Strategic Management Journal
Strat. Mgmt. J., **37**: 2589–2610 (2016)
Published online EarlyView 14 December 2015 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2439
Received 17 November 2014; Final revision received 15 August 2015

PROBLEM-FORMULATION AND PROBLEM-SOLVING IN SELF-ORGANIZED COMMUNITIES: HOW MODES OF COMMUNICATION SHAPE PROJECT BEHAVIORS IN THE FREE OPEN-SOURCE SOFTWARE COMMUNITY

NICOLAI J. FOSS,[1,2]* LARS FREDERIKSEN,[3] and FRANCESCO RULLANI[4,5]
[1] *Department of Strategic Management and Globalization, Copenhagen Business School, Frederiksberg, Denmark*
[2] *Department of Strategy and Management, Norwegian School of Economics, Bergen, Norway*
[3] *Department of Management, BSS, Aarhus University, Aarhus, Denmark*
[4] *Department of Business and Management, LUISS Guido Carli, Rome, Italy*
[5] *Department of Innovation and Organizational Economics, Copenhagen Business School, Frederiksberg, Denmark*

Research summary: *Building on the problem-solving perspective, we study behaviors related to projects and the communication-based antecedents of such behaviors in the free open-source software (FOSS) community. We examine two kinds of problem/project-behaviors: Individuals can set up projects around the formulation of new problems or join existing projects and define and/or work on subproblems within an existing problem. The choice between these two behaviors is influenced by the mode of communication. A communication mode with little* a priori *structure is the best mode for communicating about new problems (i.e., formulating a problem); empirically, it is associated with project launching behaviors. In contrast, more structured communication fits subproblems better and is related to project joining behaviors. Our hypotheses derive support from data from the FOSS community.*

Managerial summary: *We study how the way in which individuals communicate influence the project-behaviors they engage in. We find that relatively unstructured communication is associated with the setting up new projects, while communication that is structured around an artifact is associated with joining projects. Our findings hold implications for understanding how management may influence project behaviors and problem-solving: Firms that need to concentrate on more incremental problem-solving efforts (e.g., because a sufficient number of attractive problems have already been defined) should create environments in which interaction is undertaken mainly* via *artifacts. On the other hand, if firms seek to generate new problems (e.g., new strategic opportunities), they should create environments in which open-ended, verbal conversation is relatively more important than artifact-based communication.*

INTRODUCTION

A core part of strategic thinking is envisioning a space of possible strategic opportunities, intelligently searching within that space, and deploying resources, actions, and investments in

Keywords: communities; open-source software; problem–solving; communication; project organization
*Correspondence to: Nicolai J. Foss, Department of Strategic Management and Globalization, Copenhagen Business School, Kilevej 14, 2nd floor, 2000 Frederiksberg, Denmark. E-mail: njf.smg@cbs.dk

the pursuit of promising opportunities (Afuah and Tucci, 2012; Denrell, Fang, and Winter, 2003; Foss and Klein, 2012; Mintzberg, Raisinghani, and Theoret, 1976). We address communication-related aspects of this process by focusing on the extent to which (online) dialogue and communication about new opportunities, projects, and business ideas are free and unconstrained *or* whether such dialogue and communication are more constrained and revolve around existing artifacts. We argue that the former kind of search is more likely to be associated with creating and initiating new major strategic initiatives, while the latter kind of search is associated with more incremental initiatives. Our empirical setting is a community for free open-source software ("FOSS") development.

The research literature on communities conceptualizes these as self-organizing organizational forms that access and combine relevant knowledge that is distributed among individuals (Hayek, 1945) in the production of public goods, often relying on IT-mediated communication (Frey, Luethi, and Osterloh, 2012). Community members establish new projects aimed at formulating and solving problems within an overall framework characterized by voluntarism, unpaid contributions, free entry, the absence of externally imposed rules, and the presence of self-defined lateral governing rules. Although research on communities has many strides forward, little is known about the evolving pattern of problem-formulation and problem-solving activities and how these activities are embedded in the basic organizational unit of most communities, namely, projects (Dahlander and O'Mahony, 2011; O'Mahony, 2007). In many communities, such dynamics take the form of individuals launching and joining projects. The relative lack of understanding in the literature of these organizational dynamics is problematic because the ability of communities to produce private and social benefits through problem formulation and solving depend precisely on these dynamics (Nickerson and Zenger, 2004). Obtaining a better understanding of these dynamics is important because firms increasingly seek to tap into communities to increase their innovativeness (e.g., Bonaccorsi, Giannangeli, and Rossi, 2006; Dahlander and Wallin, 2006; Fosfuri, Giarratana, and Luzzi, 2008; O'Mahony and Bechky, 2008; O'Mahony and Lakhani, 2011). There may be important lessons to learn for firms from how the operation of the community stimulates the creation of new strategic initiatives.

We address communication-based antecedents of whether community members launch new projects (set up new problems) or join existing projects (solve subproblems defined by others or define *subproblems* within an overall problem structure defined by others), framing our discussion in terms of the "problem-solving perspective" (e.g., Baer, Dirks, and Nickerson, 2013; Heiman, Nickerson, and Zenger, 2009; Macher, 2006; Macher and Boerner, 2012; Nickerson and Zenger, 2004). We specifically argue that an individual's project-behavior (i.e., launching or joining) is predicted by the mode of communication (Guetzkow and Simon, 1955; Orlikowski and Yates, 1994; Zmud, Lind, and Young, 1990) and the associated artifacts (or absence thereof) that characterize the context of interaction around the focal individual. In the FOSS community, the mode of communication ranges from unstructured, informal open-ended communication to highly detailed, formal, and specific communication that revolves around a concrete artifact (i.e., a piece of code) (Carlile, 2002). We argue that different modes of communication predict individuals' behaviors as follows: Communication with less *a priori* structure (i.e., open-ended and informal communication with few constraints) is the best mode for communicating about new problems; empirically, it is therefore associated with project launching behaviors. In contrast, more structured and formal communication (i.e., communication that directly relates to well-defined objects, and therefore, has inherent constraints) fits subproblems better and is therefore more associated with project joining behaviors. We develop two hypotheses from this overall argument and test them on data from SourceForge.net (henceforth, *SF.net*), the largest repository of open-source software projects. This data set details the activity of 544,669 developers over 840 days from September 2000 to December 2002. The empirical analysis supports our hypotheses.

In sum, we develop insights into the influence of communication-related antecedents on actual problem-formulating behaviors. We argue that our theory and findings generalize to other organizational forms that—like communities—to some degree rely on spontaneous, self-organizing forces to allocate resources to the refinement of existing ideas, and to launching and developing new ideas. Examples include R&D departments and project-based firms (Felin and Zenger, 2011; Gann and Salter, 2000) as well as charities and types of NGOs (e.g.,Wikipedia) (Kuznetsov, 2006; Powell and Hwang, 2009). Thus, this research helps to understand the communication-based antecedents of the process of generating and realizing new opportunities, projects, and business ideas.

THEORY AND HYPOTHESES

Communities and projects: a problem-solving approach

Research on communities spans several fields in management, and takes place under headings such as commons-based peer production (Benkler, 2006; Lessig, 2001), virtual organizations (Wasko and Faraj, 2005), open-source science (Jeppesen and Lakhani, 2010; Lakhani *et al.*, 2007), private-collective innovation (von Hippel and von Krogh, 2003), and community-based innovation (David and Rullani, 2008; Lakhani and von Hippel, 2003; Shah, 2006; West and O'Mahony, 2008). A substantial part of the research literature deals with such issues as individuals' motivations for participating in FOSS development efforts (von Krogh *et al.*, 2012; Lakhani and Wolf, 2005; Lerner and Tirole, 2002) and the skill profiles of innovative individuals in FOSS communities (Giuri, Rullani, and Torrisi, 2008; Giuri *et al.*, 2010; Von Hippel and Von Krogh, 2006). However, there is relatively little attention paid to the organizational dynamics of communities, and how such dynamics unfold in terms of individuals launching and joining projects. Somewhat similar to the way in which firm dynamics of entry and exit influence overall knowledge creation and knowledge spillovers in a region or industry (Delmar, Wennberg, and Hellerstedt, 2011), project-related behaviors matter to the dynamics of communities (David and Rullani, 2008; Von Krogh, Späth, and Lakhani, 2003).

The emerging problem-solving perspective (e.g., Afuah and Tucci, 2012; Baer *et al.*, 2013; Felin and Zenger, 2014; Macher, 2006; Macher and Boerner, 2012; Nickerson, Silverman, and Zenger, 2007; Nickerson, Yen, and Mahoney, 2012; Nickerson and Zenger, 2004) is a useful lens for understanding these important behaviors. The problem-solving perspective builds from complexity theory (Kauffman, 1993; Simon, 1962; Stacey, 1995) to characterize problems and their associated solution landscapes. The key idea is that different governance forms (different kinds of firms, markets, and hybrids, cf. Williamson, 1996) have different problem-solving and solution-search capabilities (Macher and Boerner, 2012; Nickerson and Zenger, 2004). Problems are dimensionalized in terms of their complexity (i.e., the number of variables, their connections, and possible changes in these; cf. Simon, 1962), the extent to which they are well structured (i.e., the extent to which established approaches to framing the problem and making relevant decisions exist) (Baer *et al.*, 2013; Macher and Boerner, 2012), and the extent to which solving the problem requires access to dispersed knowledge (Felin and Zenger, 2014; Hayek, 1945; Heiman *et al.*, 2009). Governance forms have different capacities to handle problems because they have differential access to mechanisms of dispute resolution (Williamson, 1996); can enforce property rights over solutions to varying degrees; and have different abilities to deploy modes of communication and knowledge-matching mechanisms (Felin and Zenger, 2014; Nickerson and Zenger, 2004).

However, the treatment of such communication mechanisms is not fully addressed within the problem-solving perspective. Additionally, while the problem-solving perspective takes problems and their associated solution landscapes as given, the *formulation* of problems is logically prior to their resolution. Problem-formulation is a generic human activity (Popper, 1972) that converts ill-structured problems into solvable problems (Simon, 1973, 1987). However, problem-formulating behaviors have so far featured less prominently in theoretical and empirical work in this perspective (cf. Baer *et al.*, 2013: 200).

Problem-formulation and problem-solving in the FOSS community

Members of the FOSS community can formulate new problems by launching a new project. In turn, other community members can then join the new project to engage in formulating subproblems within the project or work on problems defined by the project founder. While three kinds of behavior are thus involved, there is a fundamental distinction between setting up a new overarching problem and the more incremental behaviors dedicated to formulating smaller problems within the given problem structure (i.e., leaving this structure unchanged) or even accepting to work on predefined problems.

DOI: 10.1002/smj

A similar distinction exists in the related literature on modularization (e.g., Baldwin and Clark, 2003), where, once an overall architecture with interface standards and so on have been defined, problem-solving and problem-definition can be encapsulated in the modular parts of the structure. The distinction is also similar to the one in the entrepreneurship literature between the entrepreneur's initial formulation of the basic business model of the start-up venture in terms of overall value proposition, segments, value chain organization, and so on, and the individuals hired by the entrepreneur, the behavior of whom takes places within the framework defined by the entrepreneur (e.g., Foss and Klein, 2012). Additionally, it is a distinction that can be anchored in the complexity theory that forms the foundation of the problem-solving approach.

Like other organizational forms, the FOSS community can be viewed as a complex system, as it is composed of a large number of individuals and objects that "interact in a non-simple way" (Simon, 1962: 468; Macher and Boerner, 2012; Felin and Zenger, 2014). Complexity is often organized by a "hierarchy," that is, a system that consists of interrelated subsystems. In turn, each subsystem is hierarchical (Simon, 1962: 473). Interactions among subsystems may be few (decomposable systems), many and essential (nondecomposable systems), or weak but not negligible (nearly decomposable systems) (Simon, 1962).

While there is a conceptual distinction between problem formulation and problem solving, this distinction is somewhat blurred in practice (von Hippel and von Krogh, Forthcoming). The overall problem may be well defined, so that there is a "symptom or a web of symptoms recognized as needing to be addressed" (Baer *et al.*, 2013: 81). However, the solution of an overarching problem usually proceeds by decomposing the problem into more manageable subproblems that are easier to address given the scarcity of attention (Simon, 1973). They may, for example, be delegated to other firms or individuals (Simon, 1991). As Felin and Zenger (2014: 923) explain:

> … firms have increasingly invested in innovatively transforming previously complex, non-decomposable problems into simpler, more decomposable problems often guided by the composition of theory. Firms make upfront investments in high-level design, commit to high-level architecture that reflects this theory, and then outsource sub-problems. Such modularization allows firms to decompose innovation into smaller sub-problems around which the firm can engage the productive (and sometimes voluntary) efforts of external constituents.

In turn, "innovatively transforming rather previously complex non-decomposable problems into simpler decomposable problems" usually means *formulating* those "more decomposable problems" (Felin and Zenger, 2014:923).

The interdependence of problem-formulation and problem-solving also follows from the impossibility of getting a decomposition of a problem exactly right initially (see Simon, 1973: 191). As unforeseen interdependencies emerge, an iterative approach becomes necessary (Staudenmeyer and Cusumano, 1998). This often calls for some measure of governance, such as authority. Simon (1973) argues that a specific individual (he uses the example of an architect) formulates the initial statement of the problem that is to be solved, including a specification of the problem architecture (i.e., the specification of subproblems, their interdependencies, and the constraints on problem-solving efforts). In the FOSS community, this role is assumed by the individual who launches a project.

Project behaviors in communities

In the FOSS community, problem-formulation and solving activities form around self-organizing projects, problem-formulation resulting in the formation of multi-individual projects that are aimed at solving a particular development problem in a team-based manner. As an example, here is how the (anonymous) founder of Project #18765 in our data provides the following description of the purpose of the project:

> [Project #18765] is a new Stealth Portscanner for Linux. It is currently in the first alpha version and some advanced scan techniques have already been implemented (SYN, FIN, NULL and XMAS Scan, for example). With the help of some people I hope to find here, I will try to create an alternative to [Project #543]. Some people told me [Project #18765]

> was already faster than [Project #543], but for now it is not very stable. I try to avoid a really BIG code like [Project #543]. The code will stay clean and small, and I hope that I do not need the libcap.

Other community members may join a project around a formulated problem and work on subproblems. Thus, the overall problem-formulation and solving process within FOSS communities has hierarchical features that are reflected in the division of problem-related labor. Certain community members provide the initial formulation of a problem, which involves defining a project and its initial architecture (i.e., the specification of interdependencies), and they then launch that project. Other community members join the project to work on its subproblems with the intention to solve all or parts of the problem, often in this process formulating subproblems in the context of the overarching problem first identified by the project launcher. As Fleming and Waguespack (2007: 170) explain:

> The boundaries within open innovation communities, like those in private firms, usually correspond to the interfaces between technological subsystems (Henderson and Clark, 1990). Each boundary demarcates a distinct technological area or module. The boundaries are defined directly by leaders' architectural decisions and then implemented by followers' choices of where to volunteer their efforts.

The question, then, is what explains this division of project-related labor. Communities rely on unpaid contributions and have no imposed central hierarchical authority. Furthermore, there are no significant direct monetary rewards (or often, any monetary rewards at all) associated with setting up projects in an entrepreneurial fashion, property rights are largely undefined, and there is no management team formally issuing directions on such issues as project formation, tasks, or staffing. This suggests that antecedents exist outside the realm of the conventional drivers of behaviors in markets and organizations. We specifically theorize determinants of project-related behaviors that are rooted in communication modes within the community and in attention allocated to projects by members of the community.

Modes of communication in communities

The problem-solving perspective suggests that governance forms differ with respect to the communication and knowledge-matching mechanisms they can deploy (Nickerson and Zenger, 2004). However, the treatment of these mechanisms in the problem-solving perspective is rather stylized. Yet, it is well-established that the communication of information is essential for knowledge creation, sharing, integration, application, and storage (Grant, 1996; Kogut and Zander, 1992). Communities are highly communication intensive (Monge and Contractor, 2003) and they match knowledge by means of mechanisms of self-organization in which community members voluntarily propose, launch, and join projects. Communities heavily rely on different kinds of communication in order to coordinate this self-organizing process. In particular, IT-based communication plays a major role (e.g., Frey *et al.*, 2012).

Daft and Lengel (1986) emphasize how the communication context influences individual behaviors. Other researchers argue that the notion of a "genre" is an apt metaphor for understanding communication in communities. While traditionally applied to literature, music, and other modes of art and entertainment, the notion of a genres is applicable in this context, because it rests on the notion that certain kinds of style fit some themes better than others (e.g., in classical theater, speech patterns appropriate for comedy are not appropriate for tragedy). Because communication (i.e., themes) differs, modes of communication (i.e., genres) differ. Orlikowski and Yates (1994: 542) stress that: "[m]embers of a community rarely depend on a single genre for their communication. Rather, they tend to use multiple, different, and interacting genres over time." However, as these scholars argue, the mode chosen for conveying a particular type of content is closely related to the purpose of that communication because a given mode of communication has a certain level of functionality within a specific context. This is an idea we rely on in the following. Thus, modes of communication influence the way information is presented, structured, transmitted, framed, received, and encoded by individuals (Zmud *et al.*, 1990). So do the *artifacts* that may be associated with specific modes of communication.

Research into computer-mediated communication in the field of computer science (e.g., Bly,

DOI: 10.1002/smj

Harrison, and Irwin, 1993; Carstensen and Nielsen, 2001) and work in organization theory (DeSantis and Monge, 1999) explore how artifacts—that is, analogue and digital objects—support mediated communication. A key insight is that artifacts matter because they are tools that can be physically or virtually shared by individuals, thereby providing "similar sorts of representation and manipulation capabilities to all the involved participants, making the shared artifacts a common ground amongst the participants" (Vyas, Dix, and Nijholt, 2008: 3). As they offer a "common ground," artifacts allow for economizing with the amount of attention that can be devoted to problem-formulation and problem-solving (Carlile, 2002). Relatedly, artifacts constrain and demarcate communication because they define key parameters of the communicative interaction.

Following Holmquist *et al.* (2001), we talk about "artifact-based communication" when we refer to the "epistemic machinery" (Kaplan, 2011) that supports ongoing communication. This includes not only the actual *means* of communication, but also—and importantly in our context—artifacts around which communication revolves. Such artifacts enable and constrain communication (Sapsed and Salter, 2004; Star and Griesemer, 1989). For example, when a team of product developers communicate about a concrete product prototype, that prototype defines a common ground that helps to coordinate the team's communicative acts (Carlile, 2002; Lyytinen and Newman, 2008). However, the prototype also places constraints on and boundaries around the communicative space. In general, such artifact-based communication is inherently constraint laden.

Within the context of the FOSS community, the relevant artifacts are existing lines of formal code. Communication revolves around these artifacts in the forms of patch files, bug reports, feature requests, and support requests. Artifacts enable and constrain problem solving. As Haefliger, von Krogh, and Spaeth (2008) show, existing code not only provides a starting point for subsequent problem solving, but also introduces path dependencies and diminishes flexibility because choosing a specific starting point makes certain constraints binding (see Simon, 1973). To illustrate, when discussing the development of Emacs, the program that gave birth to the FOSS phenomenon, Lin (2004:7) notes that:

> Stallman did not sit down and write [Emacs] immediately […]. Instead, he looked up the database and found that Mikkelson had made a WYSIWYG feature for TECO. He then integrated his idea into that. If Mikkelson's work had not existed, we may have seen a different technical option taken, as the "problem" may have been defined differently.

Interaction in a community may therefore occur with close reference to an existing artifact that functions as a boundary object (Carlile, 2002; Star and Griesemer, 1989). The resulting processes of problem-formulation and problem-solving align closely with that artifact because of its role as epistemic machinery that constrains and enables cognitive processes and interaction (Woods, 1998); thus, subproblems are formulated that relate directly to the existing artifact (e.g., handling a bug in an existing piece of software). In the empirical context of a self-organizing, project-based community, this is manifested in member's joining already existing projects to work on subproblems defined by someone else.

In contrast, communication behaviors that do not revolve closely around an existing artifact have, by definition, fewer artifact-based constraints (e.g., there is no explicit code base at the outset). This communication is informal as it tends to be spontaneous, interactive, and rich (March and Simon, 1958). Hence, communication can focus on envisioning the definition of, structure for, and possible solutions to an imagined problem. The space available for mental exploration is therefore larger. In this regard, open-ended and informal communication among individuals in a community form is a low-cost way of carrying out mental hypothesis testing (i.e., cognitive search; Gavetti and Levinthal, 2000) related to the formulation of new problems and possible resolutions to those problems (i.e., a problem architecture). As new problems are translated operationally into new projects, such communication is particularly likely to drive project-launching behavior. In addition, project launching includes a set of communication activities concerned with "managerial" tasks of designing the new project architecture, defining *subproblems*, and trying to recruit people to join the endeavor. Communication of this type is typically open-ended because no real artifact has (yet) been produced. The above reasoning motivates the following hypothesis:

Hypothesis 1: The more an individual is exposed to open-ended communication, the higher the number of new projects the individual will launch.

Project launch manifests that an overarching problem and a problem architecture have been formulated. Once launched by one or more individuals within a community, other community members can join the project. In the context of the FOSS community, this means that project-related artifacts are likely already to be posted online (e.g., code relating to various subproblems) at the time when other individuals may consider joining the project. Such artifacts may stimulate conversations related to the subproblems that the artifacts represent. This type of conversation attracts individuals toward these subproblems and increases the probability that they will join the existing project. Thus, we proffer the following hypothesis:

Hypothesis 2: The more an individual is exposed to artifact-based communication, the higher the number of existing projects the individual will join.

RESEARCH SETTING AND METHODS

Data sources

To empirically address the above hypotheses, we examine a data set of individual contributions to a FOSS community. Specifically, we employ data from SF.net to describe the activities of 544,669 individual developers from September 2000 to December 2002. This period is an early one in the context of the FOSS community. At the time, the influence of firms was still marginal. The marginal role of firms is important fact as we address *self-organizing* problem-formulation, and given the marginal role played by firms in this period, we do not risk including the influence of firms on individuals' project-related behaviors. SF.net is an online platform for the development of various types of software. It hosts contributions from software users and developers. Contributions include specific pieces of code and many other types of content. Individuals registered on SF.net can launch their own projects, browse the platform, identify projects that they may be interested in joining, submit bug reports or code patches, suggest new program features, ask for support, participate in discussions on forums or mailing lists, or simply "lurk" (David and Rullani, 2008). All of these activities are reported in the dataset, various versions of which have been widely exploited in the literature (e.g., Comino, Manenti, and Parisi, 2007; Lerner and Tirole, 2005).[1]

Dependent variable: project-related behaviors

A new project team is usually directed by the project's initiator (Lerner and Tirole, 2002) who at the outset formulates the problem, and outlines, and perhaps, also defines the sequence of the first steps toward solving the problem (Raymond, 1998).[2] In other words, the initiator designs the problem architecture. In contrast, members who join a project adopt a second-tier role within an existing team, in which many routines, roles, tasks, and cognitive structures are predefined.

Community members who wish to enlarge the scope of their activities (i.e., moving beyond the projects they are already working on) may focus on their own ideas and interests, and invest time and resources in the formulation of a new project that they will lead. Hence, the initiator may learn additional things beyond the code related to, for example, problem-formulation, task structure design, and task sequencing to ensure the best problem-solving effort. Alternatively, an individual may join an existing project. Such project members may be able to learn more specific heuristics regarding code development. An individual who joins a project that has already been formulated in terms of the main problem it seeks to address and the overall architecture of the problem-solving efforts typically joins a more or less well-defined subproblem within the architecture. In fact, joining behavior may be a matter of formulating a subproblem. To measure these two distinct behaviors in the FOSS community, we create two variables: *projects launched*$_{it}$, which refers to the number of new projects launched by individual i in period

[1] For detailed discussions of this dataset, see Howison and Crowston (2004), Comino *et al.* (2007), Giuri *et al.* (2010), and David and Rullani (2008).

[2] When a new FOSS project is created, the code on which it is based may not be entirely novel (Haefliger *et al.*, 2008; Sojer and Henkel, 2010). However, the problem-formulations the projects adopt are likely different as the teams of the two projects are different (David and Rullani, 2008), implying a difference in the way that initial code is conceived and further developed.

DOI: 10.1002/smj

Excerpt from forum "Developers"

Project: WebGUI (A perl-based content management system designed to let the people who create the content manage it, and let the technical folks get back to tech stuff.)

Thread: WYSIWG Development (Architecture and Implementation Discussion)

Dcolling	*Fri, 26 Apr 2002 21:01:28 GMT*	*Who was the "rich" editor developed by? There are some features I'd like to see in there, such as integration with the image gallery. If it was developed by PBS then it makes sense to work on it; […]*
→rizen	*Fri, 26 Apr 2002 23:01:37 GMT*	*We got the editor from ZDNet Devhead (if I remember correctly) and had to do major mods to even make it work. Now Devhead is no longer. We also want to integrate with the image manager as well as the macros. […]*
→→dcolling	*Sat, 27 Apr 2002 06:33:58 GMT*	*Yeah, I ran across that discussion in my travels looking for a WYSIWYG. Mozilla Composer is quite a decent little HTML editor these days, though I'm wondering who they think will be using it. Yes, it would be great if they could turn it into a component*
→→→rizen	*Sat, 27 Apr 2002 15:24:41 GMT*	*you may or may not be aware of this, but this fall the browser wars (ie vs mozilla) will be back on again.*
→flaviocu	*Sat, 27 Apr 2002 22:13:03 GMT*	*well, i actually thought about writing a java-editor for webgui so we could have a browser independent editor... i'l look at realobjects (but it is payware, no?) what do you think? would a java applet be an option?*
→→dcolling	*Sat, 27 Apr 2002 22:35:47 GMT*	*I just tried out RealObjects and found it to work very well. Yes, it's a pay-for item but that may be ok for a commercial site, which most of mine are. It's Java, so it's kind of heavy but it worked well enough for me. […]*
→→→flaviocu	*Sat, 27 Apr 2002 22:39:08 GMT*	*yeah, from a commercial viewpoint this may be true. i for myself would be happy if mozilla and co could at least get the functionality of the current IE editor for free... this is what i'm aiming for... […]*
→→→rizen	*Sun, 28 Apr 2002 06:53:15 GMT*	*yeah. it totally sux. that's why i'm hoping that the mozilla crew comes through for me.*

Figure 1. An example of a forum discussion

t, and *projects joined*$_{it}$, which indicates the number of existing projects joined by individual i in period t.

Independent variables

To operationalize communication-related antecedents of project-related behaviors, we use two main tools provided on the SF.net platform, namely, forums and tracker systems. On a project's forum, individuals may read other people's messages and submit their own. Figure 1 provides an example of a forum discussion. Tracker systems are an accessible repository of software-related artifacts, such as bug reports, feature or support requests, and patch files, which individuals can browse (see the example in Figure 2). The individual can select one of these artifacts and then work on it, upload an improved version of it, or comment on it.

Clearly, the two environments are associated with different modes of communication. Open-ended communication is typical of forums, while bug reports, feature or support requests, and patch files revolve around the code, that is, an artifact. These data allow us to identify the individuals that are members of the same project as the focal individual i (henceforth, *colleagues*). We can also identify how they communicate, that is, whether they send messages to forum (i.e., a form of open-ended communication) or submit anything to the tracker system (i.e., a form of artifact-based communication). In this way, we capture to what extent i's colleagues are engaged in each of the two modes of communication.

A potential concern is that i's colleagues may communicate simply in reaction to i's stimuli. However, we avoid this source of endogeneity by isolating and employing in our regressions only colleagues' communication that we are sure is not induced by i's behavior. We do this by focusing on the communication of project colleagues that only pertains to *projects in which individual* i *is not involved*, and discarding colleagues' forum messages and tracker system submissions sent to projects that i is also a member of. Notice that this approach in addition to addressing the above endogeneity concern allows us to focus on the communication i is exposed to in the community context, independently of her or his role in it.

These steps motivate defining our independent variables as *open-ended communication*$_{it}$ and *artifact-based communication*$_{it}$, defined as the

Excerpt from tracker system

Project: WebGUI (A perl-based content management system designed to let the people who create the content manage it, and let the technical folks get back to tech stuff.)

Artifact title: editUserProfile: wrong default in select

Submitted by: *cmarcant*
(Mon, 29 Jul 2002 19:15:19 GMT)

Assigned to: *rizen*

Tracker category: *current stable release*

Priority: *5*

Resolution: *Fixed*

Status: *Closed*
(Wed, 31 Jul 2002 02:43:58 GMT)

Artifact details:

While editing a user profile (not necessarily once's own), the default for a select is taken out of the session user profile rather than the edited user's profile.

Lines 368-369 of WebGUI/Operation/User.pm should be changed from:

```
} elsif ($session{user}{$data{fieldName}}) {
    $default = [$session{user}{$data{fieldName}}];
```

to:

```
} elsif ($user{$data{fieldName}}) {
    $default = [$user{$data{fieldName}}];
```

This change is totally in line with the behavior exercised in non select fields demonstrated at line 376:

```
|| $user{$data{fieldName}}
```

Christophe

Figure 2. An example of an artifact (a patch uploaded to a tracker system)

number of messages *i*'s colleagues have sent to the forums of projects in which individual *i* is not involved, and the number of artifacts (i.e., bug reports, patches, feature or support requests) colleagues have submitted to the tracker system of projects in which individual *i* is not involved, respectively. Figure 3 helps visualize the steps we undertook to generate each measure.

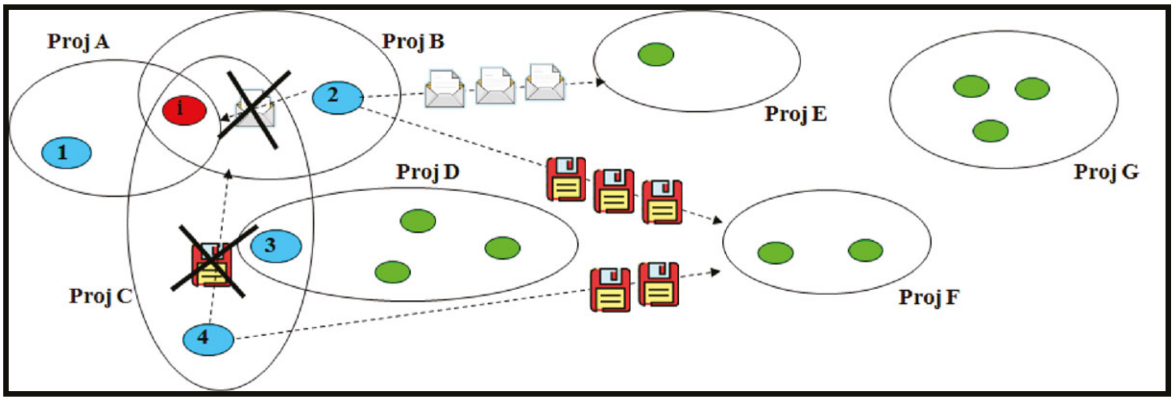
To be sure, forums and track-systems are not the only modes of communication. Mailing lists, blogs, chat, and face-to-face, are also modes of communication individuals engaged in FOSS would use (Kuk, 2006; O'Mahony and Ferraro, 2007). As it would be impossible to trace all these communications, we were forced to make a choice, and look for communication modes that could be considered good representatives of the quantity (how intense) and quality (open-ended or artifact-based) of the communication around developers *i*. We decided to focus on forum messages and artifacts sent to the tracker systems on SF.net as it seems safe to assume that they are a good proxy of how much *i*'s colleagues are prone to communicate *via* open-ended communication (forum messages) or through modes centered around virtual objects such as software, lines of code, features, and so on (artifacts). Moreover, a project's forums and tracker systems are usually the main discussion arenas that are open to external contributors and participants. Other means of communication, such as mailing lists, may be visible but perceived as not being open to external participation as they typically involve mainly the registered members of the project. Thus, focusing on forums and tracker systems, we can capture how much *i*'s colleagues communicate in the open arenas of the SF.net environment, the arenas composing the main backbone of individuals' interaction on SF.net.

Controls

As we estimate a fixed-effect model, we automatically control for any time-invariant characteristics of the individuals. We also control for their time-variant characteristics, such as the number of 30-day periods following their registration on the platform ($tenure_{it}$),[3] messages in forums ($N\ forum\ messages_{it}$), artifacts in tracker systems ($N\ artifacts\ in\ track\ system_{it}$), and posts on a project website ($N\ posted\ news_{it}$) in period t. The latter variables serve as proxies for the individuals' level of communication and participation. We also count the number of projects to which each individual belongs in period t ($N\ participated\ projects_{it}$) and use $dummy\ in\ SF\ staff_{it}$ to track whether the

[3] Given the correlation between this variable and the variable capturing the actual months in which the registration has happened (0.70 at 1% level of significance), $tenure_{it}$ is capable of accounting for the whole time dimension.

DOI: 10.1002/smj



Key

Proj = Project

= Focal individual *i*

= Colleague of the focal individual (i.e., member of the same project as the focal developer)

= Member of the same project as *i*'s colleagues but in projects *i* does not belong to.

= Forum message: captures the *propensity to communicate using open-ended communication*

= Bug reports, patches, feature and support requests, and the like: captures the *propensity to communicate using artifact-based communication*

Regressors

In this case, the two main regressors are computed as:

-) *Open-ended communication*$_{it}$ = number of forum messages sent by *i*'s colleagues to projects to which *i* does not belong = **3**

-) *Artifact-based communication*$_{it}$ = number of bug reports, patches, feature and support requests (and the like) sent by *i*'s colleagues to projects to which *i* does not belong = **5**

Figure 3. Building the main variables

individual belongs to the project managing SF.net itself.

Additionally, we take the average characteristics of each individual's projects into account, such as registration period (*Avg projects registration time*$_{it}$), performance measured as number of produced file releases (*Avg projects file releases*$_{it}$), and the ideological attitude toward FOSS based on the number of projects adopting the most restrictive OSS license (the General Public License—GPL: *N projects with GPL*$_{it}$).

Finally, we move to a relational level. In the FOSS world, working on a project implies potential exposure to all other project members' communication about the project, so that capturing ties via collaboration in the same projects is a well-established practice in the literature (e.g., Xu, Christley, and Madey, 2006). Thus, an individual's degree centrality is simply the number of other individuals working on her projects at time t, that is, *N colleagues*$_{it}$. *Brokerage*$_{it}$ captures the brokerage opportunities built in *i*'s ego network at time t measured as the complement to one of Burt's (1992) measures of constraint (Bruggeman, 2008).[4]

The following tables report the descriptions (Table 1), the descriptive statistics (Table 2), and the correlations (Table 3) for the variables described in this section.

[4] In special cases Burt's (1992) measure of constraint can exceed 1 (Bruggeman, 2008). In those cases, we normalized it to 1 as this is the value normally assigned to isolated vertices. Including directly Burt's (1992) measure of constraint without any normalization confirms our results. Notice that both network measures, degree centrality (*N colleagues*$_{it}$), and brokerage opportunities (*Brokerage*$_{it}$) are centered everywhere but in Table 1.

Table 1. Variable description

Variable	Description
Dependent variables	
Projects launched$_{it}$	The number of new projects launched by individual *i* at time *t* in SF.net
Projects joined$_{it}$	The number of existing projects joined by individual *i* at time *t* in SF.net
Independent varibles	
Open-ended communication$_{it}$	The number of messages individual *i*'s colleagues have sent to the forums of SF.net projects in which *i* is not involved at time *t*
Artifact-based communication$_{it}$	The number of artifacts (i.e., bug reports, patches, feature or support requests) individual *i*'s colleagues have submitted to the tracker system of SF.net projects in which *i* is not involved at time *t*
Control variables	
N colleagues$_{it}$	The number of other individuals belonging to the same SF.net projects as *i* at time *t*
Brokerage$_{it}$	Considering collaboration to the same SF.net project as a tie, this captures the brokerage opportunities built in individual *i*'s ego network at time *t*. It is measured as the complement to one of Burt's (1992) measures of constraint (Bruggeman, 2008).
Avg projects registration time$_{it}$	The average registration period of the SF.net projects individual *i* belongs to at time *t*
N projects with GPL$_{it}$	The number of SF.net projects adopting the most restrictive OSS license (the General Public License) among those individual *i* belongs to at time *t*.
Avg projects file releases$_{it}$	The average number of file releases produced by the SF.net projects individual *i* belongs to at time *t*
N forum messages$_{it}$	The number of messages individual *i* has sent to SF.net projects forum at time *t*
N artifacts in track system$_{it}$	The number of artifacts sent by individual *i* to the tracker system of SF.net projects at time *t*
N posted news$_{it}$	The number of new posts sent by individual *i* at SF.net projects' websites at time *t*
N participated projects$_{it}$	The number of SF.net projects to which individual *i* belongs to at time *t*
Dummy in SF staff$_{it}$	A dummy variable equal to 1 when individual *i* belongs to the project managing SF.net itself (i.e., she is a member of SF.net staff) and 0 otherwise.
Tenure$_{it}$	The number of 30-day periods since individual *i*'s registration to the SF.net platform

Endogeneity issues

Endogeneity issues may threaten the validity of this study. We thus employ a series of corrections that reduces this problem. First, we use a fixed-effect panel estimation. Second, we lag the regressors. Third, the way we build our main regressors (*open-ended communication$_{it}$* and *artifact-based communication$_{it}$*) diminishes the incidence of the focal individual on her colleagues' communication, while capturing at the same time the communication she is exposed to (see the previous section). Fourth, we introduce time-variant controls at both the individual and project levels. As the online communication we observe in our data is the main (often the only) type of interaction among developers, our controls capture a great part of the variance connected to such interactions. This increases the ability of our controls to "clean" the regression from possible remaining endogeneity. Fifth, we introduce project-level controls that address the main dimensions that may affect the individual's choice of projects.

These dimensions are chosen on the basis of the extant literature on individuals' motivations to participate in FOSS (e.g., David, Waterman, and Arora, 2003; Von Krogh *et al.* 2012; Lakhani and Wolf, 2005; Shah, 2006): *own use* of the code (e.g., Shah, 2006), *learning* from other developers (von Hippel and von Krogh, 2003), *reputation and signaling* (Lerner and Tirole, 2002), *prosocial-psychosocial* motivations (such as fun, Lakhani and Wolf, 2005; or we-intentions, Bagozzi and Dholakia, 2006, and *idealistic* motivations (such as free software ideology, Elliott and Scacchi, 2003). Projects with many members (*N colleagues$_{it}$*) who are very productive (*Avg projects file releases$_{it}$*) attract individuals interested in finding a lively environment in which others could improve the software they need for their own use, or individuals who are interested in learning from the larger number of possible collaborators and from the larger amount of code, possibly developed for a longer period (*Avg projects registration time$_{it}$*). A lively environment (*N colleagues$_{it}$* and

DOI: 10.1002/smj

Table 2. Descriptive statistics (number of observations = 56,496)

Variable	N	Mean	St. dev.	Median	Min	Max
Dependent variables						
Projects launched$_{it}$	56,496	0.085	0.325	0	0	14
Projects joined$_{it}$	56,496	0.080	0.301	0	0	7
Independent varibles						
Open-ended communication$_{it}$	56,496	1.058	6.720	0	0	260
Artifact-based communication$_{it}$	56,496	1.044	4.436	0	0	106
Control variables						
N colleagues$_{it}$	56,496	7.651	14.499	3	0	190
Brokerage$_{it}$	56,496	0.334	0.351	0.228	0	0.981
Avg projects registration time$_{it}$	56,496	16.702	9.657	18.250	0	38
N projects with GPL$_{it}$	56,496	1.279	1.270	1	0	20
Avg projects file releases$_{it}$	56,496	0.285	0.976	0	0	32.5
N forum messages$_{it}$	56,496	0.399	3.229	0	0	205
N artifacts in track system$_{it}$	56,496	0.286	4.608	0	0	1,030
N posted news$_{it}$	56,496	0.125	0.643	0	0	26
N participated projects$_{it}$	56,496	1.864	1.615	2	0	31
Dummy in SF staff$_{it}$	56,496	0.001	0.031	0	0	1
Tenure$_{it}$	56,496	10.796	6.642	10	1	27

Avg projects file releases$_{it}$) also attracts developers who are psychologically disposed toward the "flow" feelings that coding may bring (Lakhani and Wolf, 2005). Such an environment may also help to foster we-intentions, especially if driven also by an ideologically defined identity (*N projects with GPL$_{it}$*). Productive projects (*Avg projects file releases$_{it}$*) may become attractive for individuals seeking to signal their abilities (Lerner and Tirole, 2002). These individuals may also look for reaching positions in the network of collaboration between project members that allow them to increase their visibility, by means of joining many projects (*N participated projects$_{it}$*) or acquiring a high degree centrality or a high level of brokerage (i.e., *Brokerage$_{it}$* and degree centrality as captured by *N colleagues$_{it}$*). Notice that the controls described above, together with the individual-level controls, also serve as proxies for individuals' motivations, which can be captured in this way even if they change over time.

We thus control for strategic thinking (avoiding reverse causality), project selection (avoiding selection bias), time invariant and variant characteristics (avoiding omitted variable biases, e.g., motivations), and time structure (avoiding simultaneity). This use of controls is adapted from Jeppesen and Rullani (2013), plus additional social network controls, a dimension that may be important to capture in the present study. This allows for an unbiased interpretation of the impact of our main regressors (Hamilton and Nickerson, 2003: 61–62), while it simultaneously preserves the simplicity of a plain one-step Poisson estimation, which is preferable to other, less-direct methods (Stolzenberg and Relles, 1997).

Estimation strategy

We estimate two Poisson regressions with fixed effects and robust standard errors. Over-dispersion is handled using robust standard errors rather than estimating negative binomial models (Hausman, Hall, and Griliches, 1984). If the conditional distribution is wrongly specified, negative binomial models lead to inconsistent estimates of the conditional mean, while Poisson models with robust standard errors lead to consistent estimations of the conditional mean even with overdispersion (Hall and Ziedonis, 2001).

The logic employed by Giuri *et al.* (2010) indicates that the separate estimation of the two equations is empirically justified by the small correlation between *projects joined$_{it}$* and *projects launched$_{it}$* (significant at 1% level, but equal to only 0.12). The reason that there is so little correlation between the two variables arguably has to do with the nature of the empirical context. In FOSS, individuals typically only allocate a limited amount of resources (David and Rullani, 2008; Lakhani and von Hippel, 2003) to coding open-source software in their spare time (David and Shapiro, 2008;

Table 3. Correlation matrix (number of observations = 56,496)

Variable		1	2	3	4	5
Projects launched$_{it}$	1	1.000***				
Projects joined$_{it}$	2	0.135***	1.000***			
Artifact-based communication$_{it}$	3	−0.005	−0.006	1.000***		
Open-ended communication$_{it}$	4	0.000	−0.007*	0.244***	1.000***	
N colleagues$_{it}$	5	−0.023***	−0.041***	0.486***	0.328***	1.000***
Brokerage$_{it}$	6	−0.046***	−0.080***	0.320***	0.211***	0.669***
Avg projects registration time$_{it}$	7	−0.068***	−0.036***	0.062***	0.067***	0.171***
N projects with GPL$_{it}$	8	−0.055***	−0.066***	0.081***	0.117***	0.282***
Avg projects file releases$_{it}$	9	0.014***	0.010**	0.148***	0.060***	0.184***
N forum messages$_{it}$	10	0.004	0.007*	0.016***	0.038***	0.061***
N artifacts in track system$_{it}$	11	0.002	0.020***	0.031***	0.005	0.026***
N posted news$_{it}$	12	0.022***	0.026***	0.009**	0.027***	0.024***
N participated projects$_{it}$	13	−0.059***	−0.061***	0.193***	0.145***	0.417***
Dummy in SF staff$_{it}$	14	0.004	0.001	0.117***	0.099***	0.067***
Tenure$_{it}$	15	−0.108***	−0.094***	0.068***	0.040***	0.192***

Variable		6	7	8	9	10
Brokerage$_{it}$	6	1.000***				
Avg projects registration time$_{it}$	7	0.320***	1.000***			
N projects with GPL$_{it}$	8	0.441***	0.416***	1.000***		
Avg projects file releases$_{it}$	9	0.158***	0.107***	0.030***	1.000***	
N forum messages$_{it}$	10	0.073***	0.037***	0.043***	0.094***	1.000***
N artifacts in track system$_{it}$	11	0.030***	0.015***	0.006	0.040***	0.043***
N posted news$_{it}$	12	0.048***	0.110***	0.066***	0.249***	0.169***
N participated projects$_{it}$	13	0.570***	0.470***	0.709***	0.044***	0.033***
Dummy in SF staff$_{it}$	14	0.051***	−0.003	−0.006	−0.007	0.002
Tenure$_{it}$	15	0.319***	0.310***	0.318***	−0.048***	−0.015***

Variable		11	12	13	14	15
N artifacts in track system$_{it}$	11	1.000***				
N posted news$_{it}$	12	0.076***	1.000***			
N participated projects$_{it}$	13	0.048***	0.113***	1.000***		
Dummy in SF staff$_{it}$	14	0.025***	−0.003	0.033***	1.000***	
Tenure$_{it}$	15	−0.012***	−0.036***	0.394***	−0.008**	1.000***

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

David *et al.*, 2003). This suggests that the two processes of launching and joining only seldom compete for an individual's resources, and thus, do not show much correlation.

The aim is to predict the number of new projects a focal individual launches in the course of a month and the number of existing projects the same individual joins in the same period of time. As we use a fixed-effect model, all individuals with no variance in these two independent variables are automatically removed from the analysis, as they contribute no information to the estimation. Thus, data for 3,040 individuals are utilized for the purpose of the estimation. Each individual is observed from the month of entry into SF.net as a registered individual to December 2002. This results in an unbalanced panel spanning a maximum of 28 30-day periods and composed of a total of 56,496 observations. We aggregate our variables by month.[5]

[5] David and Rullani (2008) analyze data from the same source and develop a time-series analysis that detects the length of the cycles governing the number of projects launched (or joined) over time by SF.net participants. Their analysis shows that the shortest common time window among all of the detected cycles is 30 days, which supports our choice. In addition, we note that the joining and launching of projects are rare. If the window of observation is too short, randomness increases (if a project is launched this minute or the next is mainly a matter of chance). If the window is too large, we pool events together that should be distinguished from each other in terms of timing. A 30-day period seems a

DOI: 10.1002/smj

Table 4. Poisson regression models (fixed effect, robust standard errors)

Variable	Model 1 Projects launched$_{it}$	Model 2 Projects joined$_{it}$	Model 3 Projects launched$_{it}$	Model 4 Projects joined$_{it}$
Open-ended communication$_{(it-1)}$	0.006***		0.006***	0.001
	(0.002)		(0.002)	(0.003)
Artifact-based communication$_{(it-1)}$		0.018***	0.004	0.017***
		(0.004)	(0.006)	(0.004)
N colleagues$_{(it-1)}$	0.004*	0.001	0.004	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Brokerage$_{(it-1)}$	0.766***	−2.111***	0.761***	−2.110***
	(0.212)	(0.156)	(0.213)	(0.156)
Avg projects registration time$_{(it-1)}$	−0.032***	0.008**	−0.032***	0.008**
	(0.005)	(0.003)	(0.005)	(0.003)
N projects with GPL$_{(it-1)}$	−0.221**	−0.281***	−0.220**	−0.282***
	(0.096)	(0.066)	(0.096)	(0.066)
Avg projects file releases$_{(it-1)}$	0.052***	0.038**	0.051***	0.039**
	(0.018)	(0.016)	(0.018)	(0.016)
N forum messages$_{(it-1)}$	−0.001	0.008	−0.001	0.008
	(0.006)	(0.005)	(0.006)	(0.005)
N artifacts in track system$_{(it-1)}$	0.001	0.003***	0.001	0.003***
	(0.001)	(0.000)	(0.001)	(0.000)
N posted news$_{(it-1)}$	0.034	0.125***	0.035	0.125***
	(0.046)	(0.024)	(0.046)	(0.024)
N participated projects$_{(it-1)}$	−0.454***	−0.189***	−0.454***	−0.188***
	(0.123)	(0.065)	(0.123)	(0.065)
Dummy in SF staff$_{(it-1)}$	0.696	0.482**	0.656	0.449*
	(0.899)	(0.233)	(0.897)	(0.239)
Tenure$_{(it-1)}$	0.031***	0.045***	0.031***	0.045***
	(0.006)	(0.005)	(0.006)	(0.005)
Log likelihood	−11,345.628	−10,757.365	−11,345.152	−10,757.241
LR $\chi 2$	483.709	1, 456.446	494.568	1, 483.342
Prob > $\chi 2$	0	0	0	0
Obs. (individuals)	56,496 (3,040)	56,496 (3,040)	56,496 (3,040)	56,496 (3,040)

***p < 0.01; **p < 0.05; *p < 0.1.

Results

Table 4 reports our results of testing Hypotheses 1 and 2 separately; first in Models 1 and 2, and then jointly in Models 3 and 4. None of the hypotheses are rejected. The coefficients of our main regressors (respectively, in Models 1, 3, and 4 for *Open-ended communication*$_{(it-1)}$ and in Models 2–4 for *Artifact-based communication*$_{(it-1)}$) are all positive and significant, confirming the positive effect the exposure to open-ended communication has on project launching, and the exposure to artifact-based communication has on project joining.

With respect to Hypotheses 1 and 2, it is interesting to notice that in Model 3 (relative to the number of new projects launched), artifact-based formal communication is not significant, and a similar finding is identified for open-ended communication in the context of project joining (Model 4).

We advance a more nuanced understanding of the effects by elaborating on our results. We first transform the coefficients of our regressions into Incidence Rate Ratios (IRRs). IRRs represent the percentage increase in the expected number of projects launched (for Hypothesis 1) or joined (in Hypothesis 2) for a unit change in the regressor of interest. Using IRRs allow us to study the effect we have found along two dimensions: the intensity and the duration of the stimulus.

As the IRR represents the stimulus received by the dependent variable when the regressor varies by one unit, we can vary the intensity of the stimulus simply by multiplying the IRR by an

reasonable interval to balance these two tensions and also provides enough observation periods.

intensity parameter. In the literature (e.g., Zhang *et al.*, 2010), this parameter is usually represented by 1 (or more, i.e., N) standard deviations of the regressor σ above its mean. Considering more than one deviation is particularly meaningful in FOSS where the level of activity and communication is quite skewed (David and Rullani, 2008; Krishnamurthy, 2002; Kuk, 2006), resulting in (many) individuals surrounded by poor interaction and other (few) individuals operating in a very communicative and lively environment (Rullani and Haefliger, 2013). Focusing on the most productive part of the community means therefore focusing on the latter, and thus, evaluating what happens when communication around an individual is raised well above the mean.

As for the duration of the stimulus, we move from the need to be explicit about the relationship between the construct (i.e., individual i's exposure to a flow of interactions between collaborators) and the proxies we use to capture that construct (i's collaborators' artifact or forum message exchanges, which is a series of distinct events), and thus, transform our estimates based on the occurrence of precise events into estimates that capture the continuous flow of interaction surrounding the individual over T periods. A continuous stimulus for T periods can be obtained by the product of the IRRs for each different period $t = 1, t = 2, \dots, t = \mathrm{T}$.

This stimulus is considered subject to forces that diminish its marginal impact on the dependent variable as its intensity and duration increase. While most of these forces have been controlled for in our regressions, individuals may still face time constraints or information overload or other factors that above a certain threshold inevitably limit the effect of the stimulus they are exposed to. In other words, it is unrealistic to assume that they can join or launch projects infinitely. We account for this applying a weakening factor w to the effect of each increase in stimulus' intensity and duration.

These considerations can be merged into an indicator I_w expressing the overall percentage impact on the expected number of founded (joined) projects when open-ended communication (artifact-based communication) increases. I_w is obtained by the combination of two functions, one for the stimulus intensity *g(.)* and the other one for its duration *f(.)* (see the Appendix for the formula defining I_w and its derivation):

$$I_w = g\{\sigma; N; w\} \cdot f\{IRR; T; w\}.$$

We use I_w to perform a sensitivity analysis on our results. The following Tables 5 and 6 report the values of I_w for N and T that raise from 1 to 3, and when $w = 10\%$ and $w = 50\%$. This is done considering that IRR = 1.006 for *open-ended communication*$_{it}$ for both models relative to *projects launched*$_{it}$, while IRR = 1.018 for *artifact-based communication*$_{it}$ for both models relative to *projects joined*$_{it}$.

Our sensitivity analysis shows that when the weakening factor is 10 percent (i.e., when each increase in N or T cuts the effect of the stimulus by 10%), after three months of continuous exposure to open-ended communication an average individual is expected to launch 11 percent more new projects (Table 5, left panel). The effect is similar for those who perceive a stimulus three times more intense ($N = 3$) but for only one month. When the two effects are combined, the percentage rises to 30 percent. The same computation leads to percentages that are smaller, but still very relevant, even when the weakening factor is raised up to 50 percent (Table 5, right panel): 7 percent for $N = 3$ or $T = 3$, and 12 percent when $N = 3$ and $T = 3$. Moreover, irrespective of the level set for w, all these percentages are doubled for the effect of artifact-based communication on the expected increase in projects joining (Table 6).

In sum, for individuals exposed for a long period to very dense open-ended communication the expected number of projects lunched increases by one third, and the effect remains very relevant (more than 10%) even when $w = 50\%$. Prolonged exposure to dense artifact-base communication has an even stronger effect: it increases the expected number of project joined by an individual by two-thirds when $w = 10\%$ and by one-fourth when $w = 50\%$. This provides an idea of the importance of the mechanism we detect in this study, including its intensity and duration.

CONCLUDING DISCUSSION

Contributions to theory

The problem-solving perspective has emerged as an important analytical lens for linking dynamic phenomena in management research, such as innovation and learning, to economic organization and forms of governance (e.g., Macher, 2006; Macher and Boerner, 2012; Nickerson and Zenger, 2004).

Table 5. Impact I_w of open-ended communication on expected number of project launched over T periods for N standard deviations with a weakening factor *w* equal to 10% (left) and 50% (right)

w = 10%	N = 1 (%)	N = 2 (%)	N = 3 (%)	*w* = 50%	N = 1 (%)	N = 2 (%)	N = 3 (%)
T = 1 month	4	8	11	T = 1 month	4	6	7
T = 2 months	8	15	21	T = 2 months	6	9	11
T = 3 months	11	21	30	T = 3 months	7	11	12

IRR = 1.006; $\sigma = 6.7$.

Table 6. Impact I_w of artifact-based communication on expected number of project joined over T periods for N standard deviations with a weakening factor *w* equal to 10% (left) and 50% (right)

w = 10%	N = 1 (%)	N = 2 (%)	N = 3 (%)	*w* = 50%	N = 1 (%)	N = 2 (%)	N = 3 (%)
T = 1 month	8	15	22	T = 1 month	8	12	14
T = 2 months	15	29	41	T = 2 months	12	18	21
T = 3 months	22	42	60	T = 3 months	14	21	25

IRR = 1.018; $\sigma = 4.436$.

A "problem" is arguably a more natural unit of analysis for approaching such dynamic phenomena than, for example, the transaction, because a problem inherently contains a fundamental dynamic aspect—it is forward-looking, in need of a (future) solution. As such, the problem-solving perspective is well positioned to address issues that are crucial in strategy research related to the creation of a mental space of possible strategic opportunities and how firms undertake search in this space. However, the problem-solving perspective has thus far mainly concentrated on matching problems with their solutions in terms of the governance forms (i.e., firms and markets) that can best coordinate this matching process. In contrast, problem-formulation has only recently been brought into the purview of this perspective (Baer *et al.*, 2013), and the perspective has only recently been extended beyond the firm/market dichotomy (e.g., Felin and Zenger, 2011, 2014). Our study contributes by extending the problem-solving perspective in important ways.

First, we theorize about problem-formulation behaviors and their correlates, and we empirically operationalize these in a self-organizing context (Anderson, 1999) in terms of considering project-related behaviors and different kinds of communication as correlates. Second, classic organizational research (e.g., March and Simon, 1958; Marschak and Radner, 1972) points to the importance of communication for understanding organizational forms and performance. We theorized that individuals' different project-related behaviors are influenced by their exposure to different modes of communication in their community context. Specifically, we hypothesized that open-ended communication predominantly relates to problem formulation, that is, project launching, while artifact-based informal communication predominantly relates to project joining, that is, focusing on solving or adding a subproblem. Our panel data set derived from SourceForge.net and the technique we used allowed us to test our two hypotheses. These were supported by the data. Third, we extend the scope of the problem-solving perspective to communities. This organizational form and the behaviors driving it have fascinated scholars for more than a decade. In particular, the motivational antecedents of innovative and sharing behaviors in communities have been extensively addressed (e.g., Frey *et al.*, 2012; Lakhani and Wolf, 2005; Lerner and Tirole, 2002). In all, our study extends the problem-solving perspective by examining behaviors that proxy for problem-formulation behaviors in the empirical setting of the community form of FOSS and by providing detail on the communication-based correlates of such behaviors. Our findings extend previous research on communities in general and on FOSS in particular by showing the importance of communication-based factors for understanding project-related behaviors and for indicating that attention allocation may also play a key role in understanding such project-related behaviors. We thus link the focus on problem-solving and attention in Simon's (1987) to a specific empirical setting. In the context of understanding communities such as FOSS, we extend the motivation-centric focus of previous research in this field in a more knowledge-based direction. Our study does not imply a rejection of motivational antecedents in any way (as we have not posited motivational and cognitive antecedents as theoretical rivals), but rather suggests that future research should take communication-based antecedents in the community context into account. Indeed, basic motivation-opportunity-ability theory (Blumberg and Pringle, 1982) suggests that the two sets of determinants of behavior should be considered jointly, which is a challenge for future research in this field.

In seeking to understand project-related behaviors in more traditional organizational forms, researchers may benefit from including the antecedents of the project-related behaviors we have identified. This element adds to extant research on project-related behaviors that focuses on other antecedents, such as project longevity (e.g., Katz, 1982) or formal coordination mechanisms (e.g., Foss, 2003; Levitt *et al.*, 1999). Our study suggests that such antecedents may matter precisely because they influence the modes of communication that are adopted (e.g., project longevity drives the choice of certain mode of communication over time; formal coordination mechanisms place constraints on the mode of communication that organizational that members can choose).

Limitations and future research

The contributions of this study need to be considered in light of its limitations. While we can distinguish among different kinds of project-related behaviors, our data do not allow us to offer insight into the relative importance of such behaviors for overall community evolution and dynamics. While project launches introduce variation into communities, it is joining behaviors and the associated problem-solving that realize projects. Both matter for the performance and development of a community. However, communities may arguably exhibit too much exploration or too much exploitation (David and Rullani, 2008). Our data do not allow us to investigate this possibility. Therefore, we cannot offer direct insight into how the community-level exploitation/exploration tradeoff is influenced by the different modes of communication in the community.

Future research addressing this issue may compare different communities in terms of how they balance exploration and exploitation, and what are the observed innovation outcomes associated with this pattern. Alternatively, it may examine time series from the same community that exhibit different patterns of exploration and exploitation and examine the resulting innovation outcomes. However, such research, if undertaken as large-N research efforts, is inherently difficult to carry out, notably because of data availability problems. At the individual level, a connected interesting study could focus on individuals' specialization, posing questions such as: Do individuals specialize in one behavior (either problem formulation of solving), or do they do both to the same degree? And what are the behavioral antecedents of such specialization (or lack of)?

A final note on the empirical side of our work. While we strived to avoid endogeneity, no technique other than a controlled experiment can fully assure causality. This should be kept in mind not only in the present study, but also in future studies in the research domain of problems and organizational dynamics.

Managerial implications

Problem-formulating behaviors in communities are likely to differ from those in bureaucratic organizational forms because employees have less discretion with respect to deciding which problems to work on than individuals in a community form. Communities and those project-based firms that create considerable room for self-organization (Anderson, 1999; Felin and Zenger, 2011; Foss, 2003) remain notably different governance forms. In particular, the presence of fiat in firms, but not in communities, is a decisive difference (Williamson, 1996). Nevertheless, firms and communities share important commonalities. According to Anderson (1999: 228), in self-organizing organizational forms "managers establish and modify the direction and the boundaries within which effective, improvised, self-organized solutions can evolve … [and] tune the system by altering the constraints."

While we have addressed the antecedents of project-related behaviors in a nonfirm form, our study potentially holds implications for our understanding of how management may influence such behaviors in firms "by altering the constraints." Our analysis suggests that firms that need to concentrate on more incremental problem-solving efforts (e.g., because a sufficient number of attractive problems have already been defined) should create environments in which interaction is undertaken mainly *via* artifacts. On the other hand, if firms seek to generate new problems, they should seek to create environments in which open-ended, verbal conversation is relatively more important than artifact-based communication. Seen in this light, one advantage of firms relative to communities in the context of problem-formulation may be that the former more easily can establish the relevant proportion of different communication genres because of their access to managerial fiat. Thus, our study could potentially be extended to a comparative examination of organizational forms in terms of their access to and control over different modes of communication, and hence, the incidence and proportion of different project-related behaviors.

Finally, our study holds potential implications for understanding how communication plays into the generation of a space of strategic opportunities and the search across such a space. Kaplan (2011) explores how a particular communication mode (namely, PowerPoint) influences strategy-making. However, her study is not directly comparative, that is, she does not directly compare this mode of communication with other available modes. Our study suggests that a comparative study of alternative ways of communicating in the strategy process may be fruitful. Thus, we focus on two modes of communication, namely, if online dialogue and communication about new project ideas are free and unconstrained, or whether they are more constrained and revolve around existing artifacts. Our findings may imply that the former kind of search is more likely to be associated with coming up with new strategic initiatives, while the latter kind of search is associated with more incremental initiatives. However, future research needs to uncover whether these findings, obtained in the context of the FOSS community, also generalize to other organizational forms.

ACKNOWLEDGEMENTS

We thank Gautam Ahuja, Oliver Alexy, Carliss Baldwin, Stefano Brusoni, Lorenzo Cassi, Guido Conaldi, Linus Dahlander, Paul David, Alfonso Gambardella, Paolo Giordani, Bronwyn Hall, Joachim Henkel, Christoph Hienerth, Siobhàn O'Mahony, Alessandro Nuvolari, Woody Powell, Francesco Rentocchini, Christina Rossi Lamastra, Sebastian Spaeth, Toke Reichstein, Frederik Tell, Lorenzo Zirulia, and seminar participants at Linköping University, EGOS, the SEI workshop at ETH in Zurich, and AoM for helpful comments on, and discussions of, ideas unfolded in this study. We thank Lorenzo Cassi for his help with establishing our network measures and Salvatore Torrisi, Paola Giuri, Gaia Rocchetti, Matteo Ploner, VeLUG and PLUTO for having made our data collection possible. We thank the SourceForge.net staff for having provided our group with the data.

REFERENCES

Afuah A, Tucci C. 2012. Crowdsourcing as a solution to distant search. *Academy of Management Review* **37**: 355–375.

Anderson P. 1999. Complexity theory and organization science. *Organization Science* **10**: 216–232.

Baer M, Dirks KT, Nickerson JA. 2013. Microfoundations of strategic problem formulation. *Strategic Management Journal* **34**: 197–214.

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

Baldwin C, Clark K. 2003. Managing in an age of modularity. In *Managing in the Modular Age: Architectures, Networks, and Organizations*, Garud R, Kumaraswamy A,, Langlois RN (eds). Blackwell: Oxford, UK; 149–171.

Benkler Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press: New Haven, CT.

Blumberg M, Pringle CD. 1982. The missing opportunity in organizational research: some implications for a theory of work performance. *Academy of Management Review* **7**: 560–569.

Bly S, Harrison S, Irwin S. 1993. Media spaces: bringing people together in a video, audio and computing environment. *Communications of the Association of Computing Machinery* **36**: 28–46.

Bonaccorsi A, Giannangeli S, Rossi C. 2006. Entry strategies under competing standards: hybrid business models in the Open Source software industry. *Management Science* **52**: 1085–1098.

Bruggeman J. 2008. *Social Networks: An Introduction*. Routledge: New York.

Burt RS. 1992. *Structural Holes. The Social Structure of Competition*. Harvard University Press: Cambridge, MA.

Carlile PR. 2002. A pragmatic view of knowledge and boundaries: boundary objects in new product development. *Organization Science* **13**: 442–455.

Carstensen P, Nielsen M. 2001. Characterizing modes of coordination: a comparison between oral and artifact based coordination. *International Proceedings of Group* **1**: 81–90.

Comino S, Manenti FM, Parisi ML. 2007. From planning to mature: on the success of open source projects. *Research Policy* **36**(10): 1575–1586.

Daft RL, Lengel RH. 1986. Organizational information requirements, media richness and structural design. *Management Science* **32**(5): 554–571.

Dahlander L, O'Mahony S. 2011. Progressing to the center: coordinating project work. *Organization Science* **22**(4): 961–979.

Dahlander L, Wallin MW. 2006. A man on the inside: unlocking communities as complementary assets. *Research Policy* **35**: 1243–1259.

David PA, Rullani F. 2008. Dynamics of innovation in an "open source" collaboration environment: lurking, laboring and launching FLOSS projects on SourceForge. *Industrial and Corporate Change* **17**: 647–710.

David PA, Shapiro JS. 2008. Community-based production of open source software: what do we know about the developers who participate? *Information Economics and Policy* **20**: 364–398.

David PA, Waterman A, Arora AS. 2003. The free/libre/open source software survey for 2003. Preliminary draft, September 2003, quoted with authors' permission. Available at: http://www.stanford.edu/group/floss-us/report/FLOSS-US-Report.pdf (accessed 1st May 2015).

Delmar F, Wennberg K, Hellerstedt K. 2011. Endogenous growth through knowledge spillovers in entrepreneurship: an empirical test. *Strategic Entrepreneurship Journal* **5**: 199–226.

Denrell J, Fang C, Winter SG. 2003. The economics of strategic opportunity. *Strategic Management Journal* **24**: 977–990.

Elliott MS, Scacchi W. 2003. Free software developers as an occupational community: resolving conflicts and fostering collaboration. In *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work*. ACM; 21–30.

Felin T, Zenger T. 2011. Information aggregation, matching, and radical market-hierarchy hybrids: implications for the theory of the firm. *Strategic Organization* **9**(4): 163–173.

Felin T, Zenger T. 2014. Closed or open innovation? Problem solving and the governance choice. *Research Policy* **43**: 914–925.

Fleming LD, Waguespack M. 2007. Brokerage, boundary spanning, and leadership in open innovation communities. *Organization Science* **18**: 165–180.

Fosfuri A, Giarratana M, Luzzi A. 2008. The penguin has entered the building: the commercialization of open source software products. *Organization Science* **19**: 292–305.

Foss NJ. 2003. Selective intervention and internal hybrids: interpreting and learning from the rise and decline of the Oticon spaghetti organization. *Organization Science* **14**: 331–349.

Foss NJ, Klein PG. 2012. *Organizing Entrepreneurial Judgment*. Cambridge University Press: Cambridge, UK.

Frey B, Luethi R, Osterloh M. 2012. Community enterprises: an institutional innovation. *Managerial and Decision Economics* **33**: 427–439.

Gann DM, Salter AJ. 2000. Innovation in project-based, service-enhancing firms in the construction of complex products and systems. *Research Policy* **29**: 955–972.

Gavetti G, Levinthal D. 2000. Looking forward and looking backward: cognitive and experiential search. *Administrative Science Quarterly* **45**: 113–137.

Giuri P, Ploner M, Rullani F, Torrisi S. 2010. Skills, division of labor and performance in collective inventions: evidence from open source software. *International Journal of Industrial Organization* **28**: 54–68.

Giuri P, Rullani F, Torrisi S. 2008. Explaining leadership in open source software projects. *Information Economics and Policy* **20**: 305–315.

Grant RM. 1996. Prospering in dynamically-competitive environments: organizational capability as knowledge integration. *Organization Science* **7**(4): 375–387.

Guetzkow H, Simon HA. 1955. The impact of certain communication nets upon organization and performance in task-oriented groups. *Management Science* **1**: 233–250.

Haefliger S, von Krogh G, Spaeth S. 2008. Code reuse in open source software. *Management Science* **54**: 180–193.

Hall BH, Ziedonis RH. 2001. The patent paradox revisited: an empirical study of patenting in the U.S. semiconductor industry, 1979–1995. *RAND Journal of Economics* **32**: 101–128.

Hamilton BH, Nickerson JA. 2003. Correcting for endogeneity in strategic management research. *Strategic Organization* **1**: 51–78.

Hausman JA, Hall BH, Griliches Z. 1984. Econometric methods for count data with an application to the patents—R and D relationship. *Econometrica* **52**: 909–938.

Hayek FA. 1945. The use of knowledge in society. *American Economic Review* **35**: 319–330.

Heiman B, Nickerson J, Zenger T. 2009. Governance knowledge creation: a problem-finding and problem-solving perspective. In *Knowledge Governance: Processes and Perspectives*, Foss NJ, Michailova S (eds). Oxford University Press: Oxford, UK; 25–46.

Henderson RM, Clark KB. 1990. Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* **35**: 9–30.

Holmquist LE, Mattern F, Schiele B, Alahuhta P, Beigl M, Gellersen H-W. 2001. Smart-its friends: a technique for users to easily establish connections between smart artifacts. *Ubiquitous Computing, Lecture Notes in Computer Science* **2201**: 116–122.

Howison J, Crowston K. 2004. The perils and pitfalls of mining Sourceforge. In *Proceedings of Workshop on Mining Software Repositories at the International Conference on Software Engineering ICSE*, Edinburgh, UK.

Jeppesen LB, Lakhani KR. 2010. Marginality and problem solving effectiveness in broadcast search. *Organization Science* **21**: 1016–1033.

Jeppesen LB, Rullani F. 2013. Attention: a new perspective on contributions to open collaborative-innovation communities, Mimeo.

Kaplan S. 2011. Strategy and PowerPoint: an inquiry into the epistemic culture and machinery of strategy making. *Organization Science* **22**: 320–346.

Katz R. 1982. The effects of group longevity on project communication and performance. *Administrative Science Quarterly* **27**: 81–104.

Kauffman SA. (1993). *The origins of order: Self organization and selection in evolution*. Oxford university press: New York.

Kogut B, Zander U. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science* **3**(3): 383–397.

Krishnamurthy S. 2002. Cave or community? An empirical examination of 100 mature open source projects. *First Monday* **7**(6). http://ojphi.org/ojs/index.php/fm/article/view/960/881 (accessed 1st May 2015).

Kuk G. 2006. Strategic interaction and knowledge sharing in the KDE developer mailing list. *Management Science* **52**(7): 1031–1042.

Kuznetsov S. 2006. Motivations of contributors to Wikipedia. *ACM SIGCAS Computers and Society* **36**(2): 1–7.

Lakhani KR, von Hippel E. 2003. How open source software works: "Free" developer-to-developer assistance. *Research Policy* **32**: 923–943.

Lakhani KR, Jeppesen LB, Lohse PA, Panetta JA. 2007. The value of openness in scientific problem solving. HBS Working paper 07-050, Harvard University, Cambridge, MA.

Lakhani KR, Wolf RG. 2005. Why hackers do what they do: understanding motivations and effort in free/open source software projects. In *Perspectives on Free and Open Source Software*, Feller J, Fitzgerald B, Hissam S, Lakhani KR (eds). MIT Press: Cambridge, MA; 3–22.

Lerner J, Tirole J. 2002. Some simple economics of open source. *Journal of Industrial Economics* **50**: 197–234.

Lerner J, Tirole J. 2005. The scope of open source licensing. *Journal of Law, Economics, and Organization* **21**: 20–56.

Lessig L. 2001. *The Future of Ideas: The Fate of the Commons in a Connected World*. Random House: New York.

Levitt RE, Thomsen J, Christiansen TR, Kunz JC, Jin Y, Nass C. 1999. Simulating project work processes and organizations: toward a micro-contingency theory of organizational design. *Management Science* **45**: 1479–1495.

Lin Y. 2004. Epistemologically multiple actor-centered systems: or, EMACS at work! In *Paper Presented at the Oekonux Conference*, Vienna, Austria, 23–25 May. Available at: http://ubiquity.acm.org/article.cfm?id=985611 (accessed 1st May 2015).

Lyytinen K, Newman M. 2008. Explaining information systems change: a punctuated socio-technical change model. *European Journal of Information Systems* **17**: 589–613.

Macher JT. 2006. Technological development and the boundaries of the firm: a knowledge-based examination in semiconductor manufacturing. *Management Science* **52**: 826–843.

Macher JT, Boerner C. 2012. Technological development at the boundaries of the firm: a knowledge-based examination in drug development. *Strategic Management Journal* **33**: 1016–1036.

March JG, Simon HA. 1958. *Organizations*. Wiley: New York.

Marschak J, Radner R. 1972. *Economic Theory of Teams*. Yale University Press: New Haven, CT.

Mintzberg H, Raisinghani D, Theoret A. 1976. The structure of unstructured decision processes. *Administrative Science Quarterly* **21**: 246–275.

Monge PR, Contractor NS. 2003. *Theories of Communication Networks*. Oxford University Press: Oxford, UK.

Nickerson J, Silverman B, Zenger T. 2007. The "problem" of creating and capturing value. *Strategic Organization* **5**: 211–225.

Nickerson J, Yen CJ, Mahoney JT. 2012. Exploring the problem-finding and problem-solving approach for designing organizations. *Academy of Management Perspectives* **26**: 52–72.

Nickerson J, Zenger T. 2004. A knowledge-based theory of the firm: the problem-solving perspective. *Organization Science* **15**: 617–632.

O'Mahony S. 2007. The governance of open source initiatives: what does it mean to be community managed? *Journal of Management and Governance* **11**(2): 139–150.

O'Mahony S, Bechky B. 2008. Boundary organizations: enabling collaboration between unexpected allies. *Administrative Science Quarterly* **43**: 422–459.

O'Mahony S, Ferraro F. 2007. The emergence of governance in an open source community. *Academy of Management Journal* **50**: 1079–1106.

O'Mahony S, Lakhani KR. 2011. Organizations in the shadow of communities. In *Communities and Organizations: Research in the Sociology of Organizations*, Marquis C, Lounsbury M, Greenwood R (eds). Emerald Group Publishing Limited: Bingley, UK; 3–36.

Orlikowski WJ, Yates J. 1994. Genre repertoire: the structuring of communicative practices in organizations. *Administrative Science Quarterly* **39**: 541–574.

Popper KR. (1972). *Objective knowledge: An evolutionary approach*. Oxford University Press: New York.

Powell WW, Hwang H. 2009. The rationalization of charity: the influences of professionalism in the nonprofit sector. *Administrative Science Quarterly* **54**: 268–298.

Raymond E. 1998. Homesteading the noosphere. *First Monday* **3**(10). Available at: http://www.firstmonday.dk/issues/issue3_10/raymond/#d12 (accessed June 2010).

Rullani F, Haefliger S. 2013. The periphery on stage. Properties, functions, and dynamics of the periphery in the Free/Open Source Software community. *Research Policy* **42**(4): 941–953.

Sapsed J, Salter A. 2004. Postcards from the edge: local communities, global programs and boundary objects. *Organization Studies* **25**: 1515–1534.

Shah S. 2006. Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science* **52**: 1000–1014.
Simon HA. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* **156**: 467–482.
Simon HA. 1973. The structure of ill-structured problems. *Artificial Intelligence* **4**: 181–201.
Simon HA. 1987. Making management decisions: the role of intuition and emotion. *Academy of Management Executive* **1**: 57–64.
Simon HA. 1991. Organizations and markets. *Journal of Economic Perspectives* **5**: 25–44.
Sojer M, Henkel J. 2010. Code reuse in open source software development: quantitative evidence, drivers, and impediments. *Journal of the Association for Information Systems* **11**: 868–901.
Stacey RD. 1995. The science of complexity: an alternative perspective for strategic change processes. *Strategic Management Journal* **16**: 477–495.
Star S, Griesemer J. 1989. Institutional ecology, "translations" and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–1939. *Sociological Studies of Science* **19**: 387–420.
Staudenmayer NA, Cusumano MA. (1998). Alternative designs for product component integration. Working Paper, MIT Sloan working paper series #4016, Sloan School of Management, Massachusetts Institute of Technology.
Stolzenberg RM, Relles DA. 1997. Tools for intuition about sample selection bias and its correction. *American Sociological Review* **62**: 494–507.
von Hippel E, von Krogh G. 2003. Open source software and the "private-collective" innovation model: issues for *Organization Science*. *Organization Science* **14**: 209–223.
von Hippel E, von Krogh G. 2006. The promise of research on open source software. *Management Science* **52**: 975–983.
von Hippel E, von Krogh G. Forthcoming. Identifying viable 'need-solution pairs': problem solving without problem formulation. *Organization Science*.
von Krogh G, Haefliger S, Spaeth S, Wallin MW. 2012. Carrots and rainbows: motivation and social practice in open source software development. *MIS Quarterly* **36**(2): 649–676.
von Krogh G, Späth S, Lakhani K. 2003. Community, joining, and specialization in open source software innovation: a case study. *Research Policy* **32**: 1217–1241.
Vyas D, Dix A, Nijholt A. 2008. Role of artifacts in mediated communication. In *CHI conference*, Florence, Italy. Available at: http://www.chi2008.org/altchisystem/submissions/submission_dhaval_vyas_1.pdf (accessed May 2015).
Wasko M, Faraj S. 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly* **29**(1): 35–57.
West J, O'Mahony S. 2008. Designing a participation architecture to support firm-community collaboration. *Industry and Innovation* **15**(2): 145–168.
Williamson OE. 1996. *The Mechanisms of Governance*. Oxford University Press: Oxford, UK.
Woods DD. 1998. Designs are hypotheses about how artifacts shape cognition and collaboration. *Ergonomics* **41**: 168–173.
Xu J, Christley S, Madey G. 2006. Application of social network analysis to the study of open source software. In *The Economics of Open Source Software Development*, Bitzer J, Schröder PJH (eds). Elsevier: North-Holland, MI.
Zhang Y, Li H, Li Y, Zhou L-A. 2010. FDI spillovers in an emerging market: the role of foreign firms' country origin diversity and domestic firms' absorptive capacity. *Strategic Management Journal* **31**: 969–989.
Zmud RW, Lind MR, Young FW. 1990. An attribute space for organizational communication channels. *Information Systems Research* **1**: 440–457.

APPENDIX

We aim at deriving *g(.)* and *f(.)* in the equation for I_w:

$$I_w = g\,\{\sigma; N; w\} \cdot f\,\{IRR; T; w\}$$

We start by deriving *g(.)* on the basis of the evolution of the effect of σ for each subsequent increase in $n = 1,.., N$:

$$\begin{aligned}
\sigma_{n=1} &= \sigma \\
\sigma_{n=2} &= \sigma_{n=1} \cdot (1-w) = \sigma\,(1-w) \\
\sigma_{n=3} &= \sigma_{n=2} \cdot (1-w) \\
&= \overbrace{[\sigma\,(1-w)]}^{\sigma_{n=2}}\,(1-w) = \sigma\,(1-w)^2 \\
&\dots \\
\sigma_{n=N} &= \sigma\,(1-w)^{N-1}
\end{aligned}$$

Since *g(.)* captures the cumulated effect, we just need to computed the sum of the drifts from the mean, i.e., as the sum of the (weakened) standard deviations from the mean:

$$g\,\{\sigma; N; w\} = \sigma\left[\sum_{n=1}^{N}(1-w)^{n-1}\right]$$

Second, we derive *f(.)* on the basis of the evolution of the effect of the *IRR* for each subsequent increase in $t = 1, \dots, T$.

When positive, the effect of the regressor on the dependent variable is captured by the part of the *IRR* that exceeds one. Thus, the weakening factor must be applied only to that part. Consequently:

$$IRR_{T=1} = IRR$$

$$IRR_{T=2} = \left[\left(IRR_{t=1} - 1\right)(w-1) + 1\right]$$

$$= [(IRR-1)(1-w)+1]$$

$$IRR_{T=3} = \left[\left(IRR_{t=2} - 1\right)(1-w) + 1\right]$$

$$= \left[\left(\overbrace{\{(IRR-1)(1-w)+1\}}^{IRR_{t=2}} - 1\right)(1-w) + 1\right]$$

$$= \left[(IRR-1)(1-w)^2 + 1\right]$$

$$\ldots$$

$$IRR_T = \left[(IRR-1)(1-w)^{T-1} + 1\right]$$

Since *f(.)* captures the cumulated effect, and *IRR* from different periods must be multiplied to be cumulated, we just need to compute the product of the (weakened) *IRR* for to different periods, and then subtract 1:

$$f\{IRR; T; w\} = \prod_{t=1}^{T} \left[(IRR-1)(1-w)^{t-1} + 1\right] - 1$$

Given the equations above, we can define I_w as:

$$I_w = \underbrace{\sigma \left[\sum_{n=1}^{N} (1-w)^{n-1}\right]}_{g\{\sigma;N;w\}} \cdot \underbrace{\left\{\prod_{t=1}^{T} \left[(IRR-1)(1-w)^{t-1} + 1\right] - 1\right\}}_{f\{IRR;T;w\}}$$

Notice incidentally that when w=0% (the effect of the stimulus is *not* weakened as N and T grow) we obtain a much more intuitive formulation for the equation above:

$$I = [\sigma N] \cdot \left[IRR^{\mathrm{T}} - 1\right]$$