Std. Dev.			
Career experience	EXPR	6.82			4.66			
Career experience squared	EXSQ	68.17			74.15			
Level of education in years	LEDU	16.27			2.15			
Project contributions	CNTRB	60.84			346.92			
Number of Observations - 450								
	VARIABLE	Participant Frequencies by Observation Period						
Time varying variables - Dichotomous		1	2	3	4	5	6	
Employed by publicly traded firm	FPUB	48 (36%)	42 (30%)	8 (19%)	13 (31%)	18 (40%)	19 (42%)	
Employed in the software industry	FSWIN	53 (40%)	64 (45%)	23 (54%)	21 (50%)	16 (36%)	15 (33%)	
Student	STDNT	19 (14%)	13 (9%)	3 (7%)	3 (7%)	2 (4%)	2 (4%)	
Paid to develop Apache software	PDAPC	13 (10%)	20 (14%)	6 (14%)	4 (10%)	4 (9%)	5 (11%)	

Time varying lagged variables - Dichotomous	VARIABLE	Participant Frequencies by Observation Period					
		1	2	3	4	5	6
Observed job switch	JSWCH	48 (36%)	43 (30%)	10 (23%)	9 (21%)	6 (13%)	11 (24%)
No rank (No contributions to project)	NORANK	108 (81%)	96 (68%)	13 (30%)	0 (0%)	0 (0%)	0 (0%)
Apache rank equal to 'developer'	DEV	24 (18%)	36 (25%)	24 (56%)	32 (76%)	34 (76%)	34 (76%)
Apache rank equal to 'committer'	COM	0 (0%)	6 (4%)	5 (12%)	8 (19%)	5 (11%)	3 (7%)
Apache rank of project management committee member or higher	PMC+	1 (1%)	4 (3%)	1 (2%)	2 (5%)	6 (13%)	8 (18%)

Other variables	VARIABLE	Participant Frequencies by Observation Period					
		1	2	3	4	5	6
No college degree	NOCOL	40 (30%)	33 (23%)	8 (18%)	7 (17%)	10 (22%)	9 (20%)
College degree	COLL	56 (42%)	71 (50%)	19 (44%)	19 (45%)	21 (47%)	22 (48%)
Masters degree	MAST	25 (19%)	25 (18%)	13 (31%)	14 (33%)	11 (24%)	11 (25%)
PhD	PROF	9 (9%)	13 (9%)	3 (7%)	2 (5%)	3 (7%)	3 (7%)
Technical job such as software engineer, system administrator, etc	TECH	113 (85%)	111 (78%)	40 (93%)	38 (90%)	42 (93%)	41 (91%)
Management job such as project manager or CIO	MGT	20 (15%)	31 (22%)	3 (7%)	4 (10%)	3 (7%)	4 (9%)

The dependent variable, LWAGE, is the natural logarithm of the sum of each participant's annual wages and bonuses.[8] To account for inflation, each year's wages are expressed in constant 1998 U.S. dollars.

CNTRB is a measure of each participant's open source experience in terms of project contributions. In the human capital model, career experience is commonly measured as the length of time, typically in years, spent participating in one's vocation. In contrast, open source participation is a voluntary activity, and can be transitory or sporadic. If the relationship between open source participation and wages has a

---

[8] Delayed returns to open source participation may include other benefits, the most important of which are stock options. However, we chose not to ask for detailed information about stock options for the following reasons. First, open source participants working at start-ups told us that they were not allowed to disclose the number of stock options that they were holding, making it impossible to estimate the value of an option. Further, a significant percentage of respondents work at privately held firms, rendering it difficult to assess the value of these options.

"job training" or human capital explanation, it is important that we find a reliable proxy for the experience garnered through open source participation. Learning by doing is increasingly recognized as an important factor influencing the relationship between experience and productivity or success in software development (Orlikowski 2002). With experience, developers gain familiarity with the software application domain and increase their understanding of the structure and architecture of the modules, files, and code within the system (Banker et al. 1998). The experience in software development (the "doing") largely consists of authoring the software; that is writing lines of software code using a particular programming language. The number of lines of software code written or changed is a commonly used productivity metric in software development organizations (Boehm et al. 2000). Thus, we operationalize open source experience (CNTRB) as a participant's cumulative number of lines of code contributed and accepted by the Apache project.[9] If CNTRB is a good proxy for the learning experience of an open source developer, we expect CNTRB to be positively correlated with LWAGE.

The dichotomous variables NORANK, DEV, COM, and PMC+ (collectively referred to as rank) operationalize the observed levels of contributor rank naturally occurring within the Apache meritocracy, that is, latent contributor, developer, committer, and project management committee member or above, respectively. Promotion within the meritocracy is awarded after a positive peer review of one's tangible and intangible contributions to the project. Rank may then, in part, reflect sought after (yet hard to observe) traits valued by information technology labor markets, such as the depth of developers' understanding, their efficient designs, or their ability to persuade or to get people "on board" with their ideas and strategies. If Apache rank is a signal of productive capacity in the open source environment, we would expect our rank variables to be positively correlated with LWAGE.

The variable PDAPC is a qualitative variable that assumes the value of 1 when a participant's paying job in time period *t* involves contributing to any one of the Apache projects. Increasingly,

---

[9] To account for possible differences in the LOC measure between programming languages, each contribution is converted to a common metric using industry standard language conversion factors (Boehm et al. 2000)

companies are sponsoring employees to participate in open source projects that are seen as an integral part of the company's information technology strategy (IBM 1998). Expectations regarding this variable are unclear. If the act of participation in an open source project is a desirable (i.e., valuable) outcome, then participants may be willing to forgo some amount of earnings for the opportunity to work on the project. In this case, PDAPC would be negative and significant. If, on the other hand, open source experience represents a specialized or rare skill, then employers could be expected to pay a premium for such skills. In this case, PDAPC would be positive and significant.

EXPR and LEDU are the traditional human capital variables. EXPR is the total number of years of work experience of a contributor at time *t*-1. Consistent with the human capital literature, we expect wages to increase with work experience, but the percentage increase to decline with higher work experience. Thus we expect EXPR to be positively correlated with wages, and EXSQ to be negatively correlated with wages. LEDU is represents the number of years of schooling for a participant at time t-1. Education is typically represented as time invariant in studies of human capital accumulation; however, the presence of students in our sample makes it possible to infer accurate levels of schooling within subject by tracking a respondent's declaration of full-time student status within each observation period. Returns to schooling are expected to be positive.

STDNT, FPUB, FSWIN and MGT are qualitative variables that assume the value of 1 if the participant is, respectively, a student, works for a publicly traded firm, works for a firm operating in the software or e-commerce industry, or is a member of management in period *t* and is 0 otherwise. Students are frequent contributors to open source software projects (Lakhami et al. 2002) and, ceteris paribus, we expect students to earn low wages. As a result, we expect STDNT to be negatively correlated with LWAGE. Both firm-level characteristics, such as firm size and sector, and job-type (technical vs. non-technical) have been shown to significantly affect the earnings of employees in the software industry (Ang et al. 2002). Following the prior research, we expect that participants working in publicly traded firms and firms engaged in the production of software to have higher wages, all else equal. Likewise, we expect that employees fulfilling a managerial role in firms to have higher wages than those in technical roles.

As shown in Table 3.6, a majority of respondents (on average 85%) classify their occupation as technical in nature such as "Software Engineer/Developer" and are employed by firms working in the software or ecommerce industries. Overall, data regarding participants' wages and experience are comparable to published accounts of salary and experience for software developers (Computerworld 2002) with participants reporting an average annual salary of $80,000 and six years of industry experience. As is common in technology industries (Watson 2000), we observe frequent job switching behavior over the course of our panel with the high of 36% of respondents reporting that they switched jobs in 1998. Respondents are well educated with roughly 50% holding college degrees, 25% holding masters degrees, and 8% of respondents holding Ph.D.s. The number of developers at the various levels of rank remains fairly constant over the period of our panel. As expected, the largest number of participants can be found at the rank of developer, the entry level or base of the ASF meritocracy, while the higher status levels of rank (COM and PMC+) contain fewer members.

Estimation and Results

A common concern in the human capital literature is the potential correlation of some unobserved person-specific variable, say u_i, with one or more of the regressors. If we assume that u_i contains some time-invariant heritable characteristic, such as intelligence, and to the extent that u_i is correlated with one of the other regressors, both OLS and GLS will yield biased and inconsistent parameter estimates (Chamberlain 1984). In the present case, our concerns are focused on accounting for unobserved skill or quality differences across contributors. For example, such differences might include inherent programming and design capabilities, the ability to succinctly explain complex technical issues, or the ability to self-motivate and work in an unstructured, often chaotic, environment. Measures of work or programming experience may not adequately reflect such skills. Ideally, one would like to directly control for such individual effects, and indeed this is a goal of many empirical studies of human capital (Taubman and Wales 1973). If, however, we assume that such abilities are rooted in the individual and thus constant over time, then a fixed-effect (FE) model solves the omitted variables problem. By differencing away time-invariant variables, whether observed or unobserved, the FE model produces consistent parameter estimates, purged of heritable individual effects. Hence, we make use of our cross-sectional time-series data to fit a FE regression model to explore the relationship between open source participation and wages over time. We estimate the returns to open source participation using the following equation:

$$(2) \quad \text{LWAGE}_{i,t} = \alpha i + \beta 1\text{CNTRB}_{i,t}\text{-}1 + \beta 2\text{DEV}_{i,t-1} + \beta 3\text{COM}_{i,t}\text{-}1 + \beta 4\text{PMC+}_{i,t-1} + \beta 5\text{EXPR}_{i,t-1} + \beta 6\text{EXSQ}_{i,t-1} + \beta 7\text{LEDU}_{i,t-1} + \beta 8\text{JSWCH}_{i,t} + \beta 9\text{FPUB}_{i,t} + \beta 10\text{FSWIN}_{i,t} + \beta 11\text{STDNT}_{i,t} + \beta 12\text{PDAPC}_{i,t} + \beta 13(\text{MGT})_{i,t} + \varepsilon_{i,t} \qquad (i = 1..N;\ t = 1..T);$$

Where i represents cross-section i observed at time t. The individual effect α_i is assumed to be an estimable cross-section specific constant term.

Equation (2) is essentially a two-way specification along dimensions of rank and time. The results of estimating equation (2) using the FE estimator are presented in Table 3.7. Column 1 shows the parameter estimates, standard errors and significance levels for the FE estimates. To test the applicability of the FE estimator, we conduct an omnibus test of the null hypothesis that all α_i are equal to zero. This check is easily rejected (F=8.44, p<.001) indicating the appropriateness of the FE estimator for our data.

Table 3.7 – Regression Results: Fixed Effect, Random Effect, 2SLS, and Random Effect with AR(1)

Coefficients	(1) OLS-FE	(2) GLS-RE	(3) ML-RE	(4) AR(1)	(5) 2SLS-IV
PDAPC	.0484	.0458	.0458	.0567	.0461
	(.0425)	(.0394)	(.0389)	(.0408)	(.0391)
JSWCH	.0251	.0326	.0327	.0203	.0321
	(.0264)	(.0251)	(.0248)	(.0253)	(.0249)
FPUB	.0340	.0499*	.0501*	.0598**	.0490*
	(.0308)	(.0280)	(.0276)	(.0286)	(.0278)
FSWIN	.0119	.0336	.0339	.0400	.0322
	(.0329)	(.0291)	(.0288)	(.0299)	(.0290)
LEDU	.0540*	.0165	.0164	.0142	.0169
	(.0306)	(.0124)	(.0121)	(.0115)	(.0126)
STDNT	-0.2197***	-0.3160***	-0.3166***	-0.3166***	-0.3127***
	(.0842)	(.0637)	(.0629)	(.0625)	(.0637)
CNTRB	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
EXPR	.0781***	.0725***	.0725***	.0728***	.0729***
	(.0171)	(.0127)	(.0126)	(.0129)	(.0128)
EXSQ	-0.0024***	-0.0023***	-0.0023***	-0.0023***	-0.0023***
	(.0009)	(.0007)	(.0007)	(.0008)	(.0007)
MGT	.1114**	.0809**	.0806**	.0711*	.0824**
	(.0463)	(.0413)	(.0407)	(.0420)	(.0410)
DEV	.0051	.0115	.0114	-0.0057	.0115
	(.0375)	(.0305)	(.0301)	(.0316)	(.0307)
COM	.2163***	.2352***	.2354***	.2242***	.2333***
	(.0610)	(.0552)	(.0545)	(.0591)	(.0560)
PMC+	.2292***	.2524***	.2522***	.2529***	.2484***
	(.0830)	(.0718)	(.0707)	(.0780)	(.0871)
INTCPT	3.0576***	3.6884***	3.6898***	3.7065***	3.6802***
	(.4985)	(.2014)	(.1979)	(.1856)	(.2057)
Overall R2	.3414	.3780	N/A	.3786	.3772

Table 3.7 Notes – Dependent variable is total wages in 1998 US Dollars. Standard errors are in parentheses. *** signifies p-value < .01, ** signifies p-value < .05, * signifies p-value < .10. Number of observations = 450.

In evaluating our results, we first examine the impact of those variables in our model that are unique to our open source software setting. The coefficient for CNTRB ($\beta 1$=-.0002, p<.001) while significant, is for all practical purposes, zero, with a value of less than .001. Alternative specifications of CNTRB, such as the percent of overall contributions submitted, the median absolute deviation from the median, the number of standard deviations from the mean, and the amount of time spent working on Apache projects also failed to yield a substantive change in the coefficient. Our interpretation of this result is that open source project experience, expressed as cumulative contributions is not, per se, associated with an increase in wages. Given the size and complexity of ASF projects and the use of relatively unsophisticated software development tools and methods, it is not hard to imagine that employers would find it difficult to judge, first hand, the merit of an open source job candidate's contributions. In light of the above discussion, we find no support for H1 that open source participation is positively related to wages. The relationship between our measures of Apache rank or status within the project, however, tell a different story.

The coefficient of the rank variable, DEV, is not significant ($\beta 2$=.005, p=.89). Recall from our examination of patterns of rank progression that the careers of Apache contributors progress in different ways and at different rates. Indeed, for 3 out of 6 observation periods, the wages earned by latent contributors (i.e., those having no rank) are not statistically different from that of those who achieve rank DEV – the first rung of the Apache career ladder. While the Apache meritocracy, and consequently our conceptualization of rank, is certainly ordinal, it is not interval. As a practical matter, the requirements to move from NORANK to DEV are quite low in our formulation requiring only a single contribution, regardless of the size or significance of the contribution. In other words, the minimum threshold for attaining rank DEV, a single contribution, is not significantly different from no contributions at all and hence may not provide any additional insights into the productive capacity of the respondents at that rank.

Turning to our remaining measures of rank, we find that the coefficients for COM and PMC+ are positive and significant (β_3=.216, p<.001; and β_4=.229, p<.001, respectively). We can calculate the percentage change in LWAGE associated with respondents having rank COM as $100\cdot(e^{(.216)} - 1) = 24.14\%$.[10] Similarly, the percentage change in LWAGE associated with respondents having rank PMC+ is 25.76%. That is, after controlling for open source experience, education, work experience, job switch, job type, firm characteristics, and latent individual effects via the FE estimators, respondents having an Apache

[10] In a semilogarithmic regression equation, the percentage impact of a dichotomous variable coefficient, c, on the dependent variable is properly expressed as $100\cdot(e^c\text{-}1)$ (Halvorsen and Palmquist 1980).

rank of COM enjoy wages that are, on average, 24% higher that those having no rank at all. Likewise, respondents having an Apache rank of PMC+ enjoy wages that are, on average, 26% higher than respondents having no rank.

It would seem however, that there are some limitations to the extent that Apache rank can function as a signal mechanism – at least at a fine grain level. An ideal signal in a meritocracy would allow for fine-grained discrimination between increasing levels of the credential. An underlying assumption here is that advancement in Apache meritocracy is similar to promotion in a traditional employee/employer relationship. That is, within a given firm, increasing levels of responsibility are traditionally associated with increasing levels of remuneration. Similarly, rising through the meritocracy should convey information regarding one's increasing level of competence or other difficult to observe, yet desirable, characteristics that may be associated with more responsibility. We noted previously that rank DEV is indistinguishable from NORANK and is an unavoidable vestige of our operationalization. While the coefficients on COM and PMC+ are significantly different from NORANK ($p<.001$ and $p=.01$ respectively), they can not be distinguished from each other ($p=.88$). That is, COM and PMC+ appear to provide the same level of signal information. One explanation for this result may be found by exploring the relationship between the nature of a contributor's paid-work and their position in the meritocracy. To explore the relationship between rank and job-title we extended our model to include the interaction of the dichotomous variable MGT (1 if job title is managerial, 0 otherwise) and our 3 levels of rank (DEV×MGT, COM×MGT and PMC×MGT). If the characteristics that associated with achieving PMC+ are significant in the performance of a managerial role then we would expect PMC×MGT to be positive and significant. In fact, this is what we find. All of the coefficients in the interaction model remain ostensibly unchanged from the previous functional form. While interaction terms DEV×MGT and COM×MGT are not significant, PMC×MGT is positive and marginally significant ($\beta=.266$, $p=.09$). Thus, we do find that for rank PMC+ contributors who occupy management positions earn higher wages than those who fill technical roles. Overall, we interpret the above results in favor of H2 – that a credential signifying successful open source participation is positively related to wages.

It is of interest to note that Apache participants can be quite generous with the number of contributions they make to the Apache projects. As shown in Table 3.6, for each contribution occurrence, the average CNTRB value across all ranks is 61. More interestingly, we find that CNTRB increases significantly with each increase in rank; with highly ranked participants contributing an annual average CNTRB value of nearly 200 lines of code. Specifically, respondents at rank PMC+ contributed significantly more contributions (CNTRB) than respondents at rank DEV or COM. Recall from the previous discussion that moving from rank DEV to higher levels of rank is associated with significant increases in wages while measures of contribution hold little explanatory power for observed wages exhibiting significant, but negligible, coefficient values. One plausible interpretation of this finding is that teasing apart the relationship between contributions and success in an open source project is a difficult task. Credentials or proxies of success, on the other hand, appear to effectively convey information relating open source participation with desirable software engineering skills. Taken together, these results suggest that employers do not appear to reward participants for their learning experience in the open source projects operationalized as CNTRB. However, the significantly higher wages paid to contributors with higher rank is consistent with the notion that the rank conveys sought-after, but typically hard-to-observe, characteristics that may distinguish above average programmers.

The last variable in our model that is unique to our open source software setting is PDAPC. While respondents who were paid to develop Apache software were observed in every observation period, there is no significant relationship in our sample between being a paid Apache developer and LWAGE. Recall that there are conflicting expectations regarding the relationship between PDAPC and LWAGE. In our sample, it appears that employers pay neither a premium nor a discount on wages to employees who are compensated for their Apache participation.

We now turn our attention to those variables in our model that are commonly included in models of human capital. The coefficients of the control variables for experience are consistent in both sign and magnitude with the existing literature (cf. Murnane et al. 1999). Each year of EXPR significantly increases

LWAGE by 7.8% ($p<.001$), but with increasing work experience, increases in LWAGE are growing more slowly (as EXSQ is negative and significant; $p=.01$). We also find that level of education, LEDU, exhibits a marginally significant relationship with wages our sample. Each additional year of education is associated with a 5.55% increase in wages ($p=.08$). Surprisingly, JSWCH is not significant given both the well-established positive relationship between job switching and wage increases (Schafer 2003) and the amount of job switching behavior observed in our sample. It should be noted, however, that our data collection took place during a sharp decline in the information technology job market following the bursting of the "dot com" bubble. Given the decreases in overall information technology employment and the concomitant decreases in information technology hiring during this period, lateral job moves or even moves involving a reduction in wages could logically be expected.

We find no significant evidence that firm level factors are associated with higher LWAGE. Wages of respondents working for publicly held firms and those working for private firms are not significantly different. Likewise, average wages for respondents working for firms engaged in the production of software and those who do not are not significantly different. Lastly, consistent with expectations, we find that being a student is significantly and negatively associated with LWAGE as students earn 24.6% less than non-students ($p=.01$).

Alternative Estimations and Considerations

As Wooldridge (2002) observes, a primary motivation for using panel data is to solve the problem of omitted variables that may be correlated with other regressors in the model. We have made use of the FE estimator to address potential issues related to unobserved individual quality traits that may be correlated with other regressors of interest such as rank or CNTRB.[11] However, the FE specification treats differences between respondents as parametric shifts of the regression function and thus limits the applicability of FE estimates to out-of-sample prediction. In contrast, the random effects (RE)

---

[11] In the absence of such omitted variables, FE is consistent but not fully efficient (Hausman and Taylor 1981).

specification views individual-specific constant terms as randomly distributed across cross-sectional units and is appropriate if the sampled cross-sectional units are drawn from a large population.

While we can not reject the FE model using the omnibus test for the significance of all α_i, neither can we reject the RE model based on standard tests for random effects. To test the appropriateness of the RE model, we conducted a Hausman specification test for random effects (Hausman 1978). Under the null hypothesis of no correlation between the latent individual effect (u_i) and the other regressors, failure to reject the null implies that the RE estimates are consistent and efficient, and that the FE estimates are consistent but inefficient. The Hausman specification test for our RE model results in $\chi 2$ (d.f.=13) m-statistic of 10.95 (p=0.62). Therefore we fail to reject the null hypothesis of no correlation between u_i and the other regressors and thus accept the appropriateness of the RE estimator for our data.

We investigated a RE specification of our model, estimating our wage equation using a maximum likelihood RE estimator. The RE results are shown in column 2 of Table 3.7 and can be compared with those from our FE estimation in column 1. If the RE and FE estimates differ significantly, we may conclude the presence of a significant latent individual effect. If, on the hand, the RE and FE estimates do not differ significantly, then we are justified in rejecting claims that such latent effects exist. Note that while the FE estimates are manifestly less efficient than the RE estimates, there is still significant agreement between the two methods on the sign, magnitude, and significance of each coefficient in our model. The uniformity of the RE and FE estimates is even more pronounced among our measures of open source participation with essentially no difference between the RE and FE estimates for CNTRB and rank.

One potential threat to our results is that Apache rank may be endogenously determined. That is, if open source participants observe that rank is associated with higher wages, they may increase their "investment" in the project in order to attain higher rank and hence higher wages. First, we note that by linking our dependent variable to prior (lagged) values of rank we have diminished the possibility that rank is endogenously determined (Greene 2003). Even so, we investigate the possibility that rank is endogenously determined using an instrumental variables (IV) estimation of our model. Investigations into correlations between the FE residuals and candidate instruments reveal that median CNTRB within rank is highly and significantly correlated with the corresponding level of rank. At the same time, this variable is not significantly correlated with FE residuals, making it a suitable instrument for levels of rank. Column 5 of Table 3.7 shows the FE estimates from a 2SLS-IV regression using the instrument just described. As expected, the IV coefficients are less precisely estimated; however, coefficient estimates remain very close to original FE regression results. A Hausman test comparing the two models fails to reject the null hypothesis that the difference in coefficients is not systematic and thus provides some support for the exogeneity of rank.[12]

In both the FE and RE specifications of our model, the disturbance term (ε_{it}) are assumed to be independently distributed across individuals with no restrictions placed on the form of the within-subject autocorrelations. As observed by Arellano (1987), this formulation allows for heteroskedasticity and serial correlation of an arbitrary form. Indeed, a common challenge in using repeated measures on individual units to elucidate economic relationships is the presence of serial correlation in the error terms (Bhargava et al. 1982).

Test statistics for a first-order autoregressive process in unbalanced panels possess complex distributional properties (Baltagi and Wu 1999). As a result, establishing critical values is computationally obtuse. As is often the case, however, examination of the test statistic itself is revealing. With a value of 2.10, the Baltagi-Wu LBI[13] test statistic for our panel would almost certainly reject the

[12] Davidson and MacKinnon (1993) note that the Hausman test is not properly interpreted as a direct test of exogeneity. Wooldridge (2002, pg 285) suggests a test for strict exogeneity using the FE regression model. Following Wooldridge, we estimate Equation 2 with the addition of the leading values of the rank variables suspected of violating the assumption of strict exogeneity. The Wooldridge test fails to reject the null hypothesis that all such leads of rank are zero (p=0.731) providing additional evidence that rank is indeed exogenous.

[13] LBI is the locally best invariant test statistic for AR[1] processes in unbalanced panel data (Baltagi and Wu 1999).

null hypothesis of no serial correlation.[14] Additionally, examination of the unstructured within-subject residual correlation matrix reveals that the serial correlation of residuals decreases as the time lag increases, consistent with an AR(1) process. Indeed, when we compute the likelihood ratio comparing a fully unstructured within-subject residual covariance model to an AR(1) model, we reject the null hypothesis of the superiority of the full model at a 1% significance level (likelihood ratio = 20.3 ~ $\chi 2$ (d.f.=7) in favor of the more parsimonious autoregressive model.[15] Accordingly, we re-estimate our RE model specifying an AR(1) within-subject correlation structure. Results are shown in column 4 of Table 3.7. Again, considering our variables of primary interest there is little or no impact regarding the relationship between CNTRB, rank and LWAGE. As before, the parameter estimate for CNTRB is relatively unchanged exhibiting negligible impact on LWAGE. Likewise, the parameter estimates and standard errors for our rank variables show slight changes, however both COM and PMC+ retain significance at a less than 1% level.

Our primary interest is the relationship between open source participation and wages. To this point our regressions have included observations for participants from time periods prior to the start of their Apache career (i.e., for latent contributors). As a check of the robustness of our results, we constructed a new panel comprised solely of observations where participants have acquired at least a rank of DEV. Applying this constraint yields a cross-sectional time-series panel of 340 observations for the years 1999 through 2004. Table 3.8 shows the FE and RE parameter estimates for this new panel. Overall, the results are nearly identical to previous estimates. Although slightly less precisely estimated, the coefficients on our rank variables remain essentially unchanged.

---

[14] Critical values for both the LBI and the Bhargava et al. modified Durbin-Watson statistics for panels similar to ours are quite small and often negative (cf. Baltagi and Wu 1999).

[15] This finding is also confirmed by comparing both the Akaike's Information Criterion (AIC) and the Schwarz's Bayesian Criterion (BIC) of the two models (Raftery 1995).

Table 3.8 – Regression Results: Fixed Effect and Random Effect Model Excluding Latent Contributors

Coefficients	OLS-FE	GLS-RE	ML-RE	AR(1)	2SLS-IV
pdapc	.0450	.0442	.0442	.0549	.0444
	(.0484)	(.0450)	(.0443)	(.0465)	(.0448)
jswch	.0195	.0237	.0242	.0177	.0236
	(.0313)	(.0302)	(.0298)	(.0308)	(.0300)
fpub	.0454	.0550*	.0555*	.0612*	.0546*
	(.0342)	(.0323)	(.0318)	(.0334)	(.0322)
fswin	.0167	.0382	.0398	.0465	.0375
	(.0374)	(.0342)	(.0339)	(.0355)	(.0342)
ledu	.0495	.0127	.0123	.0092	.0132
	(.0355)	(.0152)	(.0146)	(.0138)	(.0154)
stdnt	-0.2563***	-0.3340***	-0.3359***	-0.3388***	-0.3319***
	(.1011)	(.0803)	(.0789)	(.0804)	(.0803)
cntrb	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
expr	.0703***	.0625***	.0620***	.0629***	.0631***
	(.0188)	(.0157)	(.0154)	(.0159)	(.0157)
exsq	-0.0021**	-0.0019**	-0.0018**	-0.0018**	-0.0019**
	(.0009)	(.0008)	(.0008)	(.0009)	(.0008)
mgt	.1202**	.1065**	.1060**	.1030**	.1070**
	(.0518)	(.0478)	(.0471)	(.0489)	(.0477)
dev	.0148	.0246	.0244	.0123	.0236
	(.0405)	(.0344)	(.0338)	(.0351)	(.0348)
com	.2270***	.2478***	.2488***	.2410***	.2447***
	(.0653)	(.0595)	(.0586)	(.0623)	(.0607)
pmc+	.2448***	.2674***	.2667***	.2644***	.2572***
	(.0896)	(.0772)	(.0759)	(.0815)	(.0940)
intcpt	3.1603***	3.7716***	3.7795***	3.8103***	3.7628***
	(.5805)	(.2485)	(.2391)	(.2249)	(.2524)
Overall R2	.2891	.3203	N/A	.3220	.3198

Table 3.8 Notes – Dependent variable is total wages in 1998 US Dollars. Standard errors are in parentheses. *** signifies p-value < .01, ** signifies p-value < .05, * signifies p-value < .10. Number of observations = 340.

Conclusion

The research presented here seeks to explore one of the more puzzling aspects of the open source phenomenon – why do developers participate? Specifically, we explore whether participation is consistent with well-established theories from labor economics. From this literature, we establish two plausible theoretical bases for the existence of returns to open source participation; viz., human capital and signaling theories. A human capital explanation of participation suggests that open source experience serves a "job training" function. Reasoning from a human capital perspective we hypothesized that open source participation is positively related to wages. In contrast, signaling theory suggests that successful open source participation serves a signaling or sorting function for IT labor markets. Reasoning from a signaling perspective, we hypothesized that successful open source project participation will be positively related to wages. Our analysis shows that employers do not reward the accumulation of experience in open source projects per se. Rather, successful open source participation, measured as higher open source rank, is associated with higher wages, even after controlling for work and programming experience. This finding is robust across several measures of contribution (CONTRB), sample selection excluding latent contributors, and model specifications and estimators.

That wages do not measurably increase with contributions to the Apache Project is consistent with the notion that employers find it difficult to assess the performance of programmers and hence changes in their human capital. It follows that employers would have even greater difficulty evaluating open source contributions in order to assess performance even though the source code is freely available. Even so, open source participation absent accompanying increases in rank may yet hold career advancement potential. Inasmuch as a contributor can apply his or her gained knowledge on the job, the programmer may be rewarded in the long run. Our findings suggest that in the case of the Apache Project, the open source community effectively screens programmers based on their productive capacity. Employers appear to recognize Apache's merit based ranking as a reliable proxy that is correlated with desirable, but imperfectly observably productive abilities. Our research contributes by providing empirical evidence on economic incentives for open source participation. Understanding the incentive structure is a critical first step in evaluating open source as a viable model for organizations seeking to exploit the obvious competitive advantages of "costless" open source development. As Strasser (2001) observes, questions about incentives are not merely academic: understanding how open source development actually works has profound implications for the strategies that corporations and governments should pursue over time. Providing insight into the economic incentive mechanisms underlying open source participation is a significant first step in this process.

CHAPTER 4.
STUDY 3: UNDERSTANDING THE MOTIVATIONS, PARTICIPATION AND PERFORMANCE OF OPEN SOURCE SOFTWARE DEVELOPERS: A LONGITUDINAL STUDY OF THE APACHE PROJECTS

Introduction

Open source software communities cannot exist or prosper without the contributions of highly motivated developers who are willing to donate their time and effort to the community. However, because these participants are often self-employed freelancers and volunteers, rather than traditional employees, it is not possible to solely rely on employment relationships or employment contracts to manage them. Thus, vital questions in open source software communities concern how to motivate participants and how to direct, sustain and influence their behaviors (Markus et al. 2000).

Several studies have revealed different motivations for contributing to open source projects (for a review see Rossi 2004). Often quoted motivations for participating in open source development projects cover a broad spectrum including scratching a "personal itch" with respect to software functionality, enjoyment, and a desire to be "part of a team" (Ghosh 1998). Others liken the open source community to a gift culture where the status of a participant depends on "what he gives away" (Raymond 2001a). Alternatively, Lerner and Tirole (2002) suggest that open source participation may in part be explained by existing theories of labor economics. Lastly, as commercial companies increase their involvement in open source projects, there are more developers being paid to contribute, adding the traditional incentive – pay – as a potential motivation to participate and raising the issue of how paid participation may affect other motivations to contribute.

Although there is no consensus in the open source literature as to which motivation is most dominant, individual contributors could likely have multiple salient reasons for participating. This raises an important and unanswered question that we address in this study: *how are the motivations of contributors related, i.e., are they independent, complementary or contradictory?* Answering this question is significant because an assumption in studies on open source participation is that motivations are complementary or "mutually reinforcing" (e.g., Markus et al. 2000). However, if, for example, some motivations are negatively related to others, increasing the level of those motivations may crowd out other motivations for participating. This issue is particularly relevant when considering how paid participation affects other motivations for open source contributions. In sum, understanding whether an open source participant's different motivations are in harmony or at odds matters because it is the participant's set of motivations, combined with knowledge, skills and abilities, that produces the participant's behaviors and performance (Mitchell & Daniels 2003).

In open source development, the different motivations to participate have been generally classified as either intrinsic or extrinsic (Rossi 2004). Intrinsic motivation occurs when an activity satisfies basic human needs for competence, control and autonomy. This makes the activity interesting and likely to be performed for its own sake rather than as a means to an end (Deci & Ryan 2000). In contrast, extrinsic motivation stems from the environment external to the task and is usually applied by someone other than the person being motivated (Johns 1996). Contributing to open source projects for the sheer enjoyment of coding is clearly an intrinsic motivation while being paid to contribute is the quintessential extrinsic motivation. Other motivations, such as contributing to solve a problem of personal use benefit (use value) or contributing to enhance status or career opportunities are, by definition, extrinsic, but, following Deci & Ryan (2000), contributors could *internalize* these motivations so that they are self-regulated rather than externally imposed. Following the literature in psychology (Deci & Ryan 1987), we classify these motivations as *internalized extrinsic motivations.* Distinguishing the different types of motivations (pure extrinsic, internalized extrinsic, and pure intrinsic) allows us to examine a second question: *how do differences in open source contributors' motivations relate to differences in their participation?* It is important to understand whether all types of motivations affect open source participation equally or in the same way. Studies of open source participants do not often consider whether different motivations differentially relate to participation. However, some motivations may strongly affect participation, while others may not be as salient. For leaders of open source projects who are trying to attract developers to participate in projects or to sustain their level of participation, it is imperative to understand which types of motivations are likely to generate more (or less) participation.

Our third question concerns the link between the level of participation and performance ranking. Specifically, *how do levels of participation relate to changes in performance rankings?* This question is salient for open source communities like Apache (the focus of our study). The Apache projects were not originally organized around a single person or primary contributor. As such, the success of the Apache projects depends on shared leadership and the contributions of participants. The projects are organized using a meritocracy, or as Roy Fielding explains it: "the more work you have done, the more you are allowed to do" (Fielding 1999, p. 43). For the meritocracy to be effective, promotions within the Apache community should be based on contributions to the Apache projects. In answering our third question, we relate a participant's promotion (or performance ranking) to the level of his or her prior contributions to the Apache software code. This provides an important validation of whether the meritocracy is functioning as intended.

Finally, while motivation is an antecedent of behavior and performance, research in psychology has also recognized the effect of performance feedback on motivation. For example, it has been shown that feedback considered controlling tends to decrease intrinsic motivation. On the other hand, research by Sansone (1986) suggests that feedback regarded as competence-enhancing can increase subsequent intrinsic and extrinsic motivation. One of the tenets of open source projects is the frequent provision of feedback to contributors (Moon & Sproull 2002). As we have noted, in some open source projects like Apache, continued contribution is rewarded with a change in performance ranking. Hence, the last question on our agenda: *how does a change in performance ranking affect the subsequent motivations of open source participants?* To the best of our knowledge, no studies of open source communities have considered how changes in performance rankings affect subsequent motivations for participating. However, the answer to this question has important implications for open source communities that wish to enhance or sustain the motivations of their participants over the longer term.

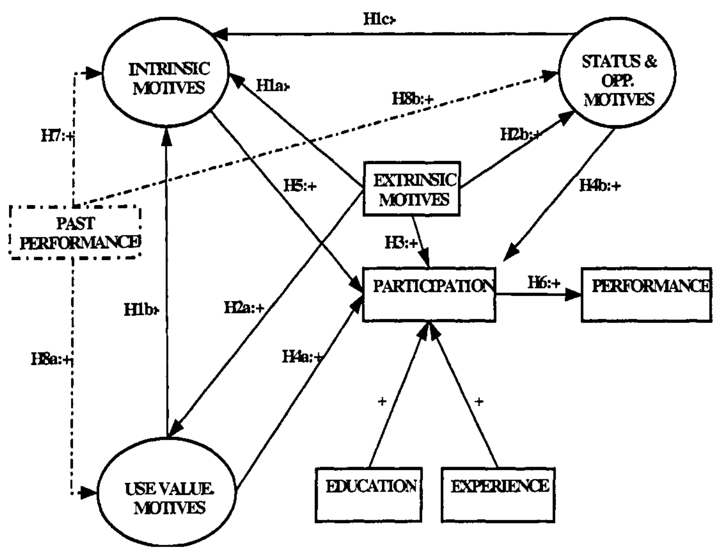
In §2 of the paper, we draw on the relevant literature in psychology to develop our theoretical model. §3 describes the empirical evaluation of our model involving a longitudinal field study of the motivations, contributions and performance of software developers in the Apache WEB server projects. §4 presents the analysis and results. We discuss our results in §5, and conclude in §6 by identifying the contributions and limitations of our study and its implications for research and practice in open source development.

Theoretical Framework

The theoretical framework for our study leverages the general model of motivation and performance in organizational and social psychology (Campbell & Pritchard 1976). In this framework, motivations vary across individuals and combine with individuals' knowledge, skills and abilities to produce task-relevant behaviors. These behaviors contribute to individual performance. It is important to distinguish motivation from behavior. Work motivation is the psychological force within an individual that determines the direction of the individual's behavior in an organization, the individual's level of effort, and the individual's level of persistence in the face of obstacles (Kanfer 1990). Motivation has an important influence on performance because it focuses attention on particular task elements and produces effort as people work harder when they are motivated. While motivation is a psychological state, the outcome or results of that state is behavior (Mitchell & Daniels 2003). Behavior also differs from performance because performance is an *evaluation* of the results of an individual's behavior usually by someone other than the individual – it involves determining how well or poorly an individual has accomplished a task (Kanfer 1990).

Applying the general model of motivation and performance to the open source context, we stipulate that motivations vary across open source contributors. Combined with developers' knowledge, skills, and abilities, motivations influence their participation in open source projects as exemplified by the level of their contributions to the source code. Over time, contributors' participation is evaluated by the open source community. This performance evaluation may lead to an increase in a contributor's rank within the community. An advance in rank can, in turn, act as feedback to influence the future motivations of contributors. Figure 4.3 illustrates our theoretical model.

Figure 4.3 – Theoretical Model



Intrinsic and Extrinsic Motivations in open source Development

open source development affords a particularly rich context in which to examine individual motivations. On the one hand, intrinsic motivations are likely to be important as contributors have a high degree of autonomy and self-determination and are valued for their competence. On the other hand, the open source community provides extrinsic motivations such as reputation or status. Indeed, empirical studies of open source contributors find that the participants report a variety of reasons for participating in projects (Hertel et al. 2003). Thus, our model includes an array of open source participation motivations ranging from strictly intrinsic to strictly extrinsic to those that have both intrinsic and extrinsic characteristics.

Research has investigated the relationship between psychological needs and intrinsic motivation. Specifically, intrinsic motivation has been linked to the satisfaction of human needs for autonomy and competence (Deci 1975). In the open source context, human needs for autonomy and competence are readily satisfied. As researchers have established, software development is an inherently motivating task as it is complicated and creative, yet difficult to observe and thus not easily monitored (Kirsch 1996; Weinberg 1998). Compared to software development in an organizational setting, contributing to open source projects allows individuals even greater opportunities to express their creativity, enjoy their work, and experience a sense of satisfaction and accomplishment (Lakhani & Wolf 2005). It is these characteristics that draw many competent, self-directed programmers into open source communities.

At the other extreme, extrinsic motivation has been linked to the operant conditioning literature (Skinner 1953) that advocates the use of incentives to reinforce desired behavior. Hence, extrinsic motivation is characterized by a strong focus on reward contingencies, which in Western-oriented economies is often represented by pecuniary compensation. While open source communities do not have a profit motive per se and hence do not offer monetary compensation to contributors, the pervasive commercial interest in many open source products has generated a thriving open source industry. It is not unusual for third parties to employ programmers for the specific purpose to contribute to open source projects. Examples include Time Warner's engagement in Mozilla and IBM's involvement in both the Linux and Apache open source projects.

Between intrinsic and extrinsic motivations there exists a continuum of motivations that are considered a mix of intrinsic and extrinsic as they are clearly not intrinsic at the outset, but undergo an internalization process that moves them away from strictly extrinsic motivations (Ryan & Connell 1989). Deci & Ryan (1987) refer to these motivations as *internalized extrinsic motivations.* According to Deci & Ryan (2000), internalization occurs when individuals assimilate and reconstitute external incentives or contingencies, transforming these external incentives into their own motives and leading to self-regulation. Self-regulation can take place in two ways: through introjection or through identification (Deci & Ryan 2000). Internalized extrinsic motivation based on *introjected regulation* is related to attainment of ego-enhancement and feelings of worth. In the open source context, an example of this motivation involves two related quests: for status and for career opportunities. Raymond (2001a) early on recognized status, or what he termed "ego-boo," as an important driver of participation in open source communities. Similarly, open source participation can be seen as a move to enhance career prospects. As described by Lerner & Tirole (2002), open source communities offer an excellent setting in which a participant motivated by career concerns can signal his or her abilities to the labor market. Likewise, von Hippel and von Krogh (2003) argue for the existence of private incentives for the provision of a public good such as open source. It is important to note that motivations based on either status seeking or career enhancement may both be internalized in an effort to increase one's standing in a reference group.

Internalized extrinsic motivation based on *identified regulation* is a more self-determined form of extrinsic motivation. In this form of regulation, individuals identify with an action and personally endorse it, leading to an identification that is accompanied by a higher degree of perceived autonomy. By definition, this type of motivation is extrinsic as it derives from the personal importance of the outcome, rather than the performance of the task (Ryan & Deci 2002). In open source communities, an example of this motivation is "use value," or the desire to fix a bug or solve a problem of immediate relevance to the contributor. Some studies have identified use value as a dominant motivation of open source contributors (e.g., Ghosh et al. 2002; Hertel et al. 2003; von Hippel 2001). While use-value is extrinsic because of the personal benefit to the user (Markus et al. 2000; Rossi 2004), from a psychological point of view, use value is internalized as a value of the open source community and transformed into a personally endorsed value.

In this study, we examine the relationships *between* different open source motivations. The literature in psychology on motivation has examined the relationships between intrinsic and extrinsic motivations (Lepper & Henderlong 2000). We consider how these relationships may apply in the open source context.

Experimental research in psychology (Deci 1971; Lepper et al. 1973) has shown that under certain conditions, extrinsic motivations displace intrinsic motivations. In cases where incentives are contingent upon performance, individuals expect to be rewarded, or incentives are tangible, external incentives undermine characteristics of intrinsic motivation such as free-choice behavior and self-reported interest. Multiple meta-analysis studies (Rummel & Feinberg 1988; Wiersma 1992; Deci et al. 1999) have found general support for this finding; Osterloh & Frey (2000) refer to this as the "crowding-out" effect. In general, previous research has found that the undermining effect of external incentives is especially powerful for monetary compensations that are perceived to be controlling. The effects are larger for monetary rather than symbolic incentives and for expected rather than unexpected incentives. The crowding out effect is also more observable for complicated rather than simple tasks (Deci et al. 1999; Lepper & Henderlong 2000).

In the context of open source projects, strictly extrinsic motivations apply directly to those participants who are remunerated for their activities. Drawing on the logic of the crowding out effect, we would expect that monetary incentives for involvement in open source projects would weaken intrinsic motives for participating. When participants are engaged in a traditional employment relationship with a firm, the employer has the right to establish the policies, rules and terms of employment, to structure the work environment, and to assign tasks to employees. This suggests that participants who are paid to contribute to open source projects are likely to have less autonomy in choosing which features they want to code and less freedom in how they do their work. If, as Weinberg (1998) argues, programming itself provides the strongest intrinsic motivation for software development, "*… if the programmer is given a chance to do it his way…*" (p.184), any restriction of participants' task autonomy should be negatively associated with their intrinsic motivations to participate in open source projects. Thus, we hypothesize that:

H1a: Contributors' intrinsic motivations to participate in open source projects are negatively related to being paid to participate.

The literature in psychology suggests that other extrinsic motivations, even those that are internalized (and thus not strictly extrinsic) could crowd out intrinsic motivations. A meta-analysis by Deci et al. (1999) examined the results of 128 laboratory studies of extrinsic and intrinsic motivation conducted over the past 25 years. The results from this meta-analysis show that most types of extrinsic motivations (and especially those specifying contingencies related to the task being performed) undermined intrinsic motivation. The authors found a positive effect of extrinsic motivations on intrinsic motivation only when the feedback was both positive and verbal. With respect to internalized extrinsic motivations that are based on the identification of values, Ryan & Deci (2002) observe that such identification is often compartmentalized and separated from one's other beliefs and values and is characterized by a reduced self-determination. Hence, a contributor may identify with the open source community through use value motivation only in a restricted, practical sense with limited self-determination and reduced pure intrinsic motivation. For example, a contributor who identifies strongly with an open source community may choose to work on tasks that are not inherently interesting because completing the tasks provides value to the community. In her review of the literature on open source motivations, Rossi (2004) states that use value motivations can provide a powerful explanation for why people do tasks that may be uninteresting and mundane and that are not "appreciated per se, for the intrinsic pleasure and enjoyment a programmer may derive from them" (p. 5). Thus, we expect:

H1b: Contributors' intrinsic motivations to participate in open source projects are negatively related to their use value motivations to participate.

Ryan (1982) finds that internalized extrinsic motivations based on introjection such as ego-enhancing motivations reduce intrinsic motivation for the target activity. Individuals may develop mastery for the purposes of gaining reputation, but find that motivation based on introjected regulation is quite controlling and less self-regulated. In the context of open source development, individuals who are highly motivated by status could find themselves working on tasks they may not necessarily enjoy but rather that are likely to enhance their reputation in the community. For example, writing open source software and helping to test and debug it are critical ways to earn respect in open source communities (Markus et al. 2000). However, a contributor who is motivated by reputation concerns may not particularly enjoy testing and debugging software, but may feel it necessary to do these activities to gain status in the open source community. This suggests that motivations grounded in status seeking motivation are likely to be negatively associated with the pure enjoyment of contributing. Thus, we posit that:

H1c: Contributors' intrinsic motivations to participate in open source projects are negatively related to their status motivations to participate.

Much of the research in psychology has focused on the effects of external incentives on intrinsic motivation. However, Ryan and Deci (2002) suggest that external incentives that promote feelings of self-determination can promote self-determined (*i.e.*, internalized) extrinsic motivations. Performance-contingent incentives (such as pay) can influence how individuals approach a task as well as their motivations during the performance period (Hennessey 2000) because such incentives can make doing well more personally important. That is, incentives that motivate individuals to strive for competence can amplify or enhance their other extrinsic motivations. Following this logic, we would expect that being paid to contribute to open source projects is complementary with other extrinsic motivations including use value and status. Thus, individuals who are being paid to contribute to open source projects could also have a high use value motivation because making contributions that improve the use value of the source code demonstrates competence for which they will be financially rewarded. Hence, we hypothesize:

H2a: Being paid to participate in open source projects is positively related to contributors' use value motivations to participate.

Similarly, individuals who are being paid to contribute to open source projects are likely to have higher status motivations because an interest in attaining status motivates individuals to demonstrate competence for which they will be financially rewarded. Therefore, we expect that:

H2b: Being paid to participate in open source projects is positively related to contributors' status motivations to participate.

Motivations and Participation in open source Development

According to the classic literature on operant conditioning (Skinner 1953), behavior that is rewarded with positive reinforcement is more likely to be repeated in the future. In Western societies, monetary compensation is the ultimate positive reinforcer to regulate economic activities in organizations. Hence, we would expect that contributors who are paid to participate in Apache projects would participate more intensely than those who are not paid. This is because their wages act as a constant positive reinforcer of their participation behavior. Thus,

> **H3: Being paid to participate in open source projects is positively related to contributors' level of participation.**

We would also expect that individuals with higher levels of use value motivations would exhibit higher levels of participation. In terms of use value motivations, one of the most often cited drivers of open source participation is the opportunity to create code that meets the specific needs of a developer (Raymond 2001a). Fixing a bug or solving a problem of immediate relevance to the programmer provides a powerful motivation to create the software code in the first place (Lerner & Tirole 2002). High levels of "use value" motivation therefore suggest a high level of participation. Thus,

> **H4a: Contributors' use value motivations to participate in open source projects are positively related to their level of participation.**

Status motivations should also be a strong driver of participation. Raymond (2001b) likened open source communities to gift cultures, where the sought-after status is determined by the programmer's contribution. In addition, human capital theory (Becker 1962) suggests that individuals, endowed with differing aptitudes and abilities, will strive to acquire additional knowledge and experience as long as the expected incentives are greater than the expected costs. In a slight variation, signaling theory (Spence 1976) presumes that individuals showcase their education and experience to signal imperfectly observable productivity characteristics to current and future employers. Human capital and signaling theories suggest that open source contributors who are motivated by status concerns will participate as a way of improving and signaling their programming abilities and competencies (Hann et al. 2002). Therefore, we hypothesize that:

H4b: Contributors' status motivations to participate in open source projects are positively related to their level of participation.

The literature in organizational psychology suggests that the "ideal" intrinsic motivation is "in the work content itself" (Calder & Staw 1975, p. 539). Tasks that are intrinsically motivating have a direct and strong association between the activity and the individual's purpose for performing the activity. Therefore, engaging in the task directly satisfies the individual's goals. In an experimental study, Shah and Kruglanski (2000) find that the strength of the activity-goal association is positively related to indices of intrinsic motivation, including the self-reported frequency of engagement in the activity and the importance one places in doing the activity. Thus, Shah and Kruglanski conclude that individuals who are intrinsically motivated to perform some activity will perform it very intensely. In the context of open source development, intrinsically motivated contributors should have higher levels of participation because they like to code, and by coding they are directly satisfying their desires.

Other psychologists have linked intrinsic motivation to task participation via its effect on creativity, because an intrinsically motivated orientation to task performance promotes characteristics that are essential for creativity (Amabile, et al. 1986). For example, individuals with high levels of intrinsic motivation focus more on the task, are more willing to take risks, and will explore alternative strategies for performing the task (Hennessey 2000; Osterloh & Frey 2000). Higher creativity should lead to higher participation in open source projects for several reasons. First, it should focus the developer's attention on the task of coding. Second, it should help developers to persist in solving difficult or challenging problems by exploring alternatives and "thinking outside the box." Indeed, a survey study by Lakhani & Wolf (2005) provides empirical support for a link between creativity and task participation. The researchers found that a personal sense of creativity has the strongest association with effort (hours worked) by contributors to open source projects. As such, we would expect contributors with higher intrinsic motivations to participate more substantially in open source projects, because intrinsic motivation promotes the characteristics needed to perform software development. Thus:

H5: Contributors' intrinsic motivations to participate in open source projects are positively related to their level of participation.

2.3 Participation and Performance in open source Development

As we have noted, psychologists distinguish performance from behavior. Performance is the outcome of an evaluation by others of an individual's behavior, and this behavior is often manifested by individuals' task output (Mitchell & Daniels 2003). In the context of open source development, several open source communities periodically evaluate the actual contributions of their members and assign each member a certain performance ranking. These rankings are based on merit and reflect the contributors' level of participation in the open source community. Advancement within the meritocracy recognizes individuals' commitment and contributions to the open source projects (Fielding 1999). Hence, we expect that:

H6: Contributors' level of participation in open source projects is positively related to their performance ranking.

In considering the antecedents of participation and performance, we control for individual contributors' knowledge, skills, and abilities in terms of their level of education and experience. In software development, both education and experience are very important antecedents of productive capacity (Ang et al. 2002). Thus, consistent with the literature on task performance (Campbell & Pritchard 1976), we expect that developers' education and experience positively relate to their level of participation in open source projects.

Past Performance and Subsequent Motivations in open source Development

Studies in psychology show that although incentives and feedback that are viewed as controlling can decrease an individual's subsequent intrinsic motivation (Deci & Ryan 2000), the opposite is true if these external evaluations are presented as indicators of personal competence (Pittman et al. 1980; Ryan et al. 1983). Incentives that have an informing aspect about task performance can increase individuals' feelings of internal control and self-efficacy, and can raise their level of enjoyment in the task (Lepper & Henderlong 2000). This is because people tend to like to do things they think they are good at (Sansone 1986). Thus, all else equal, an increase in an individual's perceived competence at an activity should increase his or her level of intrinsic motivation for the task. In the open source context, we expect that merit-based performance rankings would have more of an informing aspect than a controlling aspect. An advance in rank communicates important information to contributors about their ability and productive capacity in software development as well as the value of their contributions to the open source community. Following Sansone (1986), an increase in ranking should therefore enhance contributors' sense of competence, self-efficacy and enjoyment in participating. This implies that,

H7: An increase in contributors' performance ranking is associated with an increase in their intrinsic motivations to participate.

As early as 1943, Hull proposed that motivation can arise from the reinforcer itself. He termed this incentive motivation. Incentive motivation is dependent upon the strength of the incentive. As the size of an incentive increases, so does the level of incentive motivation and the likelihood of the individual behaving in such a way as to bring about an even greater incentive in the future. This is because the receipt of extrinsic incentives imparts information about the likelihood of receiving future extrinsic incentives for similar behaviors (Lepper & Henderlong 2000). Expectations of future incentives can thus provide continued extrinsic motivation for an individual to engage in previously rewarded activities.

In the context of open source communities, increases in performance ranking could increase contributors' internalized extrinsic motivations in several ways. From a psychological point of view, a developer who contributes code that improves the use value of the software will experience a subsequent increase in rank, and this will reinforce the internalization process of the use value motivation through identification (Ryan & Deci 2002). Thus, she could expect that future use value contributions will strengthen her identification with the community and lead to renewed internalization through future increases in rank. In this way, a prior increase in rank could increase developer's internalized extrinsic motivation to make use value contributions in the future. Hence:

H8a: An increase in contributors' performance ranking is associated with an increase in their subsequent use value motivations to participate.

In addition, a promotion in rank is certainly associated with greater status in the open source community (Raymond 2001a; Lerner & Tirole 2002). Following Ryan (1982), a promotion in rank in one time period could therefore reinforce internalization of the status motivation through introjection. The increase in a contributor's status motivation could then lead to future source code contributions. Inasmuch as an increase in status is seen as career enhancing (Lerner & Tirole 2002; Hann et al. 2002), a rank increase will also lead to a reinforcement of the status motivation through introjection. Thus, we hypothesize that:

H8b: An increase in contributors' performance ranking is associated with an increase in their subsequent status motivations to participate.

Method

We evaluate our hypotheses empirically, analyzing archival data collected from open source project records over a period of four years, and from a targeted survey of open source participants. The following describes the setting of the data collection, each data source, and our measures of key variables.

Research Setting

We investigated three major open source projects under the control of the Apache Software Foundation (ASF). The ASF includes a number of subprojects related to the development of a full-featured WEB-server product offering. We studied the largest and most significant of these projects including the Apache WEB server project which is a freely available source code implementation of an HTTP server and is the project around which the Apache Group initially formed; the Jakarta project which currently consists of 19 Apache related Java subprojects; and the XML project which currently consists of 16 Apache related XML subprojects.

The Apache context is very well suited for examining the relationships between motivation, participation and performance in open source development. As a meritocracy, status, responsibility, and benefits are commensurate with contribution (Fielding 1999). There are several observable levels of recognition or rank within the ASF. In order of increasing status, these are *developer*, *committer*, *project management committee member*, and *ASF member.* In all cases, advancement within the hierarchy is in recognition of an individual's commitment and contributions to an Apache project. Further, while the number of attainable ranks is limited, the number of promotion opportunities at any rank is not constrained. For example, there is no limit to the number of contributors who can achieve the rank of "developer" or to the number of developers who can be promoted to the rank of "committer". Similarly, a promotion to ASF member is not contingent on a number of predetermined positions (Fielding 1999). Thus, there is no rationing of promotion opportunities, and advancement reflects an objective measure of a positive peer review of one's performance.

Data Collection

<u>Archival Data</u>

All open source work products are placed in the public domain under various "free software" licensing arrangements. Apart from the source and binary codes of the actual software programs, Apache products include developer WEB sites, change logs, documentation, and developer communications in the form of email archives. From these products, we extracted two types of information: each contributor's progression along the Apache career path, and each contributor's source code contributions to the project.

To assess a contributor's performance, we captured the upward progression as a series of discrete transitions from one level to another in the ASF meritocracy. This resulted in a time line for the promotion of individuals within each project. To extract information about individual contributions, we developed tools to mine submissions of the individual developers. A submission to an open source project is known as a "patch" – an analogue to modification requests in traditional software development environments. The data encompassed contributions made and accepted into any of our three target Apache projects. Data collection was completed in January 2003 and included all contributions from 1999 through 2002.[21]

Survey Data

A secure, WEB-based survey of Apache contributors was conducted to obtain respondent motivations for participation in the project. Dr. Roy Fielding, then chairman of the ASF, introduced the survey to 1,301 uniquely identified contributors via e-mail in November 2000. Two hundred thirty-three e-mail invitations were undeliverable. Of the remaining 1,068 contributors, 325 completed the instrument, yielding a response rate of 30%. Thirty-seven responses contained one or more missing pieces of information and were thus dropped from further analysis, yielding a usable sample of 288 responses. An analysis of response bias using nonparametric tests of location and empirical distribution indicates that our sample is representative of the overall population of Apache contributors.[22]

Measures

In the following paragraphs, we first define our measures of motivation followed by a description of our measures of open source participation and performance. Our data form a panel covering years 1999 through 2002 (denoted as periods 1 through 4) with measures of individual participation (in periods 2 and 3), performance (in periods 3 and 4), and past performance (in periods 1 and 2). To this panel, we add cross-sectional survey data, collected toward the end of period 2, containing the indicators of respondent motivations. The panel provides the basis for the development of all subsequent measures and tests of hypotheses. To highlight the temporal nature of the measures and their relationships, we use the following notation. For variables denoted as $X_{t_a^z}$, the subscript t_a^z denotes the sum of variable X for periods a through z. Also note that, $X_{t_2^2}$, represents the cross-sectional value of variable X for period 2.

---

[21] Details on the research site along with a more detailed description of the archival data and the data extraction process can be found in Appendix III.

[22] Detailed results from our assessment of response bias are reported in Appendix 4-A.

Motivations

Following theoretical discussions of work motivation (Mitchell & Daniels 2003), we conceptualize open source participation behavior as being "driven" by different motivational underpinnings. These motivations can be thought of as existing on a continuum ranging from purely *extrinsic* to purely *intrinsic*. Adopting this perspective, monetary compensation is viewed as more strictly extrinsic than other types of extrinsic incentives (e.g., Calder & Staw, 1975). Individuals are extrinsically motivated if they are able to satisfy their needs indirectly, especially through monetary compensation; indeed, the "ideal" extrinsic incentive identified in the literature is strict "pay for performance" (Osterloh & Frey 2000, p. 539). Consistent with this literature, we measured *extrinsic* motivation ($\text{EXTRINSIC}_{t_2^2}$) as the average number of hours per week for which respondents were paid for their Apache development efforts.

We also conceptualized two internalized extrinsic motivations: *use value* and *status*. To measure *use value* motivations ($\text{USEVALUE}_{t_2^2}$), we draw upon the conceptualization of use value in the open source literature (e.g., von Hippel 2001), using two scale items that capture the extent to which solving bugs or problems or adding needed features is important to developers in motivating their participation. We assess *status* motivations ($\text{STATUS}_{t_2^2}$) using measures consistent with the open source literature regarding the motivating potential of status (Raymond 2001a). Four scale items capture the extent to which participants are motivated by status considerations to make contributions.

Finally, consistent with the literature on motivation (Lepper & Henderlong 2000), we operationalized *intrinsic* motivation in terms of the motivating potential of the task itself. That is, intrinsic motivation is the extent to which participants make code contributions because developing software is an activity they enjoy and one that satisfies their needs for competence, control or autonomy. We measured *intrinsic* motivation ($\text{INTRINSIC}_{t_2^2}$) using four scale items designed to capture the extent to which open source participants are motivated by aspects of the task itself to make contributions.

Following Anderson & Gerbing (1988) we assessed the psychometric properties of the motivation measurement scales used in this study for content, convergent, and discriminant validity as well as for reliability. Overall, the tests provide strong support for the reliability and validity of the motivation measures used in this study.[23]

Participation

While open source volunteers provide many different kinds of valuable services to their respective projects (Shah 2003), a principal participation behavior consists of authoring and maintaining the software, *i.e.*, writing lines of software code. Insider accounts regarding open source project organization and operation suggest that it would be improbable for participants to advance in the Apache meritocracy without substantive and sustained software code contributions.[24] Thus, we measured participation based upon the number of source code contributions submitted and accepted by the project. A potential concern with this measure is whether accepted source code contributions equal submitted contributions, that is, does consideration of only accepted contributions underestimate participation. To investigate whether this concern is salient for the Apache projects in our study, we investigated all contributions submitted by a randomly selected 10% of the contributors in our sample. Because Apache does not distinguish or track the number of "accepted" versus "rejected" submissions, we searched through email archives to follow the history of each participant's interactions with the Apache community. We found that the participation behavior of these contributors was more similar to a "revise and resubmit" process than to an "accept or reject" process. As described by other open source researchers, the participants in our sample first engaged in getting-to-know behavior by analyzing the source code and/or participating in discussions on message boards; Von Krogh et al. (2003, p. 16) have characterized this behavior as "a significant period of observation (lurking)". Often, the contributors' submissions were accepted without modification. Sometimes, contributors were asked to make changes

[23] Appendix 4-A lists the measurement scale items and their sources along with reliability and validity analyses for each measure of motivation.

[24] Personal exchange with Dr. Fielding.

before their submissions were accepted. This "revise and resubmit" procedure is similar to that described by Raymond (2001a) and Markus et al. (2000). None of the code submissions for our random sample of Apache contributors was rejected outright. This suggests that our measure of participation as code submitted *and* accepted provides a reasonable estimate of the level of participation for contributors.

To capture the immediate and subsequent effects of motivations on behavior, our measure of participation ($\text{PARTCIPATION}_{t_2}^{3}$) is derived using contribution-based metrics for periods 2 and 3 – the period covering the survey and the following period. As a check of robustness, we derived several alternative measures of contributions. The first measure is simply the cumulative number of patches submitted and accepted into the software revision control system for the particular year. The number of lines of software code written or changed is a commonly used productivity metric in software development organizations (Boehm et al. 2000); thus, our second measure is the cumulative number of lines of code submitted and accepted for the given year. Finally, to account for potential productivity differences between programming languages of the Apache subprojects under consideration, each contribution is converted to a common function point metric using industry standard language conversion factors (Boehm et al. 2000). Our primary analysis uses the function point metric as the measure of contributors' participation. Lastly, to control for project level idiosyncrasies that may influence participation, we operationalized each measure of participation as the deviation of the measure from the applicable subproject mean.

Performance

To operationalize individual performance we leveraged the fact that Apache operates as a meritocracy (Fielding 1999). Promotion to a higher rank within the Apache hierarchy is awarded after one or more cycles of contribution followed by a positive peer review and thus, is an acknowledgement of an individual's substantive contributions to the project. This operationalization of performance is consistent with the literature in psychology; as Mitchell & Daniels (2003) explain, performance is an "outside standard that is … usually assessed by others" (p. 227). Measuring performance as rank advancement in the Apache meritocracy satisfies the important criteria that someone other than the individual being evaluated is making the rating. In addition, meta-analysis studies in psychology suggest that measuring performance over time using measures of promotional progress is one of the most reliable ways to measure performance, and that measures of promotional progress also have higher validity than other types of performance measures (Meyer 1987). Rank advancement in the Apache meritocracy is clearly an indicator of promotional progress, and thus should be a reliable and valid measure of performance.

We operationalized our performance measure (PERFORMANCE $_{t_3^4}$) as the number of the changes in ASF rank experienced from period 3 to period 4 – the period after our measures of motivation and participation. This temporal distinction between our measures of motivation, participation and performance is consistent with the general model of motivation and performance in psychology (e.g., Mitchell & Daniels 2003) in which the relationship between motivation, behavior and performance is properly considered as a sequence and not as simultaneous events. As we did for our measure of participation, we control for possible subproject differences in our performance measure by operationalizing respondent performance as a deviation from the subproject mean performance.

Contributor's Knowledge, Skills and Abilities

The literature on performance has identified individual characteristics such as knowledge and skills as antecedents of participation. As described earlier, these characteristics are difficult to measure, and are frequently assessed through the use of proxies, such as the level of education and experience. Following the extant research (e.g., Ang et al. 2002), we measured these constructs using demographic survey items in which respondents reported their years of education (EDUCATION $_{t_2^2}$), and their total years of work experience (EXPERIENCE $_{t_2^2}$).

Analysis and Results

Our theoretical model stipulates measurements in various time periods. Motivation is an antecedent of participation, and participation is an antecedent of performance. Naturally, the data underlying these constructs has to reflect this sequence. In our research design this requires the collection of data of participation and performance of up to two years *after* the measurement of the motivation constructs. Using past performance as an antecedent of motivation requires data collection of up to two years *prior* to the measurement of the motivation constructs. Including past performance in this model reduces the number of data points due to 'late entry' into the Apache career. Hence, we follow a two-step estimation strategy. In the first step we estimate the motivation – participation – performance relationships with all respondents, thereby utilizing the maximum number of available data points. To test the past performance – motivation relationships of H7 and H8, we augment our primary model with our measure of past performance, analyzing data only from those respondents who had started their Apache careers as of period 1.

Model and Estimation

In order to test our hypotheses, we specified the following structural equations in a simultaneous equation model (SEM):[25]

$$\text{PERFORMANCE}_{t_3^4} = \alpha_0 + \alpha_1 {*}\text{PARTCIPATION}_{t_2^3} + \varepsilon_1$$

$$\text{PARTCIPATION}_{t_2^3} = \beta_0 + \beta_1 {*}\text{INTRINSIC}_{t_2^2} + \beta_2 {*}\text{EXTRINSIC}_{t_2^2} + \beta_3 {*}\text{USE VALUE}_{t_2^2} + \beta_4 {*}\text{STATUS}_{t_2^2} + \beta_5 {*}\text{EDUCATION}_{t_2^2} + \beta_6 {*}\text{EXPERIENCE}_{t_2^2} + \varepsilon_2$$

$$\text{INTRINSIC}_{t_2^2} = \gamma_0 + \gamma_1 {*}\text{EXTRINSIC}_{t_2^2} + \gamma_2 {*}\text{USE VALUE}_{t_2^2} + \gamma_3 {*}\text{STATUS}_{t_2^2} + \{\gamma_4 {*}\text{PERFORMANCE}_{t_1^2}\} + \varepsilon_3$$

$$\text{USE VALUE}_{t_2^2} = \eta_0 + \eta_1 {*}\text{EXTRINSIC}_{t_2^2} + \{\eta_2 {*}\text{PERFORMANCE}_{t_1^2}\} + \varepsilon_4$$

$$\text{STATUS}_{t_2^2} = \phi_0 + \phi_1 {*}\text{EXTRINSIC}_{t_2^2} + \{\phi_2 {*}\text{PERFORMANCE}_{t_1^2}\} + \varepsilon_5$$

Our analysis approach follows the factor analytic (FA) simultaneous equation model (SEM) (*i.e.*, FASEM) approach most commonly used to evaluate path analysis models with latent variables (Anderson & Gerbing 1988). Due to non-normality in our data (Normalized Mardia's Coefficient = 11.32), we estimated our model using elliptically re-weighted least squares (ERLS). ERLS has been shown to be superior to maximum likelihood when estimating models where data exhibit even moderate departures from multivariate normality (Sharma et al. 1989).

[25] Lagged PERFORMANCE (for periods 1 and 2) is in equations for INTRINSIC, USE VALUE and STATUS *only* to test H7 & H8.

We evaluate model fit using a two-index comparison strategy. Following Hu & Bentler (1997), we judged the adequacy of the hypothesized models by first examining the χ^2 for significance and the χ^2/df ratio for a value less than 3. Then, we compared the model SRMR to a cutoff value of 0.08 and either a CFI "close to" 0.95 or an RMSEA where the 90% confidence interval includes a value of 0.07 or less.

Results

We assessed model performance against several standard model instantiations (Marsh 1994).[26] The first is a saturated model where all possible paths between structural model variables, both manifest and latent, are estimated. The saturated model provides a basis for subsequent model comparisons as all other models containing the same variables will be nested within this model. The second is the null or independence model where all covariances between structural variables are set to 0 and are thus unrelated. The null model is the base model for the computation of relative fit indices such as the CFI or the NNFI as well as the theoretical model's χ^2. We computed χ^2 difference statistics ($\Delta\chi^2$) between the theoretical and alternative models under consideration as well as a comparison of model fit indices (Bentler & Bonett 1980).

Our hypothesized model fits the data very well. The theoretical model's overall χ^2 statistic is significant. The χ^2/df ratio is well within acceptable range with a $\chi^2/\text{df} = 1.28$. Model fit is judged acceptable using the SRMR (0.06) in combination with the CFI (0.98). Taken together, these statistics indicate that our hypothesized model provides a highly acceptable fit to the data. Utilizing the χ^2 difference test, we compared the performance of the hypothesized model with that of the alternative models. The hypothesized model performance is clearly superior to that of the independence model on measures of fit. Compared to the saturated model, the hypothesized model performs quite favorably as well ($\Delta\chi^2 = 1.22$, $\Delta\text{df} = 12$, $p < .001$). In this case, the hypothesized model can be viewed a constrained

[26] The means, standard deviations, correlations and covariances for the variables in our model and a summary of model fit evaluation results are reported in Appendix 4-A.

version of the saturated model where 12 of the model paths are constrained to zero. Thus constrained, the hypothesized model achieves a statistically indistinguishable level of performance from the fully saturated model. Consistent with the $\Delta\chi^2$, relevant model fit statistics are unaffected by the additional constraints placed on the model.

Figure 4.4 shows the estimated standardized path coefficients and model fit statistics for both our primary and enhanced models. H1a, H1b and H1c predicted a negative relationship between strictly *extrinsic* motivations and *intrinsic* motivations, and between the internalized extrinsic motivations (*use value* and *status*) and *intrinsic* motivations. These hypotheses are not supported. The paths from USE VALUE$_{t_2}^{2}$ to INTRINSIC$_{t_2}^{2}$ and from EXTRINSIC$_{t_2}^{2}$ to INTRINSIC$_{t_2}^{2}$ are not significant in our model. The path from STATUS$_{t_2}^{2}$ to INTRINSIC$_{t_2}^{2}$, while significant, is positive, thus in the opposite direction than predicted (γ_3 = .542, $p <$ .01). H2a and H2b predicted positive relationships between being paid to contribute to Apache and contributors' *status* and *use value* motivations. These hypotheses are partially supported. The path from EXTRINSIC$_{t_2}^{2}$ to STATUS$_{t_2}^{2}$ is positive and significant (ϕ_1 = .157, p = .04), providing support for H2b. However, the path from EXTRINSIC$_{t_2}^{2}$ to USE VALUE$_{t_2}^{2}$, while significant, is in the opposite direction than predicted by H2a (η_1 = -.132, p = .05). H3 relates the strictly *extrinsic* motivation of being a paid Apache developer to participation. This hypothesis is supported, as the path from EXTRINSIC$_{t_2}^{2}$ to PARTCIPATION$_{t_2}^{3}$ is positive and significant (β_2 = .152, p = .02). H4a and H4b relate the *use value* and *status* motivations to participation. These hypotheses are partially supported, as the path from STATUS$_{t_2}^{2}$ to PARTCIPATION$_{t_2}^{3}$ is significant and in the hypothesized direction of H4b (β_4 = .256, $p <$ .01). However, the path from USE VALUE$_{t_2}^{2}$ to PARTCIPATION$_{t_2}^{3}$, while significant, is in the opposite direction than predicted by H4a (β_3= -.208, $p <$ .01). H5 relates participants' *intrinsic* motivations to their level of participation. This hypothesis is not supported as the path from INTRINSIC$_{t_2}^{2}$ to PARTCIPATION$_{t_2}^{3}$ is not significant. Finally, H6 relates participation to changes in performance ranking. The path from PARTCIPATION$_{t_2}^{3}$ to PERFORMANCE$_{t_3}^{4}$ is positive and significant (α_1 = .178, p = .02), supporting H6.

To test H7, H8a, and H8b, we enhanced our primary model by adding measures of participants' past performance and paths to relate them to the participants' intrinsic motivations and internalized extrinsic motivations (*use value* and *status*), respectively. We then re-estimated the enhanced model including the original paths and variables and the newly added paths and variables. The enhanced model, like the primary model, exhibits an excellent fit to the data. The model's overall χ^2 statistic is significant with a χ^2/df ratio of 1.2, well within acceptable range. Model fit is judged acceptable using the SRMR (0.06) in combination with the CFI (0.98). Again, all $\Delta\chi^2$ were computed between the hypothesized and alternative models. In all cases, the $\Delta\chi^2$ indicated in favor of the hypothesized model; thus, the hypothesized model was retained as the best fitting model and serves as the basis for the examination of our remaining hypotheses.

Figure 4.4 – Model Results: Standardized Path Coefficients

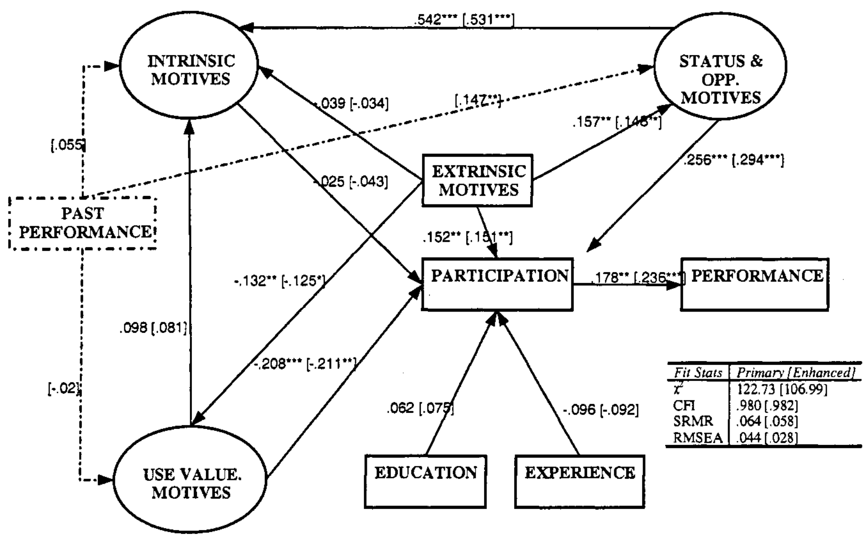


Figure 4. Notes – Coefficients are shown for both Primary and Enhanced models. Coefficients for Enhanced Model appear in [ ].

H7 positively relates past performance to *intrinsic* motivations. We find no support for H7, as the path from $\text{PERFORMANCE}_{t_1}^2$ to $\text{INTRINSIC}_{t_2}^2$ is not significant. H8a and H8b positively relate past performance to the *use value* and *status* motivations. These hypotheses are partially supported. Although the path from past $\text{PERFORMANCE}_{t_1}^2$ to $\text{USE VALUE}_{t_2}^2$ is not significant, the path from past $\text{PERFORMANCE}_{t_1}^2$ to $\text{STATUS}_{t_2}^2$ is positive and significant, as we had predicted ($\phi_2 = .147$, $p = .05$).

Discussion

This study has examined the inter-relationships between the motivations, participation and performance of open source developers. Our first finding reveals that contributors have multiple motivations to participate in open source projects, and that while some motivations are complementary, others are not. Although the previous open source literature suggests that participants have multiple motivations for contributing, these motivations are thought to be "mutually reinforcing" (e.g., Markus et al. 2000). However, our findings suggest that this is not always the case. Specifically, and contrary to our expectations, we find no evidence of extrinsic motivations crowding out strictly intrinsic motivations.

However, our results do suggest associations among some motivations. Status motivations actually enhance intrinsic motivations. In addition, being paid to contribute positively relates to participants' status motivations. A potential explanation for these complementarities is offered by Sansone and Smith (2000) who suggest that extrinsic motivations can boost, regulate and maintain interest in doing a task. For example, contributors' desires to further their careers may enhance their inherent interest in making code contributions because making contributions can also help them to achieve higher status or to obtain better career opportunities. On the other hand, we find that being paid to contribute is negatively associated with participants' use value motivations. In hindsight, this may not be too surprising. Extrinsically motivated developers may be more likely to view their contributions as part of their employment relationship and thus may have less personal use value for the Apache software. An alternative explanation is that a contributor who receives an extrinsic incentive for his or her use value motivated contribution will likely require such incentives in the future for submitting contributions that increase use value. In this sense, we do observe a crowding out effect, not of an intrinsic motivation, but of an internalized extrinsic motivation.

Our second finding suggests that not all motivations affect open source participation equally or in the same way. Some studies identify participants' most "dominant" motivations for engaging in open source projects (e.g., Ghosh et al. 2002). However, these studies do not identify how differences in actual participation levels are associated with different motivations for participating. As we predicted, being paid to contribute and status motivations are related to above average participation levels. However, contrary to our predictions, we find no significant relationship between intrinsic motivations and participation levels. At first glance, this finding is puzzling. Studies of open source contributors have revealed that contributors do enjoy participating; indeed, developers have reported flow states and losing track of time when working on open source projects (Lakhani & Wolf 2005). Why do these feelings of enjoyment in programming not translate into higher levels of code contributions? The literature on motivation and performance offers some potential explanations. Researchers have identified some disadvantages associated with intrinsic motivation (Osterloh & Frey 2000). Intrinsically motivated contributors tend to be more autonomous and self-directed, which may lead them to exhibit less desirable behaviors. In addition, Lepper & Henderlong (2000) suggest that intrinsic motivation may not be associated with better performance if the aspects of the activity that make it interesting come at the expense of attention toward some outcome – in this case, being intrinsically motivated may not positively affect participation levels. Since open source contributors are self-directed, it is possible that their intrinsic motivation is not fully aligned with the mission of the open source community. For example, a potential contributor may be intrinsically motivated to work on perfecting one feature of the code that performs a relatively minor function, and may spend inordinate amounts of time perfecting the single feature rather than implementing a large number of contributions. Also contrary to our predictions, use value motivations are associated with below average contribution levels. This result could be explained by reconsidering the nature of use value motivations. Developers who contribute because they are motivated by use value want to solve a particular bug that is causing them trouble or to add a particular feature that they need to use. Once they have solved the immediate problem or added the particular feature, they may lose interest in making future contributions if there are no further salient problems or issues to motivate them. This is consistent with the argument put forth by Sansone & Smith (2000) who assert that, without motivations that help to sustain interest, individuals could lose interest over time, even in activities they have previously found motivating. This finding has important implications for those interested in sustaining the participation of contributors to open source projects; contributors who are primarily motivated by use value considerations may need subsequent incentives to sustain their level of participation.

Our third finding suggests that the Apache meritocracy is operating effectively, and that promotions within the community are indeed based on actual contributions to the Apache projects. This is especially important for a community like Apache that depends on shared leadership and the contributions of participants for its success.

Finally, we have found that past performance rankings enhance some motivations for participating but not others. We expected that an increase in ranking would boost contributors' subsequent intrinsic and extrinsic motivations. Our findings support this expectation for status motivations, as an increase in performance ranking is associated with a subsequent increase in contributors' status motivations to participate. This finding is consistent with the notion of incentive motivation where receiving a reward increases motivation for an individual to perform previously rewarded activities. As Lerner & Tirole (2002) have noted, a promotion in rank enhances a contributor's status in the open source community and increases incentive motivation. Thus, an increase in status due to an advance in rank should amplify the contributor's subsequent status motivations to participate. However, we find no significant associations between rank increases and intrinsic motivations or use value motivations. Our results suggest that a rank increase within the Apache community is neither considered controlling and thus diminishing intrinsic motivation, nor is it viewed as informative on competence and thus enhancing intrinsic motivation. This could be the result of the performance evaluation itself; rank increases are not typically accompanied by detailed reports and feedback on performance, but rather just announced. Thus, it may not be clear to the contributor exactly what aspects of his or her performance are superior.

Conclusions

Our study makes several important contributions to the open source literature. First, our theoretical model and empirical evaluation increase the understanding of how motivations, participation and performance *interrelate* in open source projects. While some studies have examined particular aspects of motivation or participation, we are not aware of any study that has examined the *system* of inter-relationships between motivations, participation and performance. Yet, one must consider the motivational system in order to understand how successfully an open source community is functioning in terms of attracting and sustaining participation. To our knowledge, this is also the first study that investigates how past performance affects subsequent motivations in open source communities. Understanding this relationship provides insight into the motivational effectiveness of feedback in the open source setting.

As a whole, our results have several implications for attracting and sustaining participation in open source communities. First, our results suggest that open source communities should largely welcome commercial efforts by companies. While extrinsically motivated contributors have lower use value motivations, they also exhibit greater status motivations, and being paid to contribute is associated with a higher level of contributions to the source code. More importantly, in our setting, we could not detect any crowding out of (strictly) intrinsic motivations by extrinsic motivations. This is perhaps more important, as many successful open source projects (e.g., Apache, Linux, Mozilla, etc.) experience increased attention from leading software producers who pay employees to contribute to these projects. Second, developers with higher status motivations appear to be the more substantive contributors. An implication is that open source communities may want to nurture such motivations, perhaps by devoting distinct website space to recognize distinguished developers or by promoting involvement in open source communities as leverage in the labor market. Third, the positive relationship between past performance and status motivations indicates that a feedback system provides a valuable service to the open source community by increasing these motivations. However, the current feedback system, which primarily is an announcement of a rank increase, fails to enhance intrinsic motivations. Feedback that indicates personal competence should increase intrinsic motivation (Pittman et al. 1980; Ryan et al. 1983). Hence, one direct implication of our results is that open source communities should capture a competence component in the feedback system that stresses the competence of the contributor and that provides detailed information and feedback on performance. This could, for example, include a listing of achievements and extraordinary contributions.

Our study has several strengths and limitations. We have drawn upon the extensive literature on motivation in psychology and on the open source literature to develop theoretically driven hypotheses. Our research design leverages multiple data sources, matching subjective survey data on motivations with objective measures of participation and performance. This approach helps to triangulate findings and mitigate common method and source biases. In addition, we leverage archival data to capture participation and performance measures over a four-year period. This longitudinal approach allows us to tease out potential causality relationships between motivations, participation and performance and between and past performance and motivations. Our study focuses only on projects within the Apache open source community. While this potentially limits our findings in a strict sense to the Apache project, we believe that our results could be applicable to other successful open source communities such as Linux, Perl, and Mozilla that share an interesting relationship between the intrinsic and extrinsic motivations of contributors. Our approach does offer some persuasive advantages. Focusing on one open source community enables us to link developers to their actual code contributions, and to link these code contributions to objective performance evaluations by others in the Apache community. This approach enables us to cleanly capture the past performance – motivation – participation – performance relationships. In addition, eliciting broad participation in our study within one community rather than limited participation across many communities helped us to obtain a representative sample. Hence, our research design increases the internal validity, external validity and statistical conclusion validity of our results (Campbell & Cook 1979).

Our study opens up several important avenues for further research in the open source arena. In our study, we were guided by the previous literature in psychology and open source software development to identify the most likely factors leading to open source participation. However, it is possible that other factors could be salient. For example, some researchers have suggested an obligation/community-based intrinsic motivation (Lindenberg 2001) for contributing to open source projects. This motivation has certain similarities with the use value motivation we have examined as use value is internalized based on identification with the open source community. Our results for use value would suggest that obligation/community-based intrinsic motivation may be associated with below average contribution levels. However, the precise relationships between obligation motivations, participation levels and performance would need to be examined in future research. Investigating the inter-relationships among motivations, participation and performance in other open source communities is another important research extension. In addition, most of the literature in psychology on intrinsic and extrinsic motivation is validated via carefully designed and controlled experiments. An experimental approach could be especially useful in revealing the processes by which extrinsic motivations for contributing to open source projects are internalized and the mechanisms by which different internalization processes lead to differences in participation. Experiments could also be helpful for designing feedback mechanisms to maximize the effect of past performance on future motivations. Further research on the motivational mechanisms underlying participation and performance is vital for effectively leveraging the advantages of "costless" open source development.

Appendix 4-A. Descriptive Statistics of Data and Model Performance

Sample Bias Tests of Survey Data

One important aspect of any survey is determining whether respondents differ from non-respondents. A comparison of these groups can reveal how representative is the sample of the population along relevant dimensions. Participant contribution levels and rank within the ASF play a critical role in the analysis and interpretation of the results that follow. To discern whether the distribution of these two variables in the sample is comparable to that of the population, we performed a series of non-parametric tests of location and empirical distribution. Similar to prior open source research (Mockus et al. 2000), we find measures of rank and contribution are asymmetrically distributed and thus must rely on distribution free (non-parametric) statistics to compare our sample and population. The tests evaluate the hypothesis that respondents and non-respondents are drawn from the same underlying population for the variables of interest. For respondent rank, the p-value for the Kolmogorov-Smirnov (KS) statistic of 0.56 for the two-sample test was 0.91, supporting the null hypothesis that the empirical distribution of rank is identical for our sample and the Apache population. The Mann-Whitney U test for location also supports the null hypothesis of no difference ($Z=1.49$, $p = 0.14$). Likewise for contributions, both tests support the null hypothesis of no difference between our sample and the population ($KS=0.86$, $p = 0.45$; $Z=-0.18$, $p = 0.86$). These results suggest that our respondents and non-respondents are drawn from the same underlying population.

A particular strength of our research design is our collection of data from different sources and using different methods. In doing so, we have limited our exposure to common method and common data source biases (Straub et al. 1995). Even so, to assess the degree to which common method bias might present a problem for our subjectively measured motivation constructs; we subjected the motivation scale items to a principal components analysis using a varimax rotation (Harman 1967). Results indicated that the items loaded cleanly on the factors representing the expected constructs. Thus, we found no general factor that would have emerged due to common method variance; this suggests that common method bias is not an issue for the motivation constructs used in this study.

Psychometric Properties of Scales for Measures of Motivation

Following Anderson & Gerbing (1988) we assessed the motivation measurement scales used in this study for content, convergent, and discriminant validity as well as for reliability. The motivation constructs are subjected to an evaluation of measurement properties because, in contrast to participation and performance (which are measured objectively) the motivation constructs are subjectively measured using multiple indicator items. *Content validity* was established via an iterative process of expert-influenced questionnaire development. A WEB-based pilot questionnaire was created and tested using five open source developers who were either students or employees of the university where the research was conducted. In addition, a founding member of Apache Software Foundation, Dr. Roy Fielding, reviewed and commented on the instrument under development. All respondents were debriefed to assess the clarity of the questions and the adequacy of the candidate scale items in capturing their motivations for participation. All comments were incorporated, and the resulting instrument was reviewed a final time by Dr. Fielding. The remaining properties of the motivation measurement scales were assessed using confirmatory factor analysis (CFA) techniques and measures. *Scale reliability* was assessed using composite reliability and coefficient alpha scores. Similar to alpha scores, composite reliabilities reflect the internal consistency of the scale items measuring a given factor (Nunnally & Bernstein 1978; Fornell & Larcker 1981). Both composite reliability and coefficient alpha scores exceeded the recommended cutoff value of .70. *Convergent validity* was demonstrated when all scale items loaded significantly on their hypothesized latent constructs (Anderson & Gerbing 1988). Lastly, discriminant validity was assessed by computing a series of $\chi 2$ difference tests between the standard CFA model where all factors co-vary and a constrained model where inter-factor correlations are fixed at 1. All $\chi 2$ difference statistics were significant, demonstrating discriminant validity (Bagozzi & Phillips 1982). Overall, these tests provide strong support for the reliability and validity of the motivation measures used in this study.

Scale Items, Factor Loadings, and Reliabilities for Measures of Motivation

Extrinsic motivations are assessed in terms of the extent to which the respondent is paid to contribute to the Apache projects. We used the following two questions. "At any time during this year, were you paid to participate in Apache projects?" and "If 'YES', approximately how many hours a week were you paid for participating in Apache projects?" Hence, we operationalize extrinsic motivation ($EXTRINSIC_{t_2}$) as the average number of hours per week for which respondents received monetary compensation for their Apache development efforts.

Five-point Likert scales were used to measure the remaining three dimensions of motivation – intrinsic, status and use value. Respondents were given the following introduction to the motivation scale items. "People volunteer, or otherwise do unpaid work, for many different reasons. They also derive different kinds of satisfaction from such work. Following are some reasons other developers have given us regarding their participation in open source projects. Thinking of your own participation, please indicate how important each reason is to you personally." Response choices ranged from 1 (little importance) to 5 (great importance). Both pre-existing and new scale items were used. The scale items, reliabilities, and factor loadings are given in Table 4.9 below.

Model Fit Evaluation Statistics

Panel a – Primary Model (N=288)

	Chi-Square	*Df*	*CFI*	*SRMR*	*RMSEA*	*NNFI*	*GFI*
Independence Model	1230.85	105	0.00	0.22	0.19	0.00	0.57
Uncorrelated Factors Model	222.06	95	0.89	0.14	0.07	0.88	0.88
Saturated Model	86.75	67	0.98	0.05	0.03	0.97	0.95
Theoretical Model	101.45	79	0.98	0.06	0.03	0.97	0.94

Panel b – Enhanced Model (N=180)

	Chi-Square	*Df*	*CFI*	*SRMR*	*RMSEA*	*NNFI*	*GFI*
Independence Model	1153.37	120	0.00	0.20	0.18	0.00	0.59
Uncorrelated Factors Model	225.69	113	0.89	0.18	0.06	0.88	0.86
Saturated Model	101.65	77	0.98	0.15	0.03	0.96	0.93
Theoretical Model	106.99	88	0.98	0.06	0.03	0.98	0.94

Table 4.9 – Motivation Scale Items

Question No.	*Question Text*	*Standardized Loadings*
Intrinsic Scale Items (Coefficient Alpha = .82; Composite Reliability=.82) [a,d]		
Q1	It is the satisfaction of seeing the results.	.78
Q2	It gives me the chance to do things I am good at.	.74
Q3	I really enjoy it. It is fun.	.67
Q4	It gives me a sense of personal achievement.	.71
Status Scale Items (Coefficient Alpha=.73; Composite Reliability=.79) [b,c]		
Q5	It gives me the chance to attain a recognized qualification or skill.	.65
Q6	It gives me status at work.	.70
Q7	It increases my opportunities for a better job.	.74
Q8	It gives me status in the Apache community.	.68
Use Value Scale Items (Coefficient Alpha=.71; Composite Reliability=.77) [e,f]		
Q9	I fixed a bug or problem that was causing me trouble.	1.00
Q10	I added a feature I wanted or needed to use.	.55

Table 4.9 Notes:

a Items adapted from subscales of the Multi-Item Measures of Values instrument (Herche 1994). Subscales: Being Well-Respected, Fun and Enjoyment, Sense of Accomplishment.

b Items adapted from the Career sub-scale of the Volunteer Functions Inventory (Clary et al. 1998).

c Items adapted from concepts regarding the effects of recognition on volunteerism (Fisher and Ackerman 1998).

d Items adapted from subscales of the Job Diagnostic Survey (Hackman and Oldham 1974). Subscales: Task Identity, Task Significance, and Autonomy.

e Synthesized from qualitative discussions and theory regarding open source participation (Raymond 2001a, 2001b, 2001c; Dempsey et al. 1999).

f The factor loadings for the Use Value construct reflect the almost unanimous responses to both Use Value scale items. Thus, the measurement error for this construct is very small, and the error variance is constrained to zero. We re-formulated our models using each scale item individually as a single item measure of Use Value. Treating Use Value as a manifest variable using either item individually generates results consistent with the original model where Use Value is treated as latent.

Means, standard deviations, correlations, and covariances

Panel a – Primary Model (N=288)

	Mean d	*SD*	*CONTRIB* [b,e]	*EXPER*	*EDU* [f]	*EXTRINSIC* [e]	*PERF* [b]
Q1	4.26	0.85	0.14**	-0.09*	-0.09*	0.01	0.08*
Q2	3.88	1.10	0.05	-0.07	-0.07	0.06	0.06
Q3	4.04	0.97	0.08	-0.13**	-0.10*	0.08	0.04
Q4	4.02	1.05	0.07	0.01	-0.15***	-0.03	0.10*
Q5	2.95	1.30	0.20***	-0.12**	-0.07	0.14**	0.05
Q6	2.37	1.21	0.18***	-0.08	-0.10	0.19***	0.12**
Q7	2.58	1.25	0.23***	-0.09	-0.17***	0.07	0.16***
Q8	2.21	1.18	0.33***	-0.07	-0.18***	0.05	0.09**
Q9	4.35	1.05	-0.28***	0.05	0.11	-0.12**	-0.06
Q10	4.26	1.11	-0.19***	0.02	-0.01	-0.12**	-0.04
CONTRIBUTION [a b]	261.10	2681	12.77	-0.11*	-0.01	0.22***	0.19**
EXPERIENCE [c]	5.58	4.13	-1.68	17.04	0.16***	0.03	0.01
EDUCATION [c]	16.56	2.09	-1.18	39.66	3576.30	-0.01	0.05
EXTRINSIC	3.29	9.94	0.82	0.13	-0.20	1.10	0.08
PERFORMANCE [b]	0.03	0.24	0.14	0.01	0.64	0.02	0.04

Panel a – Continued

	Q1	*Q2*	*Q3*	*Q4*	*Q5*
Intrinsic Construct Indicators					
Q1	0.70	0.57***	0.50***	0.54***	0.29***
Q2	0.53	1.22	0.52***	0.49***	0.33***
Q3	0.40	0.55	0.92	0.44***	0.29***
Q4	0.47	0.56	0.44	1.08	0.31***
Status Construct Indicators					
Q5	0.32	0.47	0.36	0.41	1.70
Q6	0.24	0.31	0.19	0.32	0.70
Q7	0.21	0.33	0.19	0.49	0.78
Q8	0.08	0.12	0.07	0.15	0.19
Use Value Construct Indicators					
Q9	-0.02	-0.06	-0.05	-0.03	-0.34
Q10	0.04	0.06	0.06	0.10	-0.09
Manifest Variables					
CONTRIBUTION [a]	0.41	0.21	0.28	0.26	0.93
EXPERIENCE [c]	-0.32	-0.33	-0.51	0.06	-0.67
EDUCATION [c]	-4.27	-4.81	-5.51	-9.53	-5.38
EXTRINSIC	0.01	0.07	0.08	-0.03	0.20
PEFORMANCE [b]	0.01	0.01	0.01	0.02	0.01

Panel a – Continued

	Q6	*Q7*	*Q8*	*Q9*	*Q10*
Intrinsic Construct Indicators					
Q1	0.24***	0.20***	0.23***	-0.02	0.04
Q2	0.23***	0.25***	0.27***	-0.05	0.05
Q3	0.16***	0.16***	0.18***	-0.05	0.06
Q4	0.26***	0.38***	0.37***	-0.03	0.08*
Status Construct Indicators					
Q5	0.44***	0.49***	0.38***	-0.24***	-0.06
Q6	1.46	0.53***	0.49***	-0.13**	0.01
Q7	0.78	1.51	0.53***	-0.15**	0.01
Q8	0.23	0.25	0.15	-0.17***	-0.05
Use Value Construct Indicators					
Q9	-0.18	-0.21	-0.07	1.21	0.51***
Q10	0.02	0.01	-0.02	0.63	1.28
Manifest Variables					
CONTRIBUTION [a]	0.79	1.01	0.46	-1.10	-0.76
EXPERIENCE [c]	-0.39	-0.46	-0.12	0.21	0.07
EDUCATION [c]	-7.14	-12.29	-4.28	7.29	-0.52
EXTRINSIC	0.24	0.09	0.02	-0.13	-0.14
PEFORMANCE [b']	0.03	0.04	0.01	-0.01	-0.01

Panel b – Enhanced Model (N=180)

	Mean [d]	*SD*	*CONTRIB* [be]	*EXPER*	*EDUC* [f]	*EXTRINSIC* [e]	*PERF* [b]
Intrinsic Construct Indicators							
Q1	4.24	0.83	0.14**	-0.09*	-0.09*	0.01	0.08*
Q2	3.88	1.10	0.05	-0.07	-0.07	0.06	0.06
Q3	4.03	0.96	0.08	-0.13**	-0.10*	0.08	0.04
Q4	4.02	1.04	0.07	0.01	-0.15***	-0.03	0.10*
Status Construct Indicators							
Q5	2.96	1.30	0.20***	-0.12**	-0.07	0.14**	0.05
Q6	2.39	1.21	0.18***	-0.08	-0.10	0.19***	0.12**
Q7	2.59	1.23	0.23***	-0.09	-0.17***	0.07	0.16***
Q8	2.21	1.18	0.33***	-0.07	-0.18***	0.05	0.10**
Use Value Construct Indicators							
Q9	4.31	1.10	-0.28***	0.05	0.11	-0.12**	-0.06
Q10	4.24	1.13	-0.19***	0.02	-0.01	-0.12**	-0.04
Manifest Variables							
CONTRIBUTION [a b]	277.87	2733.51	12.77	-0.11*	-0.01	0.22***	0.19**
EXPERIENCE [c]	5.58	4.13	-1.68	17.04	0.16***	0.03	0.01
EDUCATION [c]	16.56	2.09	-1.18	39.66	3576.30	-0.01	0.05
EXTRINSIC	3.29	9.94	0.82	0.13	-0.20	1.10	0.08
PEFORMANCE [b]	0.03	0.21	0.14	0.01	0.64	0.02	0.04
PAST PERFORM [b]	0.47	0.73	0.50	-0.26	-1.25	0.03	0.01

Panel b – Continued

	PAST PERF[b]	*Q1*	*Q2*	*Q3*	*Q4*	*Q5*
Intrinsic Construct Indicators						
Q1	0.01	0.70	0.57***	0.50***	0.54***	0.29***
Q2	0.06	0.53	1.22	0.52***	0.49***	0.33***
Q3	0.08	0.40	0.55	0.92	0.44***	0.29***
Q4	-0.03	0.47	0.56	0.44	1.08	0.31***
Status Construct Indicators						
Q5	0.14**	0.32	0.47	0.36	0.41	1.70
Q6	0.19***	0.24	0.31	0.19	0.32	0.70
Q7	0.07	0.21	0.33	0.19	0.49	0.78
Q8	0.05	0.08	0.12	0.07	0.15	0.19
Use Value Construct Indicators						
Q9	-0.12**	-0.02	-0.06	-0.05	-0.03	-0.34
Q10	-0.12**	0.04	0.06	0.06	0.10	-0.09
Manifest Variables						
CONTRIBUTION [a]	0.22***	0.41	0.21	0.28	0.26	0.93
EXPERIENCE [c]	0.03	-0.32	-0.33	-0.51	0.06	-0.67
EDUCATION [c]	-0.01	-4.27	-4.81	-5.51	-9.53	-5.38
EXTRINSIC	1.10	0.01	0.07	0.08	-0.03	0.20
PEFORMANCE [b]	0.02	0.01	0.01	0.01	0.02	0.01
PAST PERFORM [b]	0.03	0.03	0.10	0.09	0.07	0.10

Panel b – Continued

	Q6	*Q7*	*Q8*	*Q9*	*Q10*
Intrinsic Construct Indicators					
Q1	0.24***	0.20***	0.23***	-0.02	0.04
Q2	0.23***	0.25***	0.27***	-0.05	0.05
Q3	0.16***	0.16***	0.18***	-0.05	0.06
Q4	0.26***	0.38***	0.37***	-0.03	0.08*
Status Construct Indicators					
Q5	0.44***	0.49***	0.38***	-0.24***	-0.06
Q6	1.46	0.53***	0.49***	-0.13**	0.01
Q7	0.78	1.51	0.53***	-0.15**	0.01
Q8	0.23	0.25	0.15	-0.17***	-0.05
Use Value Construct Indicators					
Q9	-0.18	-0.21	-0.07	1.21	0.51***
Q10	0.02	0.01	-0.02	0.63	1.28
Manifest Variables					
CONTRIBUTION [a]	0.79	1.01	0.46	-1.10	-0.76
EXPERIENCE [c]	-0.39	-0.46	-0.12	0.21	0.07
EDUCATION [c]	-7.14	-12.28	-4.28	7.29	-0.52
EXTRINSIC	0.24	0.09	0.02	-0.13	-0.14
PEFORMANCE [b]	0.03	0.04	0.01	-0.01	-0.01
PAST PERFORM [b]	0.02	0.07	0.05	-0.02	0.05

Panel Notes:

*** p =.01, ** p = .05, * p = .10. Covariances (Correlations) in lower left (upper right) triangular matrix. Variances along the diagonal.

a Items adapted from subscales of the Multi-Item Measures of Values instrument (Herche 1994). Subscales: Being Well-Respected, Fun and Enjoyment, Sense of Accomplishment.

b Function points added.

c Measured as deviation from subproject mean.

d Measured in years.

e Means and standard deviations for nominal variable values.

f Natural log of variable used in model.

g Square of variable used in model

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5.
DISCUSSION AND CONCLUSIONS

Summary of Studies

The three studies comprising this dissertation represent a research program designed to conduct a rigorous analysis of issues related to open source developer motivations and performance. This work was designed to make a significant contribution to the emerging literature surrounding the issues of motivation to participate in open source projects and the consequences for the project as a result of varying sources of motivation. This thesis is among the first studies to yield rich, detailed data and models exploring the sources and consequences of the motivation of open source developers.

In Study 1, I empirically investigated the motivations of open source contributors for three Apache projects. Based on theories of volunteerism and job-design, I identified functionally based motivations to participate in open source projects. A survey instrument and a subsequent factor analysis were used to explore the appropriateness of the hypothesized motivations. Of the motivations, five (career impact, recreation, reputation, use value, and social) were related to latent variables or factors that held significant explanatory power for the observed correlations among the observed variables. A hybrid conjoint exercise was introduced to investigate the relative importance of the motivations to contribute to open source projects. The conjoint analysis revealed motivational profiles vary across developers conditioned on exhibited levels of commitment to the project. The dominant motivation among those exhibiting a low level of project commitment is the personal use value of the software (27%) while the dominant motivation among those exhibiting a higher level of commitment is the recreational aspects of participation (20%).

Study 2 explored the motivation engendering the most controversy within the open source community – career concerns. Specifically, I explore whether participation is consistent with well-established theories from labor economics. From this literature, I established two plausible theoretical bases for the existence of returns to open source participation – human capital and signaling theories.

Reasoning from the human capital perspective leads to the hypothesis that open source experience will be positively related to wages. While reasoning from a signaling perspective, leads to the hypothesis that successful open source participation will be positively related to wages. Analysis of multiple econometric models reveals that employers do not reward the accumulation of experience in open source projects per se. Rather, successful open source participation, measured as higher open source rank, is associated with higher wages, even after controlling for work and programming experience.

Study 3 examined the inter-relationships among the elements – motivations, participation and performance of open source developers. Structural models detailing the relationships among these elements were developed and analyzed using covariance analysis models techniques (i.e., structural equation modeling). Analysis revealed that contributors have multiple motivations to participate in open source projects and that while some motivations are complementary, others are not. Although previous open source literature suggests that participants have multiple motivations for contributing, these motivations are thought to be "mutually reinforcing" (e.g., Markus et al. 2000). In contrast, the findings here suggest that this is not always the case. Contrary to expectations, I find no evidence of extrinsic motivations crowding out strictly intrinsic motivations. The results do suggest complementarities among some motivations however. For example, status motivations are found to enhance intrinsic motivations. In addition, being paid to contribute (an extrinsic motivator) is positively related to participants' status motivations and negatively related with participants' use value motivations. A second finding suggests that not all motivations affect open source participation equally or in the same way. As hypothesized, being paid to contribute and status motivations are related to above average participation levels. However, contrary to prediction, I find no significant relationship between intrinsic motivations and participation levels. Also contrary to prediction, use value motivations are associated with below average contribution levels suggesting that use value motivations may not be capable of sustaining above average levels of contribution. A third finding suggests that the Apache meritocracy is operating effectively and that promotions within the community are based on above average levels of contribution. Lastly, I found that past performance ratings enhance status motivations but not others. In this sense, status motivations can be viewed as an incentive motivation where receiving a reward increases motivation for an individual to perform previously rewarded activities. The lack of association between past performance ratings and intrinsic or use value motivations suggests that increases in rank within the Apache community is neither viewed as controlling (and thus diminishing intrinsic motivation) nor as informative on competence (and thus enhancing intrinsic motivation).

Integration of Findings

As one would expect, there is no single motivation to participate underlying the open source phenomenon. Instead, there are multiple salient motivations for participation that are interrelated in complex ways. Furthermore, these motivations are not mutually exclusive – a common assumption implicit in the early models and analysis of open source motivations. This dissertation provides empirical evidence revealing the dimensions of open source motivations and that motivations to participate in open source projects are functional in nature. That motivations are functional (i.e., serve to meet a particular psychological need) suggests that projects may be structured in some way to provide the incentives that appeal to these needs in order to attract and retain participants. There is also evidence that motivations may be used to segment the developer population and that these segments give rise to different patterns of participation. For example, group analysis of the conjoint data in Study 1 revealed the motivational profiles vary according to contribution level. Likewise, Study 3 demonstrates that different motivations *cause* different results or outcomes. It follows that identification of motivationally-based developer segments that are associated with desirable project outcomes should be a significant factor in the design and operation of an open source project.

This dissertation also demonstrates that motivations to participate have different relative importance and that participation behavior is related to motivations in important ways. Awareness of the range of participation motives coupled with information linking motives to types of behavior represents a significant step in understanding the functioning of open source projects. As previously noted, motivations are not mutually exclusive and thus one or more motivations may be more or less important for any given contribution. Where possible, open source project structures and processes should actively support contributions resulting from any of the motivational dimensions identified by this research. However, the use-value and status seeking motivations may deserve special consideration.

The evidence suggests that the use-value motivation may occupy a unique position in functioning of an open source project. Study 1 revealed the average relative importance of the use-value motivation while Study 3 revealed that use-value motivations are associated with a below average number of contributions. These results suggest that use-value plays a critical role in motivating initial or early contributions but may not sustain long term project affiliation. Survey data reflective of underlying motivation orientation is cross-sectional; thus, I can not say with confidence that motivations follow a predictable evolutionary path. That is, for example, whether use-value motivations tend to give way over time to reputation motivations or vise-versa. However, the findings here and elsewhere (e.g., Shah 2003; Lerner and Tirole 2002) are suggestive that use-value is a likely forerunner of future, and more durable, motivational orientations and should be encouraged and actively supported.

All three studies reveal the significance of reputation or status seeking motivations. Reputation based motivations were shown to be of high relative importance and associated with above average levels of contribution. Study 3 demonstrates an effective structural cycle of status motivations to contribution, contribution to performance, and performance back to increased levels of status motivation. Status motivations are also associated with above average levels of intrinsic motivation. The relationship between status seeking and intrinsic motivation may play a critical role in project outcomes. Increasing intrinsic motivation has been a goal of organizational and management theorists who have long advocated for management styles and work designs that promote intrinsic motivation. Studies have repeatedly shown that people whose motivation is intrinsic in nature have more interest, excitement, and confidence which in turn is manifest as behavior associated with improved performance (McGregor 1966; Lawler 1969, 1970; Hackman and Oldham 1980). While not specifically address by this research, the preceding discussion suggests the intrinsic motivation, while not associated with above average levels of contribution, may be linked with other desirable project outcomes such as superior software design or quality. And lastly, Study 2 shows that manifestations of open source status can operate as a credible signal of productive capacity in the IT labor markets. Taken together these studies demonstrate that a loosely-coupled system of developers, project infrastructure (organization) and IT labor markets can work in harmony as a feedback and reward system that functions to benefit of all parties.

Implications for Practice

Open source software is playing an increasingly important role in the corporate IT infrastructure. Many prominent IT vendors have openly supported open source projects and have made it part of their IT strategy. Corporations such as IBM, Sun Microsystems, and Hewlett Packard have embraced open source development as a new IT paradigm. However, open source solutions have also been viewed critically. According to a recent survey of chief technology officers, the open source community has yet to convince many IT executives to adopt the open source paradigm (Yager 2001). An often cited barrier to open source adoption is the uncertainty surrounding issues of product accountability and technical support related to the dependability of volunteers whose motivations are often non-pecuniary and difficult to discern or manage. Gaining an understanding of the motivational structure of contributors is a critical first step in evaluating open source as a viable development model for commercial software engineering endeavors (Lancashire 2001; Strasser 2001) or as a viable supplier of software products and services (Koch 2005). A primary concern, however, is how to align the unstructured and often uncertain initiatives of the open source community with the zeal for milestones and project accountability common in commercial software development. This thesis provides important empirically grounded insights for those interested in understanding the range of open participation motivations, the way in which these motivations are inter-related, and the relationship between motivations and project outcomes. These insights have implications for IT executives as well as for leaders of open source communities.

Motivational orientations may be one of the few means available to "manage" a largely volunteer open source initiative. It has long been known that motivations matter in work outcomes (Lawler 1969, 1970; Hackman and Oldham 1980). In addition, calls-to-action that make use of the functional nature of motivations have been successfully used by volunteer organizations to attract and retain volunteers. Once on board, the ability to of an activity to assuage a volunteer's motivational "need" is associated with long term affiliation (Clary et al. 1998). It follows from the above points that, open source projects as largely volunteer organizations, will benefit from a deeper understanding of developer motivations and how these motivations are related to project participation.

This thesis has shown that different motivations differentially influence participation. The use-value motivation was confirmed as an important motivation for attracting participants; however, use-value is not associated with long term project commitment. The fact that contributors are likely to have multiple salient motivations suggests that projects may be able greatly improve the "conversion rate" of those initially motivated by use-value to a more long-term or durable motivational affiliation. For example, early contact between use-value contributors and the project should be viewed as opportunity to appeal to potential latent motivations, such as status, via processes and cultural norms intended to appeal to likely potential motivational orientations. Appealing to status seeking motivations would involve actively and visibly recognizing their contribution, etc. Actively appealing to latent motivations early in the contribution relationship should result, all else equal, in higher retention rates of use-value contributors possessing other salient latent motivations.

Participants motivated by status, either outside (career concerns) or inside (ego) the project, are an important constituency as they exhibit above average commitment to the project and also have higher levels of intrinsic motivation – a proven positive for work related outcomes. Thus, firms who are involved in or depend on open source communities where a non-employment relationship is the norm may rely on rewards that directly appeal to the status of participants to influence participation and performance. Such recognition is also important in open source communities where employment relationships exist. The research presented here suggests that employers do recognize Apache's merit based ranking as a reliable proxy for desirable, but imperfectly observably productive abilities. Thus projects courting volunteers having status or career concern motivations should consider the implementation of formal credentialing or reward systems reflective of a performance-based meritocracy. Open source communities may want to nurture such status/career motivations, perhaps by devoting distinct website space to recognize distinguished developers or by promoting involvement in open source communities as leverage in the labor market.

Lastly, there is a question as to the long-term stability of open source communities with a mix of volunteer and paid developers. One risk to the project is that developers who may work full time on the project may yield a disproportionate influence over the project's goals and technical direction. However, Study 3 reveals that being paid to contribute is associated with a sustained volume of contributions at an above average level. Concerns regarding crowding out effects of intrinsic motivation notwithstanding, I find that extrinsic incentives do not necessarily compromise intrinsic enjoyment of contributing to OSS projects. This bodes well for many successful OSS projects (e.g., Apache, Linux, Mozilla, etc.) experiencing increased attention from leading software producers whose paid employees contribute to these projects. Even so, the results of this thesis do not address the influence (positive or negative) that the extrinsically motivated developers may have in the long-run on the motivations of more intrinsically motivated volunteers. Experimental evidence from psychology suggests however, the perception that others are extrinsically motivated is associated with decreased levels of intrinsic motivation (Wild and Enzle 2002).

Contributions

Open source software plays an increasingly important role in the corporate information technology decisions. The ultimate impact of open source as a new paradigm for either software development or software supply is in its early and formative stages. This thesis contributes to the emerging literature surrounding the issues of motivation to participate in open source projects and the consequences for the project as a result of varying sources of motivation. This thesis is among the first studies to yield rich, detailed data and models exploring the sources and consequences of the motivation of open source developers and has implications for IT executives as well as for leaders of open source communities. All three studies comprising this thesis make use of unique longitudinal panel matching responses from multiple surveys with archival data collected at the level of the individual open source contributor. Study 1 contributes to the growing literature on open source development by providing insight into the underlying motivational profiles of open source participation. An additional and significant contribution of this study has been to establish the relative importance of underlying motivations through the novel application of conjoint analysis in the open source domain. Study 2 contributes in a significant way to the emerging literature surrounding provisioning of open source projects. By exploiting repeated individual observations to exclude participants' time invariant endowments in a fixed-effects model I provide empirical evidence of pecuniary returns to open source participation and thus validation of career concerns motivation.

Lastly, Study 3 makes several important contributions to the open source literature. First, the theoretical model and empirical evaluation increase the understanding of how motivations, participation and performance interrelate in open source projects. While some studies have examined particular aspects of motivation or participation, I am not aware of any study that has examined the system of inter-relationships between motivations, participation and performance. Yet, one must consider the motivational system in order to understand how successfully an open source community is functioning in terms of attracting and sustaining participation. This research also contributes by revealing how the different motivations of open source programmers differentially influence their actual source code contributions. To my knowledge, this is also the first longitudinal study of open source participation, resulting in the first work to incorporate the relationship of past performance as an antecedent to motivations. Understanding this relationship provides insight into the motivational effectiveness of feedback in the open source setting.

Limitations and Future Research

This dissertation has several strengths and limitations. In this work I have drawn upon the extensive literature on motivation in psychology and organizational behavior, labor economics, and the emerging open source literature to develop theoretically driven research questions and hypotheses. All three studies in this dissertation utilize a research design that leverages multiple data sources, matching subjective survey data on motivations with objective measures of participation and performance. This approach helps to triangulate findings and mitigate common method and source biases. In addition, I leverage archival data to capture participation and performance measures over a 6-year period. This longitudinal approach allows one to tease out potential causal relationships among motivations, participation and performance. This work focuses solely on projects within the Apache Software Foundation community of open source projects. While this potentially limits the findings in a strict sense to the Apache OSS community, I believe that the results to be applicable to other successful OSS communities such as Linux, Perl, and Mozilla that share an interesting relationship between the intrinsic and extrinsic motivations of contributors. This approach does offer some persuasive advantages. Focusing on one OSS community enables me to link developers to their actual code contributions, and to link these code contributions to objective performance evaluations by others in the Apache community. This approach enables me to cleanly capture the performance – motivation – participation relationships. In addition, eliciting broad participation within one community rather than limited participation across many communities has facilitated the assembly of a representative sample. Hence, the research design increases the internal validity, external validity and statistical conclusion validity of my results (Campbell & Cook 1979).

My results also have interesting implications for future research. First, this work would benefit from further validation in other open source contexts and methodologies. This research approach is characterized by a deep sampling of the three major Apache projects. The results obtained are therefore strongest in internal and external validity with respect to the Apache projects sampled. While I have not encountered any systemic differences across the three projects that I have covered and have no indication why this should be the case with the other sub-projects, only a similar analysis in other open source projects can provide assurance. In addition, a validation through other methods would strengthen my results. For this work, I have relied on survey instruments to provide data. Although the same individuals provided responses via the survey for independent and dependent variables, I see no reasonable alternative than to obtain the data from the same source. Motivation is a subjective construct. Only the individual involved can report on his or her own motivations (motivations can neither be observed nor inferred). Similarly, only the individual can express their personal utility for a particular motivation to contribute. However, future studies could relate self-reported motivations to objectively observed behaviors (such as documented source code contributions) to mitigate same-source bias. Lastly, future studies relating motivational orientation to other project outcomes would be beneficial. The link between motivation and software quality may be especially important for open source projects as the quality model for open source projects is not well documented or understood.

In summary, the three studies comprising this thesis conducted a rigorous analysis of several issues related to open source developer motivations and performance. This research program was designed to make a significant contribute to the emerging literature surrounding the issues of motivation to participate in open source projects and the consequences for the project as a result of varying sources of motivation. By rigorously developing and testing theories of open source participation, this research seeks to broaden the knowledge regarding the feasibility of the open source model to more widespread commercial adoption.

REFERENCES

Amabile, T.M., Hennessey, B. and Grossman, B. 1986. Social influence on creativity: The effects of contracted-for reward. *Journal of Personality and Social Psychological,* 50(1): 14-23.

Anderson, J.C., and Gerbing, D.W. 1988. Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin,* 103(3): 411-423.

Ang, S., and Beath, C.M. 1993. Hierarchical Elements in Software Contracts. *Journal of Organizational Computing*, 3(3): 329-362.

Ang, S., Slaughter, S.A., and Ng, K.Y. 2002. Human capital and institutional determinants of information technology compensation: Modeling multilevel and cross-level interactions. *Management Science*, 48(11): 1427-1426.

Anonymous. 2003. Computerworld 2002 Salary/Skills Survey. Computerworld, Accessed March, 2003: http://www.computerworld.com/departments/surveys/skills?from=bsm.

Anonymous. 1998. IBM Helps Companies Turn Simple Web Sites into Powerful E-Business Solutions. *New York Business Wire.*

Ante, S.E. 2001. Big Blue's Big Bet on Free Software. *Business Week*, Issue 3761: 78.

Apache Software Foundation. 2001. How the ASF Works. The Apache Software Foundation, WWW document. Accessed February, 2001: http://www.apache.org/foundation/how-it-works.html.

Apache Software Foundation. 2001. Press Release. *Apache Software Foundation.* Accessed October, 2003: http://www.apache.org/foundation/:Apache Software Foundation.

Arellano, M., 1987. Computing Robust Standard Errors for within-Groups Estimators. *Oxford Bulletin of Economics and Statistics*, 49(4): 431-434.

Bagozzi, R.P. and Phillips, L.W. 1982. Representing and testing organizational theories: A holistic construal. *Administrative Science Quarterly*, 27(September): 459-489.

Baltagi, B.H. and Wu, P.X. 1999. Unequally Spaced Panel Data Regressions with Ar(1) Disturbances. *Econometric Theory*, 15(6): 814-823.

Banker, R.D., Davis, G.B., and Slaughter, S.A. 1998. Software Development Practices, Software Complexity, and Software Maintenance Performance: A Field Study. *Management Science*, 44(4): 433-450.

Basili, V.R. and Caldiera, G. 1995. Improve Software Quality by Reusing Knowledge and Experience, *Sloan Management Review*, 37: 55-64.

Becker, G. S. 1962. Investment in human capital: A theoretical analysis. *Journal of Political Economy,* 70(5, Part 2: Investment in Human Beings): 9-49.

Belman, D., and Heywood, J.S. 1991. Sheepskin Effects in the Returns to Education: An Examination of Women and Minorities. *The Review of Economics and Statistics*, 73(4): 720-724.

Benjamin, J.D., Jud, G.D., Roth, K.A. and Winkler, D.T. 2002. Technology and Realtor Income. *Journal of Real Estate Finance and Economics*, 25(1): 51-65.

Benson, P., Dohority, J., Garman, L., Hanson, R., Hochschwender, M., Lebold, C., Rhor, R., and Sullivan, J. 1980. Intrapersonal correlates of non-spontaneous helping behavior. *Journal of Social Psychology,* 110(1): 87-95.

Bentler, P. M. and Bonett, D.G. 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3): 588-606.

Bezroukov, N. 1999. Open Source Software Development as a Special Type of Academic Research. *First Monday* 4(10).

Bhargava, A., Franzini, L. and Narendranathan, W. 1982. Serial Correlation and the Fixed Effects Model. *The Review of Economic Studies*, 49(4): 533-549.

Blau, P. 1964. Exchange and Power in Social Life. New York, Wiley.

Blaug, M. 1976. The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey, *Journal of Economic Literature*, 14(3): 827-855.

Boehm, Barry W., Ellis Horowitz, Ray Madachy, donald Reifer, Bradford K. Clark, Bert Steece, A. Winsor Brown, Sunita Chulani and Chris Abts, *Software Cost Estimation with Cocomo 2*, (Upper Saddle River, NJ:Prentice Hall, 2000).

Boehm, B., Abts, C., Brown, A.W., Chulani, S., Clark, B.K., Horowitz, E., Madachy, R., Reifer. D., and Steece, B. 2000. Software cost estimation with COCOMO II. Upper Saddle River, NJ: Prentice Hall.

Boh, W., Slaughter, S.A. and Espinosa, J.A. 2003. Learning from Experience in Software Development: A Multi-Level Analysis, GSIA Working Paper.

Bollinger, T. 1999. Linux and Open-Source Success: Interview with Eric S. Raymond. *IEEE Software,* 16(1): 85-89.

Bostrom, J. 2005. EBay sets up open-source community, IDG New Service, June 2005. Accessed: July, 2005:
http://www.computerworld.com/developmenttopics/development/story/0,10801,102684,00.html?source=x10

Browne, C. B. 1998. Linux and Decentralized Development. *First Monday* 3(3).

Calder, B. and B. Staw 1975. The self-perception of intrinsic and extrinsic motivation. *Journal of Personality and Social Psychology*, 31(4): 599-605.

Campbell, D. and T. D. Cook 1979. Quasi-experimentation: Design & analysis issues for field settings. Boston, Houghton Mifflin.

Campbell, J.P., and Pritchard, J.P. 1976. Motivation theory in industrial and organizational psychology. In M.D. Dunnette (Ed). Handbook of industrial and organizational psychology, Chicago, IL: Rand McNally

Cattell, R. B. 1988. The Meaning and Strategic Use of Factor Analysis. In J.R. Nesselroade and R.B. Cattell (Eds). Handbook of Multivariate Experimental Psychology. Chicago, IL: Rand McNally.

Chamberlain, G. 1984. Panel Data. In Z. Griliches and M.D. Intriligator, (Eds). Handbook of Econometrics. Amsterdam: North Holland.

Clary, E. G. and Orenstein, L. 1991. The amount and effectiveness of help: The relationship of motives and abilities to helping behavior. *Personality and Social Psychology Bulletin*, 17(1): 58-64.

Clary, E.G., Ridge, R.D., Stukas, A.A., Snyder, M., Copeland, J., Haugen, J. and Miene, P. 1998. Understanding and Assessing the Motivations of Volunteers: A Functional Approach. *Journal of Personality and Social Psychology*, 74(6): 1516-1530.

Clary, E.G. and Snyder, M. 1991. A Functional analysis of altruism and prosocial behavior: The case of volunteerism. In M. Clark (Eds). Review of personality and social psychology. Newbury Park, CA: Sage.

Cohen, S. 2005. No Renegade Group Behind Linux. *Business Week*, May 2005. Accessed October, 2005: http://www.businessweek.com/technology/content/may2005/tc2005056_1017.htm?campaign_id=search.

Day, K.M., and Devlin, R.A., 1998. The Payoff to Work without Pay: Volunteer Work as an Investment in Human Capital, *The Canadian Journal of Economics*, 31(5): 1179-1191.

Deci, E.L. 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychological* 18(1): 105-155.

Deci, E. L. 1975. Intrinsic motivation. New York, NY: Plenum Press.

Deci, E. L., R. Koestner and R. M. Ryan 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin 125(6): 627-668.

Deci, E.L. and Ryan, R.M. 1987. The support of autonomy and the control of behavior journal of personality and social psychology. *Journal of Personality and Social Psychological* 53(6): 1024-1037.

Deci, E.L. and Ryan, R.M. 2000. The what and why of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4): 227-268.

Dempsey, B.J., Weiss, D., Jones, P. and Greenberg, J. 1999. *A quantitative profile of a community of open source Linux developers*. School of Information and Library Science, University of North Carolina at Chapel Hill, Working Paper, Accessed April, 1999: http://www.ibiblio.org/osrt/develpro.html.

Detsky, A.S., Naglie, G., Krahn, M.D., Naimark, D. and Redelmeier, D.A. 1997. Primer on medical decision analysis. Part 1: Getting started. *Medical Decision Making,* 17(2): 123-126.

DiBona, C., Ockman, S. and Stone, M. (Eds). 1999. *Open Sources: Voices from the Open Source Revolution*. Sebastopol, CA: O'Reilly & Associates.

Fielding, R. 1999. Shared Leadership in the Apache Project. Communications of the ACM 42(4): 2.

Fisher, R. and Ackerman, D. 1998. The effects of recognition and group need on volunteerism: A social norm perspective. *Journal of Consumer Research*, 25(3): 262.

Garvey, M. J. 1998. "Linux takes off." *InformationWeek*, 709: pp. 181.

Ghosh, R. A. 1998. Interview with Linus Torvalds: What motivates free software developers? *First Monday* 3(3).

Ghosh, R.A., Glott, R., Krieger, B., and Robles, G. 2002. Free/libre and Open Source Software: Survey and Study. University of Maastricht. The Netherlands, International Institute of Infonomics. Accessed March, 2004: http://www.infonomics.nl/FLOSS/report/index.htm.

Goldberg, M.S. and Warner, J.R. 1987. Military Experience, Civilian Experience, and the Earnings of Veterans. *Journal of Human Resources*, 22(1): 62-81.

Goleman, Daniel, Working with Emotional Intelligence. New York, NY: Bantam Books.

Gorsuch, R. L. 1983. Factor Analysis. 2nd Ed. Hillsdale, NJ: L. Erlbaum Associates.

Green, P. 1984. Hybrid models for conjoint analysis: An expository review. *Journal of Marketing Research*, 21(May): 155.

Green, P. and Krieger, A.M. 1996. Individualized hybrid models for conjoint analysis. *Management Science*,42(6): 850.

Green, P. and Rao, R. 1971. Conjoint Measurement: A New Approach for Quantifying Judgmental Data. *Journal of Marketing Research*, 8(3): 355-363.

Green, P. and Wind, Y. 1975. New Techniques for Measuring Consumer Evaluations of Products and Services. *Harvard Business Review*, July-August.

Greene, W.H. 2003. Econometric Analysis. Upper Saddle River, NJ: Prentice Hall.

Hackman, J.R. and Oldham, G. 1980. Work Redesign. Reading, MA: Addison-Wesley.

Hackman, J.R. and Oldham, G. 1974. The job diagnostic survey: An instrument for the diagnosis of jobs and the evaluation of job redesign projects. Killingworth, CT: Yale University, Technical Report No. 4 Department of Administrative Sciences.

Halloran, T.J., and Scherlis, W.L. 2002. High Quality and Open Source Software Practices. *Meeting Challenges and Surviving Success: 2nd Workshop on Open Source Software Engineering. 24th International Conference on Software Engineering* (ICSE 2002): Orlando, FL.

Halvorsen, R. and Palmquist, R. 1980. The Interpretation of Dummy Variables in Semilogarithmic Equations. *The American Economic Review*, 70(June): 474-475.

Hann, I., Roberts, J.A., Slaughter, S.A. and Fielding, R. 2002. Economic Incentives for Participating in Open Source Software Projects. *Proceedings of the Proceedings of the 22nd International Conference on Information Systems* (ICIS 2002): Barcelona, Spain.

Hanrahan, T. 2005. What's Driving Global Linux Adoption? *ComputerWorld*, July 2005

Harman, H. H. 1967. Modern factor analysis (2nd ed). Chicago, IL: Univ. of Chicago Press.

Hars, A. and Ou, S. 2000. Why is Open Source Software Viable? - A study of intrinsic motivation, personal needs, and future returns. *Proceedings of the Americas Conference on Information Systems 2000* (AMCIS 2000): Long Beach, CA.

Hars, A. and Ou, S. 2001. Working For Free? - Motivations of Participating in Open Source Projects. *Proceedings of the 34th Hawaii International Conference on System Science* (HICS 2001): Hawaii.

Hausman, J.A. 1978. Specification Tests in Econometrics. *Econometrica*, 46(6): 1251-1271.

Hausman, J.A. and Taylor, W.E. 1981. Panel Data and Unobservable Individual Effects, *Econometrica*, 49(6): 1377-1398.

Heckman, J.J. 1979. Sample Selection Bias as Specification Error. *Econometrica*, 47(January): 153-161.

Hennessey, B.A. 2000. Rewards and creativity. In C. Sansone and J. Harackiewicz (Eds). Intrinsic and Extrinsic Motivation: The search for optimal motivation and performance. San Diego, CA: Academic Press: 55-78.

Herche, J. 1994. Measuring social values: A multi-item adaptation to the list of values (MILOV). Marketing Science Institute, Working Paper Report No. 94-101, Cambridge, MA: p: 25.

Hertel, G., Niedner, S., and Herrmann, S. 2003. Motivation of software developers in open source projects: An internet-based survey of contributors to the linux kernel. *Research Policy*, 32(7): 1159.

Hochmuth, P. 2004. Linux looks for new worlds to conquer. *Network World*, 21(3): pp. 47.

Hu, L. and Bentler, P.M. 1997. Selecting cutoff criteria for fit indexes for model evaluation: Conventional criteria versus new alternatives. Univ. of California. Technical Report, Santa Cruz, CA.

Humphrey, W.S. 1995. A Discipline for Software Engineering. Reading, MA: Addison-Wesley.

Jakubson, G. 1991. Estimation and Testing of the Union Wage Effect Using Panel Data. *The Review of Economic Studies*, 58(5): 971-991.

Johns, G. 1996. Organizational Behavior: Understanding and managing life at work (4th Ed). New York, NY: HarperCollins.

Kanfer, R. 1990. Motivation theory and industrial and organizational psychology. In M. D. Dunnette and L. Hough (Eds.), Handbook of industrial and organizational psychology. Volume 1. Theory in industrial and organizational psychology. Palo Alto, CA: Consulting Psychologists Press. pp. 75-170.

Kang, S. and Bishop, J. 1986. Effects of Curriculum on Labor Market Success Immediately after High School. *Journal of Industrial Teacher Education*, 23: 14-29.

Kay, F.M. and Hagan, J. 1995. The Persistent Glass Ceiling: Gendered Inequalities in the Earnings of Lawyers. *British Journal of Sociology*, 46(2): 279-310.

Kemerer, C.F. and Slaughter, S.A.1999. An Empirical Approach to Studying Software Evolution. *IEEE Transactions on Software Engineering*, 25(4): 493-509.

Kirsch, L.J. 1996. The Management of Complex Tasks in Organizations: Controlling the Systems Development Process. *Organization Science*, 7(1): 1-21.

Koening, M. Guptill, B. McNee, B. 2005. IT Trends 2005-2009: An Executive Perspective. Saugatuck Technology, SSR-175: 1-29.

Koch, C. 2005. Who Is Open Source? CIO Magazine, Nov 2005.

Lakhami, K. and von Hippel, E. 2002. How Open Source Software works: Free user-to-user assistence. *MIT Sloan Open Source Project.* Accessed July, 2002: http://opensource.mit.edu/papers/lakhanivonhippelusersupport.pdf.

Lakhami, K., Wolf, B., Bates, J., and DiBona, C. 2002. The Boston Consulting Group Hacker Survey. Boston Consulting Group. Accessed October , 2002: http://www.osdn.com/bcg/bcghackersurvey-0.73.pdf.

Lakhani, K. and Wolf, B. 2005. Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In J. Feller, B. Fitzgerald, S. Hissam and K. Lakhani (Eds). *Perspectives on Free and Open Source Software.* Boston, MA: MIT Press.

Lancashire, D. 2001. The Fading Altruism of Open Source Development, *First Monday*, 6(12): URL: http://firstmonday.org/issues/issue6_12/lancashire/index.html.

Lawler, E.E., III and Hall, D.T. 1970. Relationship of job characteristics to job involvement, satisfaction, and intrinsic motivation. *Journal of Applied Psychology*, 54: 305-312.

Lawler, E.E., III. 1969. Job design and employee motivation. *Personnel Psychology*, 22: 426-435.

Lepper, M.R., Greene, D., and Nisbett , R.E. 1973. Undermining children's intrinsic interest with extrinsic reward - test of overjustification hypothesis. *Journal of Personality and Social Psychological*, 28(1): 129-137.

Lepper, M.R. and Henderlong, J. 2000. Turning 'play' into 'work' and 'work' into 'play': 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone and J. Harackiewicz (Eds). Intrinsic and Extrinsic Motivation: The search for optimal motivation and performance. San Diego, CA: Academic Press: 257-307.

Lerner, J. and Tirole, J. 2002. Some simple economics of open source. *The Journal of Industrial Economics* 50(2): 197.

Levitt, J. 2003. How to Get Hired as an Open-Source Developer. *The Open Enterprise.* Accessed October, 2003: http://www.techweb.com/wire/story/TOE20021202S0001.

Lindenberg, S. 2001. Intrinsic motivation in a new light. *Kyklos.* 54(2/3): 317-342.

Litwin, G.H. and Stringer, R.A., Jr. 1968. Motivation and organizational climate. Boston, MA: Harvard University Press.

Loehlin, J.C. 2004. Latent variable models : An introduction to factor, path, and structural equation analysis (4th Ed). Mahwah, NJ: L. Erlbaum Associates.

Markus, L., Manvile, B. and Agres, C. 2000. What makes a virtual organization work? *MIT Sloan Management Review*, 42(1): 13-26.

Marsh, H.W. 1994. Confirmatory factor analysis models of factorial invariance. A multifaceted approach. *Structural Equation Modeling* 1(1): 5-34.

McGregor, D. 1966. Leadership and motivation. In W.G. Bennis, E.H. Schein, and C. McGregor (Eds). Leadership and motivation essays. Cambridge, MA: MIT Press.

Mears, J. 2003. Apache earning its stripes. *Network World*, 20(21): pp. 23.

Mauss, M. 1967. The gift; forms and functions of exchange in archaic societies. New York: Norton.

Meyer, H. 1987. Predicting supervisory ratings versus promotional progress in test validation studies. *Journal of Applied Psychological*, 72(4): 696-697.

Mincer, J. 1974. Schooling, Experience, and Earnings. New York: Columbia University Press.

Mitchell, T.R. and Daniels, D. 2003. Motivation. In W.C. Borman, D.R. Ilgen, R.J. Klimoski (Eds), I.B. Weiner (Editor-in-Chief). Handbook of Psychology. Volume 12: Industrial and Organizational Psychology. New York, NY: Wiley. pp. 225-254.

Mockus, A., Fielding, R. and Herbsleb, J.D. 2000. A Case Study of Open Source Software Development: The Apache Server. *Proceedings of the 22nd international conference on Software engineering* (ICSE 2000). Limerick Ireland: ACM Press. June 4-11.

Moon, J.Y. and Sproull, L. 2002. Essence of distributed work: The case of the Linux kernel. In P. Hinds and S. Kiesler (Eds). Distributed Work, Cambridge, MA: MIT Press. pp. 381-404.

Murnane, R.J., Willett. J.B. and Boudett, K.P. 1999. Do Male Dropouts Benefit from Obtaining a GED, Postsecondary Education, and Training? *Evaluation Review*, 23(5): 475-504.

Netcraft. 2005. The Netcraft WEB server survey. Bath, England: Netcraft, Inc.. Accessed May 2005: http://news.netcraft.com/archives/web_server_survey.html.

Nunnally, J. and Bernstein, I.H. 1978. Psychometric Theory (3rd Ed). New York, NY: McGraw Hill.

O'Reilly, T. 2000. Open Source: The Model for Collaboration in the Age of the Internet. Proceedings of the Computers, Freedom and Privacy, Toronto, Canada. April 6.

Orlikowski, W.J. 2002. Knowing in Practice: Enacting a Collective Capability in Distributed Organizing, *Organization Science*, 13(3): 249-273.

OSI. The Open Source Definition. The Open Source Initiative. Accessed May, 2002: http://opensource.org/docs/definition_plain.html.

Osterloh, M. and Frey, B. 2000. Motivation, knowledge transfer, and organizational forms. *Organization Science*,11(5): 538-550.

Pfeffer, J. and Konrad, A.M. 1991. The Effects of Individual Power on Earnings. *Work and Occupations*, 18(4): 385-414.

Pittman, T.S., Davey, M.E., Alafat, K.A., Wetherill, K.V. and Kramer, N.A. 1980. Informational versus controlling verbal rewards. *Personality and Social Psychological Bulletin*, 6(2): 228-233.

Raftery, A.E. 1995. Bayesian Model Selection in Social Research. In P.V. Marsden (Ed). Sociological Methodology. Cambridge, MA: Basil Blackwell. pp. 111-163.

Raymond(a), E. 2001. Chapter 2. The Cathedral and the Bazaar. The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. Cambridge, MA: O'Reilly.

Raymond(b), E. 2001. Chapter 3. Homesteading the Noosphere. The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. Cambridge, MA: O'Reilly.

Raymond(c), E. 2001. Chapter 4. The Magic Cauldron. The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. Cambridge, MA: O'Reilly..

Roberts, J.A, Hann, I. and Slaughter, S.A. Forthcoming. Understanding the Motivations, Participation and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects. *Management Science.*

Robillard, P.N. 1999. The Role of Knowledge in Software Development. *Communications of the ACM*, 42(1): 87-92.

Rossi, M.A. 2004. Decoding the free/open source (F/OSS) puzzle - a survey of theoretical and empirical contributions. Siena, Università degli Studi di Siena. Dipartimento Di Economia Politica: pp. 1-41.

Rousseau, D. 1995. Psychological contracts in organizations: understanding written and unwritten agreements. Thousand Oaks, CA: Sage Publications.

Rummel, A. and Feinberg, R. 1988. Cognitive evaluation theory - a meta-analytic review of the literature. *Social Behavior and Personality*, 16(2): 147-164.

Ryan, R.M. 1982. Control and information in the intrapersonal sphere - an extension of cognitive evaluation theory. *Journal of Personality and Social Psychological*, 43(3): 450-461.

Ryan, R.M. and Deci, E.L. 2002. An overview of Self-Determination Theory: An organismic-dialectical perspective. In E. L. Deci and R. M. Ryan (Eds). Handbook of Self-Determination Research. Rochester, NY: The University of Rochester Press. pp. 3-33.

Ryan, R.M. and Connell, J.P. 1989. Perceived locus of causality and internalization - examining reasons for acting in 2 domains. *Journal of Personality and Social Psychological*, 57(5): 749-761.

Ryan, R.M., Mims. V. and Koestner, R. 1983. Relation of reward contingency and interpersonal context to intrinsic motivation - a review and test using cognitive evaluation theory. *Journal of Personality and Social Psychological*, 45(4): 736-750.

Sansone, C. 1986. A question of competence: The effects of competence on task feedback on intrinsic interest. *Journal of Personality and Social Psychological*, 51(5): 918-931.

Sansone, C. and Smith, J.L. 2000. Interest and self-regulation: The relation between having to and wanting to. In C. Sansone and J. Harackiewicz (Eds). Intrinsic and Extrinsic Motivation: The search for optimal motivation and performance. San Diego, CA: Academic Press: 341-372.

Schreyer, P. and Pilat, D. 2001. Measuring Productivity. OECD Economic Studies, 33(2001/2): URL: http://www.oecd.org/dataoecd/27/42/1959006.pdf

Sethi, V. and King, W. R. 1991. Construct Measurements in Information Systems Research: An Illustration in Strategic Systems. *Decision Sciences*, 22(3): 455-472.

Settoon, R., Bennett, N. and Liden, R. 1996. Social Exchange in Organizations: Perceived Organizational Support, Leader-Member Exchange, and Employee Reciprocity. *Journal of Applied Psychology*, 81(3): 21-227.

Shah, S., 2003. Understanding the Nature of Participation and Coordination in Open and Gated Source Software Development Communities, MIT Sloan School of Management Working paper: Cambridge, MA.

Shah, Journal and Kruglanski, A. 2000. The structure and substance of intrinsic motivation. In C. Sansone and J. Harackiewicz (Eds). Intrinsic and Extrinsic Motivation: The search for optimal motivation and performance. San Diego, CA: Academic Press. pp. 106-123.

Shah, S. 2004. Understanding the nature of participation and coordination in open and gated source software development communities. In D.H. Nagao (Ed). *Proceedings of the Sixty-third Annual Meeting of the Academy of Management* (CD), ISSN 1543-8643.

Sharma, S., Durvasula, S. and Dillon, W.R. 1989. Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data. *Journal of Marketing Research*, 26(2): 214-221.

Skinner, B. F. 1953. Science and human behavior. New York, NY: Macmillan.

Slaughter, S.A., Harter, D.E. and Krishnan, M.S. 1998. Evaluating the Cost of Software Quality. *Communications of the ACM*, 41(8): 67-73.

Slaughter, S.A. and Ang, S. 2003. Returns to Human Capital for It Professionals. *Proceedings of the 2003 ACM SIGCPR conference on Computer personnel research*. Philadelphia, PA: ACM Press.

Smith, M.B., Brunder, J.S. and White, R.W. 1956. Opinions and personality. New York, Wiley.

Snyder, M. and DeBono, K.G. 1989. Understanding the functions of attitudes: Lessons from personality and social behavior. In A. R. Pratkanis, S. J. Breckler and A. G. Greenwald(Eds). Attitude structure and function. Hillsdale, NJ: Lawrence Erlbaum. pp. 339-359.

Spence, M. 1973. Job Market Signaling. *Quarterly Journal of Economics*, 87: 355-374.

Spence, M. 1974. Market Signaling: Information Transfer in Hiring and Related Screening Processes. Cambridge, MA: Harvard University Press.

Spence, M. 1976. Competition in salaries, credentials, and signaling prerequisites for jobs. *Quarterly Journal of Economics*, 90(1): 51-74.

Stallman, R. Philosophy of the GNU Project. GNU – Free Software Foundation, Accessed October, 2001: http://www.gnu.org/philosophy/philosophy.html.

Statz, J. 1999. Leverage Your Lessons. *IEEE Software*, 16(2): 30-32.

Staub, E. 1978. Positive social behavior and morality. New York, NY: Academic Press.

Strasser, M. 2001. A New Paradigm in Intellectual Property Law? The Case Against Open Sources. *Stanford Technology Law Review*, 4: 2-70.

Straub, D., Limayen, M. and Karahana-Evaristo, E. 1995. Measuring system usage: Implications for is theory and testing. *Management Science*, 41(8): 1328-1342.

Taubman, P.J. and Wales, T.J. 1973. Higher Education, Mental Ability, and Screening, *Journal of Political Economy*, 81(1): 28-55.

Thurstone, L.L. 1947. Multiple-factor analysis: a development and expansion of the vectors of mind. Chicago, IL: The University of Chicago Press.

Torvalds, L. and Diamond, D. 2001. Just For Fun: The story of an accidental revolutionary. New York, NY: Harper Business.

Tyler, J.H., Murnane, R.J. and Willett, J.B. 2000. Estimating the Labor Market Signaling Value of the GED. *Quarterly Journal of Economics*, 115(2): 431-468.

Verbeek, M. and Nijman, T. 1992. Testing for Selectivity Bias in Panel Data Models. *International Economic Review*, 33(3): 681-703.

von Krogh, G., Spaeth, S. and Lakhami, K.R. 2003. Community, Joining, and Specialization in Open Source Software Innovation: A Case Study. *Research Policy*, 32(7): 1217-1241.

von Hippel, E. 2001. Innovation by user communities: Learning from open-source software. *MIT Sloan Management Review*, 42(4): 82-86.

von Hippel, E. and von Krogh, G. 2003. Open source software and the private-collective innovation model: Issues for organization science. *Organization Science*, 14(2): 209-223.

Watson, S. 2000. End of Job Loyalty. *Computerworld*. pp. 52-53.

Weinberg, G. 1998. The psychology of computer programming. Silver Anniversary Edition. (1st Ed). New York, NY: van Nostrand Reinhold.

Weiss, A. 1995. Human-Capital vs Signaling Explanations of Wages. *Journal of Economic Perspectives*, 9(4): 133-154.

Whang, S. 1992. Contracting for Software Development. *Management Science*, 38(3): 307-324.

Wiersma, U.J. 1992. The effects of extrinsic rewards in intrinsic motivation - a meta-analysis. *Journal of Occupational and Organizational Psychology*, 65(2): 101-114.

Wild, T.C. and Enzle, M.E. 2002. Self-Generated Changes in Intrinsic Motivation as a Function of Social Perception. In E.L. Deci and R.M. Ryan (Eds). Handbook of self-determination research. Rochester, NY: University of Rochester Press. Ch 7.

Wigton, R.S., Hoellerich, V.L. and Patil, K.D. 1986. How Physicians use clinical information in diagnosing pulmonary embolism: an application of conjoint analysis. *Medical Decision Making*, 6(1): 2-11.

Wooldridge, J.M. 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

Yager, T. 2001. Open source takes hold. *InfoWorld.* WWW document. Accessed October, 2002: http://www.infoworld.com/articles/tc/xml/01/08/27/010827tcintro.xml

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX I.
EXPERIENCE AND EMPLOYMENT HISTORY INVENTORY

GSIA Graduate School of Industrial Administration Carnegie Mellon

HELP US UNDERSTAND OPEN SOURCE PARTICIPATION

Faculty and Doctoral Students at Carnegie Mellon University have created this survey to explore the question of open source participation. By participating in this survey, you can share your experiences and improve the understanding of the open source movement.

The survey should take about 15 minutes to complete.

STATEMENT OF CONFIDENTIALITY

The survey will ask you for some demographic information, your position in your company, your compensation, and other facts about your participation in the Apache project. This information will only be used to help us aggregate responses and will **NOT** be used to identify individuals. Only summary statistics for large groups of respondents will be reported.

Carnegie Mellon University's Institutional Review Board (IRB) has reviewed this research project regarding violations of privacy and morally or ethically objectionable procedures or requirements. The IRB has found no objections and given its approval for this research project.

Under **no circumstances** will individual responses be revealed. Individual responses will be kept **completely confidential**.

BENEFIT THE APACHE SOFTWARE FOUNDATAION

In appreciation of your participation in this research, for each survey completed before November 25th, 2001 we will make a contribution of $10 (upto a maximum of $10,000) to the Apache Software Foundation.

Take The Survey

GSIA *Graduate School of Industrial Administration* **Carnegie Mellon**

Open Source Participation

In this section of the survey we are interested in learning more about your contributions to open source or so-called free software projects in general and the Apache family of projects in particular.

While contributions to open source may take many important forms, for the purposes of this survey, we define an open source "contribution" as participation in the development of contributed source code. In other words, a patch.

Q:1	Using the definition of contribution above, have you ever made a contribution to any Apache Software Foundation (ASF) project?	☐ Yes ☐ No

If "NO" Skip to Question 18.

Q:2	Below we have listed some milestone events that appear to occur in the "career" of an Apache developer. Please indicate which of these levels of recognition that you have achieved and if achieved, approximately when this event occurred.

[illegible]	[illegible]	[illegible]
Submitted first patch to an ASF project.	☐ Yes ☐ No	- MM- -YYYY-
Granted commit access to any ASF project CVS tree.	☐ Yes ☐ No	- MM- -YYYY-
Awarded Project Committee Membership.	☐ Yes ☐ No	- MM- -YYYY-
Awarded ASF membership.	☐ Yes ☐ No	- MM- -YYYY-

Q:3	For this question, we have attempted to isolate a few contributions (patches) from Apache CVS archives that have been attributed to your efforts. For each contribution, please confirm your involvement. For each confirmed contribution, please (1) provide an estimate of the total number of hours that <u>you</u> devoted to the actual coding or development of the patch, (2) indicate the nature of the patch, (3) indicate your role in the development of the patch and (4) indicate the number people, other than yourself, directly involved in the development of the patch.

MADE CONTRIBUTION	[illegible]

☐ Yes ☐ No	**Log Message:**
	`Add getCustomPathInfo() which has code that seems to work for Tomcat.`
	
	Affected Files:
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/util/ServletUtil.java,v`
	<u><ViewComplete Listing></u>

If "YES", please answer the following questions.

Estimate of the total number of hours that <u>you personally</u> devoted to the actual coding or development of the contribution listed above.	☐ Hours
Which of the following best describes the purpose or reason for the contribution or patch? Select all that apply.	☐ Fix Bug ☐ Enhance Existing Function ☐ Add New Functionality <Other - Type Over>
Please indicate your role in the development of the patch. Select all that apply.	☐ Primary Developer ☐ Second Developer ☐ Committed Changes ☐ Architect/Designer ☐ Code Reviewer <Other - Type Over>
How many people, not including you, were directly involved in the development of the contribution or patch.	- Number of People -

MADE CONTRIBUTION	[illegible]

☐ Yes ☐ No	**Log Message:**
	`no message`
	
	Affected Files:
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/util/encoding/Attic/CharacterEncoder.java,v`
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/util/encoding/Attic/CharacterDecoder.java,v`
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/util/encoding/Attic/CEStreamExhausted.java,`
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/util/encoding/Attic/CEFormatException.java,`
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/util/encoding/BASE64Enc`
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/util/encoding/BASE64Dec`
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/util/base64/Base64.java`
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/util/StreamUtil.java,v`
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/net/http/object/impl/debug/DebugHTTPObjectS`
	`/home/cvspublic/xml-xang/java/src/org/apache/xang/net/http/object/impl/auth/http/HttpHandler.`
	<ViewComplete Listing>

If "YES", please answer the following questions.

Estimate of the total number of hours that you personally devoted to the actual coding or development of the contribution listed above.	Hours	
Which of the following best describes the purpose or reason for the contribution or patch? Select all that apply.	☐ Fix Bug ☐ Enhance Existing Function ☐ Add New Functionality <Other - Type Over>	
Please indicate your role in the development of the patch. Select all that apply.	☐ Primary Developer ☐ Second Developer ☐ Committed Changes	☐ Architect/Designer ☐ Code Reviewer <Other - Type Over>
How many people, not including you, were directly involved in the development of the contribution or patch.	- Number of People -	

[NOTE: Two additional contribution questions similar Q:3 were redacted]

Q:4	For each of the Open Source projects listed below, please indicate those projects to which you have made contributions. For those projects selected, please indicate (1) approximate # of contributions (patches) that you have made. (2) the average amount of time that you devoted to your patches and (3) the approximate date of your last contribution. Please exclude Apache projects from consideration when answering this question.

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
GNOME	☐ Yes ☐ No	- # Patches -	--- Hours ---	- MM-	-YYYY-
KDE	☐ Yes ☐ No	- # Patches -	--- Hours ---	- MM-	-YYYY-
Linux Kernel	☐ Yes ☐ No	- # Patches -	--- Hours ---	- MM-	-YYYY-
Mozilla	☐ Yes ☐ No	- # Patches -	--- Hours ---	- MM-	-YYYY-
Open Office	☐ Yes ☐ No	- # Patches -	--- Hours ---	- MM-	-YYYY-
Perl - CPAN	☐ Yes ☐ No	- # Patches -	--- Hours ---	- MM-	-YYYY-
Samba	☐ Yes ☐ No	- # Patches -	--- Hours ---	- MM-	-YYYY-
<Other - Type Over>		- # Patches -	--- Hours ---	- MM-	-YYYY-
<Other - Type Over>		- # Patches -	--- Hours ---	- MM-	-YYYY-
<Other - Type Over>		- # Patches -	--- Hours ---	- MM-	-YYYY-

Q:5	People volunteer, or otherwise do unpaid work, for many different reasons. They also derive different kinds of satisfaction from such work. Following are some reasons other developers have given us regarding their participation in open source projects. Thinking of your own participation, please indicate how important each reason is to you personally.

	<================================>				
I meet people and make friends through it.					
It gives me the chance to attain a recognized qualification or skill.					
It gives me status at work.					
It is the satisfaction of seeing the results.					
It gives me the chance to do things I am good at.					
Political belief in "free software".					
I really enjoy it. It is fun.					
It broadens my technical experience.					
It gives me a sense of personal achievement.					
It increases my opportunities for a better job.					
It gives me status in the Apache community.					
My employer values my participation.					
My employer <u>requires</u> my participation.					
I fixed a bug or problem that was causing me trouble.					
I added a feature I wanted or needed to use.					
	<================================>				

Education

Q:6	What is the highest level of education you have completed?	--- Please Select One ---

If you have not yet earned your bachelors degree, please <u>Skip to Question 9</u>.

Q:7	What was your baccalaureate discipline or major?	--- Please Select One ---
Q:8	If you have a graduate, masters, Ph.D. or professional degree, what was your area of specialty?	--- Please Select One ---

General Information

In this section of the survey we are interesting in learning more about your general demographics.

Please note, there are those who may feel that some of the following questions are of a sensitive nature. While we understand this feeling, the answers are an important part in understanding participation in open source software projects. The questions are not intended to offend and <u>all responses will be held in the strictest of confidence.</u>

Q:9	What was your age on January 1, 2001	--- Please Select One ---	
Q:10	What is your gender?	☐ Male ☐ Female	
Q:11	Which best describes your race? Select all that apply.	☐ White, Caucasian ☐ Black,African Am. ☐ Hispanic	☐ Asian ☐ Other <Other - Type Over>

Experience & Job History

In this section of the survey we are interesting in learning more about your work experience and earnings history. Please answer the following questions relating to your regular paid job. For the purposes of this survey, we consider your regular paid job to be your primary income producing activity during the period in question.

Q:12	How much work experience do you have in the general area of technology, technology development, or similar since beginning to work full time as an adult? In other words, technology related experience since you entered the work force.	- Yrs - / - Mths -
Q:13	How much total work experience do you have since beginning to work full time as an adult? In other words, total work experience since you entered the work force.	- Yrs - / - Mths -

Q:14 Please indicate whether or not you have experience with each of the following technologies. If so, how much experience do you have?

TECHNOLOGY	EXPERIENCE	AMOUNT OF EXPERIENCE
Active X	☐ Yes ☐ No	
ASP	☐ Yes ☐ No	
C++ and C	☐ Yes ☐ No	
CASE Tools, Any	☐ Yes ☐ No	
COBOL	☐ Yes ☐ No	
COM / COM+	☐ Yes ☐ No	
CORBA Development	☐ Yes ☐ No	
HTML	☐ Yes ☐ No	

Java	☐ Yes ☐ No	
Linux	☐ Yes ☐ No	
.Net	☐ Yes ☐ No	
Network Security	☐ Yes ☐ No	
OOA / OOD	☐ Yes ☐ No	
Oracle	☐ Yes ☐ No	
Other RDBMS	☐ Yes ☐ No	
Other Unix	☐ Yes ☐ No	
Perl (Core or CPAN)	☐ Yes ☐ No	
Python	☐ Yes ☐ No	
TCP/IP	☐ Yes ☐ No	
Visual Basic	☐ Yes ☐ No	
WAP / WML	☐ Yes ☐ No	
Win32 / NT Development	☐ Yes ☐ No	
XML / XSLT	☐ Yes ☐ No	

Please note, there are those who may feel that the following questions are of a sensitive nature. While we understand this feeling, the answers are an important part in understanding participation in open source software projects. The questions are not intended to offend and <u>all responses will be held in the strictest of confidence.</u>

Q:15	Please describe your employment and compensation history for the calendar years listed.

COMPENSATION AND EQUITY HOLDINGS

If, for any reason, you worked for more than one company at any time during the period in question, please report "Salary" as the sum of wages for all regular paid jobs or contracts held during that period.

If you were self-employed during the periods in question: interpret the "Salary" as total wages resulting from your primary income producing activity.

	2000 CALENDAR YEAR	[illegible] CALENDAR YEAR	[illegible] CALENDAR YEAR
Please specify the currency to be used for all compensation figures during this period.	United States, Dollars	United States, Dollars	United States, Dollars
Annual Salary *(Exclude Bonus, Incentive and/or Commission)*	- Wages in USD - (In Stated "**Currency**")	- Wages in USD - (In Stated "**Currency**")	- Wages in USD - (In Stated "**Currency**")
Total Annual Compensation *(Include <u>Annual Salary</u>, Bonus, Incentive and/or Commission)*	- Wages in USD - (In Stated "**Currency**")	- Wages in USD - (In Stated "**Currency**")	- Wages in USD - (In Stated "**Currency**")
Did your wages increase during this year for any reason?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No

If "YES", which of following best describes the reasons for the increase. Please estimate the approximate % of your Total Annual Compensation represented by the increase.	-Primary Reason - -Secondary Reason - <Other Reasons -% Increase - (Using "**Currency**")	-Primary Reason - -Secondary Reason - <Other Reasons -% Increase - (Using "**Currency**")	-Primary Reason - -Secondary Reason - <Other Reasons -% Increase - (Using "**Currency**")
Did you receive or hold any stock options in your company during the period in question?	☐ Yes ☐ No If "NO", skip remaining questions in this column. Proceed to column for 1999	☐ Yes ☐ No If "NO", skip remaining questions in this column. Proceed to column for 1998	☐ Yes ☐ No If "NO", skip remaining questions in this column. Skip to "Employment Information"
What % of total outstanding shares in your company did your options represent ?	-% Options-	-% Options-	-% Options-

EMPLOYMENT INFORMATION

If, for any reason, you worked for more than one company at any time during the period in question, please answer the following questions by referring to the position and the firm where you spent the majority of the hours worked that year.

If you were unemployed or a full time student with less than $15,000 of annual wages during the period in question, please put "Unemployed" or "Student" as the Job Title.

If you were unemployed or a full time student with more than $15,000 of annual wages during the period in question, please answer the following questions by referring to the position and the firm where you spent the majority of the hours worked that year.

If you were self-employed during the periods in question: please put "Self Employed" as the Job Title

	[illegible]	[illegible]	[illegible]
Job Title or Job Name			
On average, how many hours a week did you work for pay during this year?	--- Hours ---	--- Hours ---	--- Hours ---
Is this a public or private firm?	--- Select One ----	--- Select One ----	--- Select One ----
Firm Size in People	--- Select One ----	--- Select One ----	--- Select One ----
Firm Industry	--- Select One ----	--- Select One ----	--- Select One ----
City Where Employed			
Country Where Employed	United States	United States	United States
Did you leave your job for another job for any reason during this period? If YES, what was the primary reason.	☐ Yes ☐ No - Reason - <Other - Type Over>	☐ Yes ☐ No - Reason - <Other - Type Over>	☐ Yes ☐ No - Reason - <Other - Type Over>
At any time during this year were you considered a full time student? If YES, what grade or level were you	☐ Yes ☐ No - Level or Grade -	☐ Yes ☐ No - Level or Grade -	☐ Yes ☐ No - Level or Grade -

attending?			
Did you receive any service awards or other forms of recognition for your achievements at your regular paid job during this time period? Check all that apply.	☐ Promotion ☐ Formal Achievement Award ☐ Mgmt Recognition or Award ☐ Merit Pay Raise ☐ Performance Bonus ☐ Gain Sharing Distribution ☐ Profit Sharing Distribution <Other - Type Over>	☐ Promotion ☐ Formal Achievement Award ☐ Mgmt Recognition or Award ☐ Merit Pay Raise ☐ Performance Bonus ☐ Gain Sharing Distribution ☐ Profit Sharing Distribution <Other - Type Over>	☐ Promotion ☐ Formal Achievement Award ☐ Mgmt Recognition or Award ☐ Merit Pay Raise ☐ Performance Bonus ☐ Gain Sharing Distribution ☐ Profit Sharing Distribution <Other - Type Over>
At any time during this year, were you <u>paid</u> to participate in Apache projects?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
If "YES", approximately how many hours a week were you paid for participating in Apache projects?	--- Hours ---	--- Hours ---	--- Hours ---

Q:16	Since you first began contributing to Apache projects, how many times have you moved or switched, for any reason, from one regular paid job to another? For the purposes of this survey we define a "job switch" to be a change in employers. This does not include internal transfers or promotions within the same company or parent company.	- # Switches -

If "ZERO" <u>Skip to Question 18</u>.

Q:17	Please answer the following question by recalling the circumstances surrounding your three (3) most recent job switches. If you have fewer than 3 job switches since you began contributing to Apache projects, complete only as many columns as required starting with the column labeled "Most Recent Job Switch".

	[illegible]	[illegible]	[illegible]
Approximate month and year you switched jobs.	- MM- -YYYY-	- MM- -YYYY-	- MM- -YYYY-
Approximately how many <u>different</u> recruiting firms contacted you regarding available jobs in the 90 days prior to the actual job switch.	--- Please Select One ---	--- Please Select One ---	--- Please Select One ---
Approximately how many <u>different</u> companies did you interview with before before you made your job selection.	--- Please Select One ---	--- Please Select One ---	--- Please Select One ---
Approximately how many <u>different</u> companies extended you a job offer before you made your job selection.	--- Please Select One ---	--- Please Select One ---	--- Please Select One ---

Odds and Ends

Please indicate if you wish to receive a summary of the results from this research and whether you would be interested in participating in future open source research.

Q:18	"I would like to receive an email summary of the research findings."	☐ Yes ☐ No
Q:19	"I would be interested in participating in future open source research."	☐ Yes ☐ No
If "YES" to either of the above, please indicate your preferred email address.		

Before you submit your results, please take a minute to provide us feedback on this survey.

Q:20 How difficult was this survey to complete?

Very Easy	Easy	Acceptable	Difficult	Very Difficult
☐	☐	☐	☐	☐

Q:21	Approximately how long did it take you to complete the survey?	-Please Select One-
Q:22	When completing an on-line survey, which of the following GUI elements do you prefer to use when providing answers?	-Please Select One-
Q:23	Please use the comments box below to let us know what it was like for you to take this survey. For example: How was the length? Were the questions straight forward or otherwise clearly worded? Etc.	

Comments:

APPENDIX II-A.
MODIFIED VOLUNTEER FUNCTIONS INVENTORY

GSIA Graduate School of Industrial Administration Carnegie Mellon

Apache Survey - Part One

Thank you for taking the time to help us understand of how open source projects attract and retain high quality development talent.

This survey consists of three parts. The first part of the survey explores potential motivations for open source participation. The second part of the survey is designed to help us understand the relative importance of different aspects of open source participation. The last part of the survey requests update information regarding your recent work history and experience.

Specific instructions are given at the start of each section. **Please read the instructions carefully.** Your individual answers will be kept completely confidential. Please respond to every question.

Motivations

On this page, you will find statements regarding possible motivations for participation in open source software projects. Specific instructions are given at the start of the following section. **Please read the instructions carefully.** Your individual answers will be kept completely confidential. Please respond to every question.

Motivations - Section One

Thinking of your own open source participation, please indicate how accurately each statement describes <u>your</u> motivations for open source participation. Choose a number to the right of each statement based on the following scale:

How accurate is each statement in describing your motivations for open source participation?

1	2	3	4	5	6	7
Very Inaccurate	Mostly Inaccurate	Slightly Inaccurate	Uncertain	Slightly Accurate	Mostly Accurate	Very Accurate

	1	2	3	4	5	6	7
Contributing to an open source project can help me get my foot in the door at a place where I would like to work.	☐	☐	☐	☐	☐	☐	☐
My employer requires my participation.	☐	☐	☐	☐	☐	☐	☐

Contributing to open source projects gives me the chance to do things I am good at.	☐	☐	☐	☐	☐	☐	☐
Making contributions to open source projects gives me a sense of personal achievement.	☐	☐	☐	☐	☐	☐	☐
People I know share an interest in open source or free software.	☐	☐	☐	☐	☐	☐	☐
Others in this project are aware of my technical skills.	☐	☐	☐	☐	☐	☐	☐
I am genuinely concerned about or believe in open source or "free" software.	☐	☐	☐	☐	☐	☐	☐
I am motivated when my technical achievements are recognized by others in this project.	☐	☐	☐	☐	☐	☐	☐
By making contributions to an open source project, I can make new contacts that might help my business or my career.	☐	☐	☐	☐	☐	☐	☐
Developing software is fun, I enjoy it.	☐	☐	☐	☐	☐	☐	☐
I make contributions to open source projects because of the satisfaction I receive from seeing the results.	☐	☐	☐	☐	☐	☐	☐
My patches are for my or my company's benefit. If it helps others that's just a bonus.	☐	☐	☐	☐	☐	☐	☐
If a bug isn't affecting me or my work, I'm not likely to fix it.	☐	☐	☐	☐	☐	☐	☐
Others with whom I am close place a high value on making contributions to open source projects.	☐	☐	☐	☐	☐	☐	☐
Writing software is a hobby of mine.	☐	☐	☐	☐	☐	☐	☐
Making contributions to an open source project lets me learn things through direct, hands on experience.	☐	☐	☐	☐	☐	☐	☐
I feel it is important to help others.	☐	☐	☐	☐	☐	☐	☐
The main reason I contribute to this open source project is to make my paid job easier.	☐	☐	☐	☐	☐	☐	☐
Open source projects give me a chance to showcase my technical skills.	☐	☐	☐	☐	☐	☐	☐
I can do something for a cause that is important to me.	☐	☐	☐	☐	☐	☐	☐
Making contributions to an open source project is an important activity to the people I know best.	☐	☐	☐	☐	☐	☐	☐
People who know me, know that I like to write software.	☐	☐	☐	☐	☐	☐	☐
I can learn how to deal with a variety of people and personalities.	☐	☐	☐	☐	☐	☐	☐
Open source experience will look good on my resume.	☐	☐	☐	☐	☐	☐	☐
I can explore my own strengths within an open source project.	☐	☐	☐	☐	☐	☐	☐
My employer encourages my participation.	☐	☐	☐	☐	☐	☐	☐

APPENDIX II-B.
CONJOINT EXPERIMENT

GSIA Graduate School of Industrial Administration

Carnegie Mellon

Apache Survey - Part Two

In this part of the survey, you will be asked to suspend belief for a brief moment and imagine that you are about to begin your Apache career. This part of the survey has relatively few questions but requires some careful thought. Specific instructions are given at the start of each section. **Please read both the introduction section and the section instructions carefully.** Please respond to every question.

Open Source Participation Tradeoffs

Tradeoffs - Introduction

For this segment of the survey we ask you to imagine that you are a developer who has just decided to play an active role in one of the Apache projects. Over the course of the next 12 months you expect to submit several patches to that project. As you know, contributing to an open source project has both costs and benefits. Costs include the time and effort to be an active project participant. Benefits include feelings of accomplishment, developing expertise in a particular technology or just plain old fun. Your 12-month participation experience on this project can vary across several dimensions described below.

1st Dimension: Rank within ASF

Contributions you submit are reviewed by your peers for appropriateness of the solution and design. Periodically, your service to the project is evaluated by a group of highly respected Apache project participants. Accumulating sufficient positive evaluations during the year may increase your status within the project. A participant's status can increase from *Developer* - someone without write access to the source tree to *Committer* - someone with write access to the project source tree. In some cases, very involved developers have even climbed to the status *ASF member* within one year. At year's end there are three possible outcomes. If you fail to accumulate sufficient positive evaluations to advance your status, you may remain a *Developer*. On the other hand, if your efforts have been well received, you may have been promoted to *Committer* or even *ASF Member*.

To summarize, at the end of your first year of contribution, your rank within the ASF will have one of the following levels:

- **Developer** - Someone who has not been granted write access to the CVS source tree.
- **Committer** - Someone who has been granted write access to the CVS source tree.

- **ASF Member** - Someone who has been invited to become a member of the Apache Software Foundation

2nd Dimension: Type of Contribution

There are many ways to describe open source contributions. We limit our interest here to a simple classification of contributions produced and submitted. For our purposes, your contributions are characterized either as *add new functionality*, *enhance existing functionality* or *fix a bug*.

Adding new functionality puts you in charge of all aspects of a typical software development project. Thus, you are responsible for the architecture, design, implementation and testing of your contribution.

Enhancing existing functionality involves making a substantial change to a component or existing feature of the software. With this type of contribution you are responsible for the implementation and testing of your code but the original architecture and overall design of the component remain unchanged.

Fixing a bug involves making a change to an component for the purpose of removing a fault condition present in the software.

To summarize, at the end of your first year of contribution your entire portfolio of contributions will be characterized as belonging predominantly to one of the following levels:

- **Adding New Functionality**
- **Enhancing Existing Functionality**
- **Fixing Software Bugs**

3rd Dimension: Use Value

One purpose for contributing to an open source project is to enhance the usefulness of the product, in some way, to you personally. That is to increase your, so-called, use value of the product. One way to measure your increase in use value is to estimate the time that you might expect to save as a result of the contribution.

Typical time savings resulting from a contribution might include increased efficiency in the performance of your job or less downtime. For our scenarios, we ask you to think of the use value of all your contributions as the time you expect to save in the performance of your paid job. For example, you might submit a patch that resolves a race condition responsible for your server crashing at least once a month. Suppose over the entire year, you estimate that your contributions could save you 20 minutes per week, you will save approximately 17 hours per year. If your contributions save you 40 minutes per week, you will save approximately 34 hours per year. If your contributions save you 60 minutes per week, you will save approximately 51 hours per year.

To summarize, at the end of your first year of contribution, the use value resulting from your contributions will have one of the following levels:

- **17 hrs/yr (20 min/week)**
- **34 hrs/yr (40 min/week)**
- **51 hrs/yr (60 min/week)**

4th Dimension: Career Impact

Participation in the Apache project is sometimes rewarded by current and future employers. Employers, who rely on Apache software in their day-to-day operations, may value and encourage participation in the Apache project. Other employers, who rely on your programming abilities in other areas, may welcome your efforts to hone your technical skills. Lastly, some employers have no regard for your participation in open source projects.

Assume the following conditions exist. Employers who value your Apache participation have rewarded participants with an additional 3% salary increase. Assuming your current salary is 70,000 USD, you could expect to receive an additional increase in your annual compensation of 2,100 USD entirely attributable to your Apache participation.

Likewise, employers who welcome your efforts to hone your technical skills through participation in an Apache project increase your salary by an additional 1%. Assuming your current salary is 70,000 USD, you could expect to receive an additional increase in your annual compensation of 700 USD entirely attributable to your Apache participation.

Lastly, you would expect to receive NO additional increase in your annual compensation from employers who don't value open source participation.

To summarize, at the end of your first year of contribution, the career impact resulting from your contributions will have one of the following levels:

- **Expected gain of 2,100 USD**
- **Expected gain of 700 USD**
- **Expected gain of 0 USD**

5th Dimension: Effort

Participation in an open source project requires that you devote your personal time to develop and test your code. As mentioned before, you anticipate making several contributions in the following 12 months. By your estimates, you may spend 100, 35, or 15 total hours in the following year on contributions to the project.

To summarize, at the end of your first year of contribution, the amount of effort that you put into your contributions will have one of the following levels:

- **100 total hrs/yr**
- **35 total hrs/yr**
- **15 total hrs/yr**

6th Dimension: Recreational Value

An often cited benefit of open source participation is its value as a form of recreation. Some developers derive a sense of enjoyment from the creative process required to develop their contributions. For our purpose, we classify the enjoyment derived from the contribution experience in terms of the recreational value you experience as a result of your contribution. The recreational value of your participation will have one of three values. *Very enjoyable* - where there is a very high degree of enjoyment, *Somewhat enjoyable* - where there is moderate enjoyment, and *Not enjoyable* - where there is a little or no enjoyment derived from the process.

To summarize, at the end of your first year of contribution, the recreational value of your contribution experience can be classified as follows:

- **Very enjoyable** - a high recreational value results from the contribution experience
- **Somewhat enjoyable** - a moderate recreational value results from the contribution experience
- **Not enjoyable** - little or no recreational value results from the contribution experience

7th Dimension: Social / Community

Open source participants, like those in other virtual communities, may develop social (non-task related) relationships or experience a shared sense of purpose or community as a result of their interactions with other contributors. While the precise nature of the social experience is, of course, unique to each participant, we classify the experience in three ways: *Very social* - where there is a high degree of social communication and interaction, *Somewhat social* - where there is a moderate degree of social communication and interaction, and *Not social* - where there is a little or no social communication or interaction.

To summarize, at the end of your first year of contribution, your contribution experience can be classified as follows:

- **Very social** - having a high degree of social communication and interaction
- **Somewhat social** - having a moderate degree of social communication and interaction
- **Not social** - little or no social communication or interaction

Tradeoffs - Section One

Each of the 7 dimensions of open source participation outlined above may play a role in your decision to participate in an open source project. In this part of the survey we would like to measure how important each of the 7 previously discussed dimensions are to you personally.

To express the level of importance of each dimension, please allocate 100 points across the 7 dimensions so as to best reflect their relative importance to you. You may assign points in any way you wish (including zero points to one or more dimensions). Please ensure that your points total to 100.

Reputation - Your status, reputation or rank within Apache Software Foundation.	0
Contribution Type - The type of contribution you make.	0
Use Value - The potential value or usefulness you personally derive from your contributions.	0
Career - The impact on your career or earnings potential.	0
Effort - The effort or time required to participate.	0
Recreational Value - The recreational value you experience from participation.	0
Social - Your perception of Apache as a place to develop social relationships.	0
Allocate 100 Total Points	

Tradeoffs - Section Two

Each of the 7 dimensions of open source participation outlined above may play a role in your decision to participate in an open source project. Recall that each dimension also has 3 levels or sub-dimensions.

Please rate the desirability of each level for each dimension to you personally, <u>assuming that the level will result from your open source participation experience.</u> Choose a number to the right of each level based on the following scale:

How desirable are each of the following dimension levels in describing your ideal Apache participation experience?

1	2	3	4	5	6	7
Very Undesirable	Undesirable	Slightly Undesirable	Neutral	Slightly Desirable	Mostly Desirable	Very Desirable

Reputation: Your status, reputation or rank within Apache Software Foundation.	1	2	3	4	5	6	7
1. Developer - No write access to the project source tree.	☐	☐	☐	☐	☐	☐	☐
2. Committer - Granted write access to the project source tree.	☐	☐	☐	☐	☐	☐	☐
3. ASF Member - Member of the Apache Software Foundation.	☐	☐	☐	☐	☐	☐	☐

Contribution Type: The type of contribution you make.	1	2	3	4	5	6	7
1. Add new functionality - Lead architecture, design & implementation efforts on a new feature.	☐	☐	☐	☐	☐	☐	☐
2. Enhance existing functionality - Implement incremental improvement to existing component.	☐	☐	☐	☐	☐	☐	☐
3. Fix a bug - Provide patches to correct software defects or faults.	☐	☐	☐	☐	☐	☐	☐

Use Value: The potential value or usefulness you personally derive from your contributions.	1	2	3	4	5	6	7
1. 17 hrs/yr - Expected personal time savings resulting from your contributions.	☐	☐	☐	☐	☐	☐	☐
2. 34 hrs/yr - Expected personal time savings resulting from your contributions.	☐	☐	☐	☐	☐	☐	☐
3. 51 hrs/yr - Expected personal time savings resulting from your contributions.	☐	☐	☐	☐	☐	☐	☐

Career: The impact on your career or earnings potential.	1	2	3	4	5	6	7
1. Expected gain of 2,100 USD - Incremental salary increase resulting from participation.	☐	☐	☐	☐	☐	☐	☐

	1	2	3	4	5	6	7
2. Expected gain of 700 USD - Incremental salary increase resulting from participation.	☐	☐	☐	☐	☐	☐	☐
3. Expected gain of 0 USD.	☐	☐	☐	☐	☐	☐	☐

Effort: The effort or time required to participate.	1	2	3	4	5	6	7
1. 15 hrs/yr - Annual cost in time to participate.	☐	☐	☐	☐	☐	☐	☐
2. 35 hrs/yr - Annual cost in time to participate.	☐	☐	☐	☐	☐	☐	☐
3. 100 hrs/yr - Annual cost in time to participate.	☐	☐	☐	☐	☐	☐	☐

Recreational Value: The recreational value you experience from participation.	1	2	3	4	5	6	7
1. Very enjoyable - a high recreational value results from the contribution experience.	☐	☐	☐	☐	☐	☐	☐
2. Somewhat enjoyable - a moderate recreational value results from the contribution experience.	☐	☐	☐	☐	☐	☐	☐
3. Not enjoyable - little or no recreational value results from the contribution experience.	☐	☐	☐	☐	☐	☐	☐

Social: Your perception of Apache as a place to develop social relationships.	1	2	3	4	5	6	7
1. Very social - having a high degree of social communication and interaction.	☐	☐	☐	☐	☐	☐	☐
2. Somewhat social - having a moderate degree of social communication and interaction.	☐	☐	☐	☐	☐	☐	☐
3. Not social - little or no social communication or interaction.	☐	☐	☐	☐	☐	☐	☐

Tradeoffs - Section Three

Every day, people make tradeoffs between both similar and dissimilar things. These tradeoffs may or may not have explicit monetary value. For example, you might prefer to spend a few days on the beach as opposed to purchasing a new LCD monitor. In this section, we are interested in the way you might value hypothetical outcomes resulting from your open source participation.

Below is a list of choices that describe various possible outcomes resulting from your participation in an open source project. This exercise asks you to rank order your preferences for 8 possible outcomes from your open source participation. There are two steps to this exercise. Please **complete** step one before beginning step two.

Please note, this exercise is implemented as a Java applet. Once you have begun this exercise, please do not navigate away from this page. Since we do not use 'cookies', doing so will result in a reset of the applet resulting in the loss of your previous work.

STEP ONE
Rank the following 8 outcome descriptions from **"Most Preferred"** to **"Least Preferred"** by assigning an eight(8) to your most preferred outcome and a one(1) to your least preferred outcome. Please assigning a unique ranking to each of the eight outcomes. That is, any particular rank number should be used only once.

Once you have **completely finished** your rankings, please proceed to STEP TWO.

Rank	ASF Rank	Contribution Type	Use Value	Effort	Recreation Value	Social Value	Career Impact
-	Committer	enhance existing function	34 hrs/yr	35 Hours	Not enjoyable	Not social	700 USD
-	Developer	bug fix	51 hrs/yr	15 Hours	Not enjoyable	Not social	2,100 USD
-	ASF Member	add new function	17 hrs/yr	35 Hours	Somewhat enjoyable	Not social	2,100 USD
-	Developer	enhance existing function	34 hrs/yr	100 Hours	Very enjoyable	Very social	700 USD
-	Committer	bug fix	17 hrs/yr	15 Hours	Very enjoyable	Very social	0 USD
-	ASF Member	bug fix	17 hrs/yr	35 Hours	Very enjoyable	Somewhat social	2,100 USD
-	Committer	enhance existing function	51 hrs/yr	15 Hours	Somewhat enjoyable	Somewhat social	0 USD
-	ASF Member	add new function	34 hrs/yr	100 Hours	Somewhat enjoyable	Very social	0 USD

STEP TWO
Now that you have ranked the outcome descriptions according to your preference, use the pull-down menu to the left of each description to assign a numeric score between 0 and 100 percent. Choose a score to represent the chance (probability) that you would participate in the open source project if you knew in advance that your participation would result in the outcome you are

scoring.

Score	ASF Rank	Contribution Type	Use Value	Effort	Recreation Value	Social Value	Career Impact
-	Committer	enhance existing function	34 hrs/yr	35 Hours	Not enjoyable	Not social	700 USD
-	Developer	bug fix	51 hrs/yr	15 Hours	Not enjoyable	Not social	2,100 USD
-	ASF Member	add new function	17 hrs/yr	35 Hours	Somewhat enjoyable	Not social	2,100 USD
-	Developer	enhance existing function	34 hrs/yr	100 Hours	Very enjoyable	Very social	700 USD
-	Committer	bug fix	17 hrs/yr	15 Hours	Very enjoyable	Very social	0 USD
-	ASF Member	bug fix	17 hrs/yr	35 Hours	Very enjoyable	Somewhat social	2,100 USD
-	Committer	enhance existing function	51 hrs/yr	15 Hours	Somewhat enjoyable	Somewhat social	0 USD
-	ASF Member	add new function	34 hrs/yr	100 Hours	Somewhat enjoyable	Very social	0 USD

APPENDIX III.
OVERVIEW OF ARCHIVAL DATA COLLECTION

Research Site – The Apache Software Foundation

The Apache Software Foundation (ASF) is a not-for-profit corporation that provides the legal, organizational and financial infrastructure for the software projects gathered under the ASF open-source umbrella. Each of the ASF projects operate autonomously controlling all aspects of product development including project management, requirements specification, architecture, design, development, testing, and configuration management. ASF projects are characterized by a "collaborative, consensus based development process, an open and pragmatic software license, and a desire to create high quality software that leads the way in its field" (Apache 2001).

At the time this research was undertaken, the ASF encompassed eight subprojects related to the development and support of a full featured web server product offering. The Apache server project is a freely available source code implementation of an HTTP (Web) server. The Apache HTTP (web) server project is the project around which Apache Group initially formed. The Apache Portable Runtime project is a free library of C data structures and subroutines designed to facilitate porting the Apache HTTP Server to a host of disparate operating systems. The Jakarta project consists of all Apache related server side Java projects . Jakarta consists of over 18 Java related subprojects. The Apache/Perl project is the integration of the Perl programming language implemented as an Apache HTTP server module. Similarly, the PHP project is an embedded scripting language implemented as an Apache HTTP server module. The Tool Control Language (Tcl) project is an umbrella for Tcl-Apache integration efforts. These projects combine the Apache web server with the Tcl scripting language. The Apache XML project is home for Apache XML related activities. There are over 9 XML related subprojects.

The projects of the Apache Software Foundation (ASF) enjoy wide acceptance both in the marketplace and in the OSS development community. According to a recent web server survey by Netcraft (2005), the Apache web-server and its derivatives have attained a dominant 70% share of the web-server market, and the Apache web-server software is installed at more than 44 million sites worldwide. Similarly, the ASF projects consistently attract and retain the large number of participants vital for OSS project success (Von Krogh et al. 2003). The three ASF projects in this study accepted 100,000 changes from over 1,300 different open source developers from 1999-2002. Membership in the ASF is by invitation only and is based on a strict meritocracy.

The Apache Career

A common characteristic OSS projects is presence of a strong project leader. (Raymond 2001a) Apache, however, is unique among OSS projects in this regard. Since its inception the Apache project has operated under a model of shared leadership and responsibility (Fielding 1999). This model of shared responsibility is reflected in the principles of the ASF meritocracy that defines advancement within the ASF (Apache 2001). As a meritocracy status, responsibility, and benefits are commensurate with contribution. There are five observable levels of recognition or rank within the ASF. In order of increasing status, these are developer, committer, ASF member, project management committee member, and ASF board member. In all cases advancement is in recognition of an individual's commitment and contributions to an Apache project.

Individual reasons for initial involvement in any Apache project vary. Typical reasons cited include reporting a problem or "bug", fixing a problem in the software that has become a nuisance or impairs usage. Another reason is to extend existing functionality or add new features required by the user or their organization. For the majority of contributors there is a single encounter with the project. Some developers however choose a deeper level of involvement and continue to make contributions. If a developers contributions are significant and consistent over a period of time they may be nominated for an increase in rank from developer to "committer". The practical significance of becoming attaining the rank of committer on any Apache project is the privilege of submitting code changes directly to the source code repository as opposed to going through an intermediary to have your changes included in the product. Committers who continue their involvement in the project may be nominated for ASF membership by an existing ASF member. ASF membership is a largely a matter of recognition and carries with it a certain prestige in the Apache community. ASF members are eligible to be nominated by the ASF board of directors or to serve on a project committee. Project committee members are responsible for all aspects of managing an Apache subproject including project plans and roadmaps, release schedules, etc. ASF members may be elected to the ASF board of directors.

Archival Data Collection

One of the basic tenets of OSS is that the development process and resulting products are "open" and freely available. In keeping with free and open access, all OSS work products are placed in the public domain under various "free software" licensing arrangements.

To assess a contributor's performance ranking requires the construction of an Apache career path for each contributor. We capture the upward progression as a series of discrete transitions from one level to another in the ASF meritocracy. In our study, a participant's first source code contribution is considered a consummation of the joining decision, signaling entry into the meritocracy and the beginning of the Apache career. To determine each individual's Apache career progression we used archival data from three sources: the Apache developer web site, contribution meta-data from the Concurrent Versioning System (CVS) revision control software, and minutes from the Apache Board of Directors meetings. Each Apache subproject maintains a separate developer website that includes a list of contributors and project management committee members. By observing changes to these files over time we are able to construct a timeline for the promotion of individuals within each project.

To extract information about individual contributions from the data, we developed tools to mine contribution information at the level of the individual developer. A submission to an OSS project is commonly referred to as a "patch." Patches are analogous to modification requests in traditional software development environments. Unlike modification requests, however, patches result from the nearly random arrival of developer submissions and have no formal designation or means of tracking. Our research follows the method used by Mockus et al. (2000) to reconstruct patches from source code archives. For each patch we extracted and retained common software metrics including lines of code added and deleted, the date of submission, the names and number of source code files affected by the change, change log entries, and the list of patch authors.