Where and When Can Open Source Thrive?
Towards a Theory of Robust Performance

Sheen S. Levine[1] and Michael J. Prietula[2]

[1]Singapore Management University
50 Stamford Road
Singapore 178899
sslevine@sslevine.com

[2]Emory University
Atlanta, GA 30322
United States of America
prietula@bus.emory.edu

Abstract. While the economic impact of, and the interest in, open source innovation and production has increased dramatically in recent years, there is still no widely accepted theory explaining its performance. We combine original fieldwork with agent-based simulation to propose that the performance of open source is surprisingly robust, even as it happens in seemingly harsh environments with free rider, rival goods, and high demand. Open source can perform well even when cooperators constitute a minority, although their presence reduces variance. Under empirically realistic assumptions about the level of cooperative behavior, open source can survive even increased rivalry and performance can thrive if demand is managed. The plausibility of the propositions is demonstrated through qualitative data and simulation results.

Keywords: Innovation, Exchange, Performance, Agent-based Modeling

1 Introduction

Interest in open source is booming. With its roots in freely shared software [1], the term has been expanding to include broader instances of product development [2], process innovation [3], and knowledge exchange among end-users [4].[1] All in myriad fields, including technology, science [7], medicine [10], and law [11], among others.

Open source software has become a viable alternative to commercial software. What has once been the domain of computer hobbyists (or "hackers"), has gained

---

[1] We use "open source" as shorthand to refer not only to open source software, but also to what scholars have called "community-based innovation" [e.g., 5, 6], "commons-based peer production" [7, 8], "free software" and "software libre" [9]. We are not attempting to diminish the differences between them.

acceptance with major corporations and governments [12] and created hundreds of millions of dollars in value [13]. In software, where open source has frequently been studied, developers were found to contribute through collective action organized on-line and in absence of direct monetary compensation [1, 14, 15]. Not only that the produced software is shared among contributors, but it is also freely available to non-contributors for personal and (often) commercial use.

The actions of such collectives, working in unison yet often without ever meeting each other to create products and services of economic value, have captured the attention of the media, general public, business practitioners and academics. Yet despite its growing economic impact, a theoretical explanation for its performance is still incomplete. While many have documented cases of open source and theorized about the motivation of participants, internal organization and market dynamics, there is little understanding of its performance. For instance, while open source has been discussed often in the case of software, recent accounts suggest that it may be much more widespread. In which industries, then, can we expect open source to compete with firms? Which goods can be successful produced by open source? Which environment does it require to succeed? Ultimately, what affects the performance of open source?

We build on previous explanations, mostly pertaining to open source software, and combine original fieldwork with agent-based simulation to generate propositions that offer analytical and predictive power as to the performance of open source. These can aid in pinpointing the elements that make such entities successful in achieving their collective goals. We proceed by reviewing the relevant literature and briefly presenting our fieldwork and the agent-based model and. We then present the computational experiments we conducted and the resulting propositions, the implications of which are later discussed.

2 Literature Review

The last decade saw booming interest in open source as a mechanism of production and innovation initially in software and more recently in realms beyond it [16, 17] A bibliometric query reveals a pattern of rapid growth in the use of "open source" and "open innovation" as terms in the academic literature. While in the three year period 1995-1998 "open source" was mentioned only 12 times, it appeared 32 times in 1999, 40 times in 2000, and more than double that in the following year. In 2009, the term appeared 687 times, representing an impressive seventeen-fold growth in a decade. [2]

---

[2] The citation count was carried out in January 2010, using the Web of Science database, which provide access to current and retrospective bibliographic information, author abstracted, and cited references from over 10,000 leading journals of science, technology, social sciences, arts and humanities and over 100,000 book-based and journal conference proceedings. It provides access to seven databases: Science Citation Index (SCI), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Index Chemicus, Current Chemical Reactions, Conference Proceedings Citation Index: Science and Conference Proceedings Citation Index: Social Science and Humanities. See http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science.
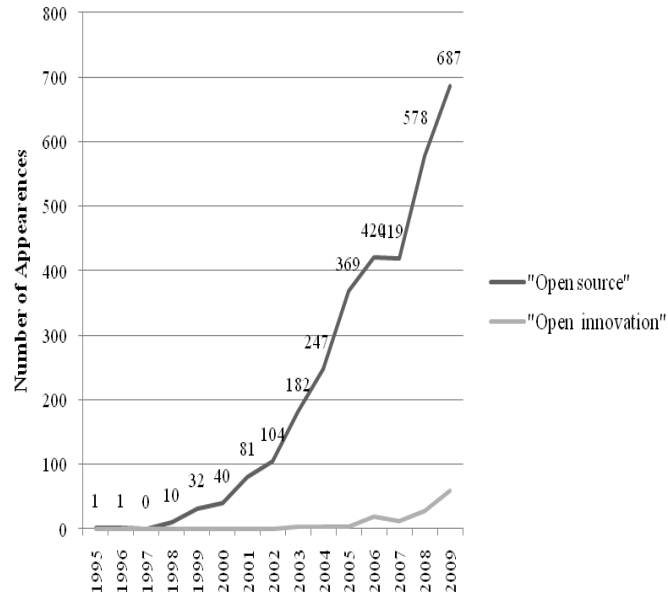
Fig. 1. Appearances of the terms "open source" and "open innovation" in the academic literature, 1995-2009

2.1 Open Source Differs from Firm-based Innovation

Scholars generally agree that open source represents a distinctive way of innovating as well as producing goods and services (hereafter: "goods"). Open source is "a fundamentally different organizational model for innovation and product development" [18]. While open source may complement (or replace) firm-based innovation [5, 19], it is emphatically different in elements as diverse as the individual motivation to participate [20], production [21], governance [18, 22] and market behavior [23, 24].

The characteristics that distinguish open source from firm-based (and market-based) innovation are related to its unique benefits. First is the ability to build upon others' work in the most direct way because the architecture of the good is visible and openly available, such as the lines of code in software or the circuits of a mobile phone [25, 26]. Second, open source may enhance innovation and problem solving by "removing barriers to entry to non-obvious individuals" [27], allowing users to contribute as much or as little as they may wish (or able) to contribute, without requiring upfront commitment or pre-specified roles, unlike in most firm settings [7, 28]. For firms, open source can generate rents from innovations created not by its personnel, but by users or consumers [29-31]. Better performance can also come from

engaging in open collaboration with other firms [32, 33]. At the societal level, open source can benefit public welfare [7, 34] as well as foster virtue [8].

2.2 A Lingering Puzzle of Performance

Yet those unique benefits of open source also pose several important puzzles [16]. It is intriguing that open source thrives on the internet [but also elsewhere, cf. 35], which is largely devoid of "the social signaling, cues, and relationships that tend toward moderation in the absence of law" [36]. These were thought to be necessary for such collaborative sharing behavior to occur [cf. 37]. The same conditions that give rise to open access and consumption with minimal intellectual property rights should have also caused it to quickly collapse under the weight of free riding, as theoretician have predicted [38, 39, cf. 40]. It is all the more puzzling because participants in open source effort could have easily regulated the sharing of costs and benefits by restricting access to those able and willing to contribute. Sharing does not *necessitate* open source. [cf. the notion of club goods, 41]. Nevertheless, open source seems resistant to free riding and highly skewed contributions [28, 42, 43].

While there is a large and growing body of knowledge on innovation in firms, the same has been developing for open source production and innovation. It is hard to imagine how the same conditions that promote innovation in firms would apply unchanged to open source, given the dramatic differences in the why innovations are produced [for a review see 5, 42]. Hence, we see a need to identify variables that exert important effect on the performance of open source organization, be it a coding community, file sharing or an advice forum [44]. Such theory would be essential in encouraging it, as some have urged to do [45, 46].

2   From Qualitative Data to a Model of Open Source

We build on qualitative data originating in a non-software environment, where a product of economic value is openly shared, to develop agent-based modeling. We use the two approaches jointly to investigate the phenomenon, utilizing the complementary nature of a grounded theory and formally expressed theory [47]. Qualitative accounts are rich in detail, evocative, describe processes lucidly, and possess high external validity [48, cf. 49]. However, qualitative data is hardly parsimonious, affords little in generalizability, and allows only limited field experimentation [cf. 50, 51]. Formal models have been employed in some of the pivotal studies in organizational theory [52, 53]. Agent-based models, of the kind we employ here, are increasingly seen as a promising way to study complex phenomena, whether social or organizational [54-56]. They have been used in the study of open source [57]. As recently demonstrated [58], agent-based models can capture features of the social order, such as embeddedness of actors and the emergence and dynamics of norms, that are difficult to represent by traditional analytic models.

3.1 Specifying the Model: Building Blocks of Open Source

In line with the boarder conceptualization of open source, which extends beyond software to other realms of innovation and production, we sought to indentify elements that were present in the specific field setting and also theoretically distinguishing between open source and non-open-source systems. While an all encompassing definition is beyond the scope of this paper, we present a working definition that marks the boundaries of our framework. It is applicable to all systems that feature the following elements.

Table 1. Modeled Elements of Open Source

Elements	Description	Corresponding Model Manipulation	Theoretical and Empirical Referents
Open access to contribute and consume	Anyone can join the development process or partake in its outcome, regardless of their level of contribution	Any agent in the model can participate in development or consumption with no exclusions	[1, 17, 44, 59, 60]
Create products of economic value	Products have clear economic value. That is, organizations whose primary purpose is social interaction (e.g., primary social group) are excluded	The dependent variable is the overall value created by the efforts of the collective, reflecting the economic value created	[17, 19, 35, 61]
Interaction and exchange activities are central	Participants interact, exchange and reuse each other's work	Each agent can engage in its own work or exchange with others	[7, 25-28, 62]
Participants work purposefully yet loosely coordinated	Coordination, structure and hierarchy are emergent and less specified compared to a firm or market setting	Agents coordinate only when engaging in exchange events	[5, 18, 21, 22, 42, 62-64]

The elements allow the model to capture multiple instances of what is generally regarded as open source, including beyond software [35], e.g. Wikipedia [28, 36, 65], user-run support forums [44, 66] and file sharing services.

3.2 Specifying the Model: Variables Related to Open Source Performance

While the variables presented above are thought to define open source as such, the manipulated variables are likely to affect the performance of the system.

Table 2. Model variables Manipulated in the Studies

Level of Analysis	Variable	Description	Corresponding Model Manipulation	Theoretical and Empirical Referents
Individuals	**Agent & Populations Characteristics** (Study 1)	Agent characteristic is defined in terms of its likelihood to contribute; population characteristic is defined in terms of the distribution of agent characteristics in a population.	Agent characteristics are drawn from three types: cooperator, reciprocator, and free rider; Population characteristics are systematically varied and their impact on performance is assessed.	[18, 38, 40, 42, 43, 67]
	Demand (Study 2 & 3)	Demand describes the uniqueness of a sought resource.	Levels of Demand were manipulated to discern effect on performance, and how Agent, Population & Rivalry characteristics interact with it.	[7, 27, 28, 68-70]
Goods	**Rivalry** (Study 2 & 3)	Goods differ on the cost associated with providing them to a marginal individual, i.e., the cost of contributing them to others. E.g., rival good cannot be used simultaneously by more than one agent.	Levels of Rivalry were manipulated to discern effect on performance, and how Agent, Population & Demand characteristics interact with it.	[19, 41]

Degree of Rivalry. A good is considered non-rival if for any level of production the cost of providing it to a marginal (additional) individual is zero [41]. Non-rivalry does not imply that the *total* production costs are low, but that the *marginal* production costs are low.[3] Thus, sharing a pure non-rival good does not decrease the utility of any individual from consumption (e.g., sharing a digital music file or software code). Few goods are perfectly rival or non-rival and one can imagine a continuum of rivalry [18, 71].

Composition of Cooperative Types in the Population. People's behavior with regards to cooperation is heterogeneous between- and stable within-individuals [67, 72-75] with remarkable stability across cultures [76]. An individual's type is so stable

---

[3] For instance, a non-rival good such as national defense is extremely expensive to produce, but the cost is insensitive to the number of beneficiaries.

that "a group's cooperative outcomes can be remarkably well predicted if one knows its type composition" [67]. We reflect those recent findings by assigning agents to follow one of the three empirically observed types: 1) Cooperators – which contribute to others even at cost to self [77]; 2) Reciprocators – which contribute based on others' behavior [cf. 78], e.g., contribute if they observed others doing so; 3) Free Riders – which do not contribute, but still consume. Kurzban and Hauser's [67; hereafter: KH] empirical results (hereafter: KH ratio) suggest that 13% of individuals can classified as Cooperators, 53% as Reciprocators, and 20% as Free-riders. The behavior of the remaining 14% is not stable enough to be classified. In implementing the KH types in the model, a global parameter of group contribution behavior was defined in terms of the exchange ratio (exchanges/attempts), reflecting the population rate of contribution for a given period, εp.4

Demand Homogeneity. Of particular importance is the homogeneity of participants with regards to demand, which may be dependent on relevant skills [e.g., 7, 27, 28] and the motivation to contribute and consume [e.g., 4, 18, 69, 70]. Together, they both determine the extent to which participants are placing demand on the open source system.

4   Computational Experiments and Results

The model consists of a simple population of 100 agents ($a_1,....,a_{100}$), each of which follows the algorithm described in the figure.[5]
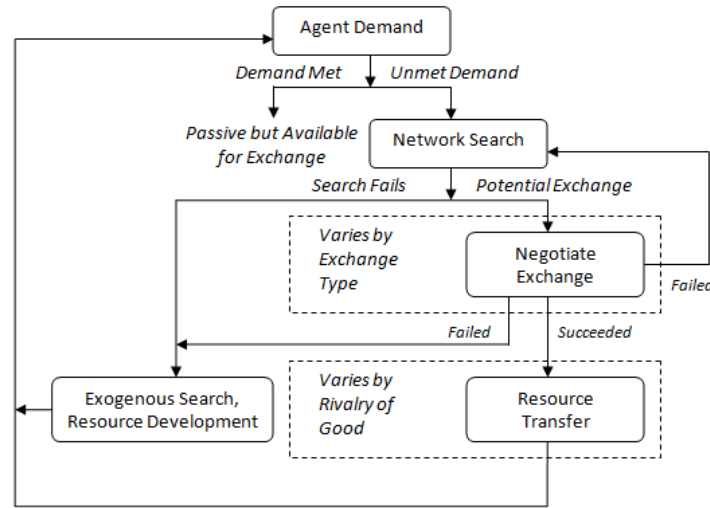
Fig. 2. Agent behavioral algorithm for the simulation

4.1 The Impact of Cooperative Types on Performance

We consider the impact of a range of cooperative type mixes, modeled after KH types, on the time (measured in steps) that it takes to complete the set of problems (goals) in an open source system. Hundred agents were randomly assigned 100 goals each to complete, with initial resources levels randomly drawn from a [0,100] flat distribution. All goals were unique (i.e., low demand for each resource) and all resources were non-rival. The values of each type were varied in steps of 10% ranging from 0.0% to 100%, with 100 runs made for each configuration.[6] Each run defines an instance of a problem faced by the population of the open source system. All problems are considered independent, such that there is no knowledge carryover from prior runs.

Results. The results are plotted the figure as overall performance means for each level of cooperator percentage in the population (solid line), with the vertical bars denoting 95% confidence intervals. In the figure, the x-axis depicts the percent of cooperators in the population (e.g., 5%). For each set of runs at a fixed cooperator percentage of the x-axis, the remaining percent (i.e., 95%) of the population contains reciprocators and free-riders whose mix is systematically varied from all reciprocators to all free-riders. An expected main effect of cooperators is apparent and confirms our intuition about cooperation in general – the more cooperators in the population, the better performance of the open source model.

Proposition 1a: Cooperators improve performance. Over a mix of reciprocator and free-rider levels, a higher ratio of cooperators leads to better performance.

Although the performance plot is increasing (see figure), that there is a distinct concavity in the graph. We continued by analyzing further the effect of cooperator ratio on performance to reveal a significant pattern of nonlinearity. The figure shows (in dotted line) the added benefit in performance due to higher percentages of cooperators. As is suggested in the figure, there is a decreasing marginal benefit to the addition of cooperators to the population.[7] The largest gain occurs when the percentage in the population jumps from 1% to 5%, but trends down as more cooperators enter the population.

---

[6] For all three studies, cell sample sizes (replications) were planned in order to detect absolute effect sizes with $\alpha = 0.05$ for all main effects and interaction contrasts with likely power $\geq$ .80 [79].

[7] An analysis of variance confirmed that differences in performance across percentage levels of cooperators were significant ($F(1,8588) = 4897.8$, $p < .001$). A post-hoc analysis (Games & Howell 1976) revealed that all means differed from each other significantly and the subtended line connecting the means had a best fit with a logarithmic model ($R^2 = .842$, SE = .075). All analyses conducted using SPSS 17.0 (www.spss.com) and Statistica 8 (www. statsoft.com).

Proposition 1b: Decreasing Marginal Returns from Cooperation. Over a mix of reciprocator and free-rider levels, increasing the ratio of cooperators has decreasing marginal positive effect on performance.

Finally, an examination of the confidence intervals in the figure indicates that the spread decreases (i.e., the intervals become "tighter") as the ratio of cooperators increases. That is, the variation in performance for a given ratio of cooperators decreases as there are more cooperators in the population.

Proposition 1c: Cooperators reduce variance. Over a range of reciprocator and free-rider levels, a higher ratio of cooperators reduces variation in performance. Stability, and thus predictability, in model performance is accommodated by increasing the ratio of cooperators in the population.
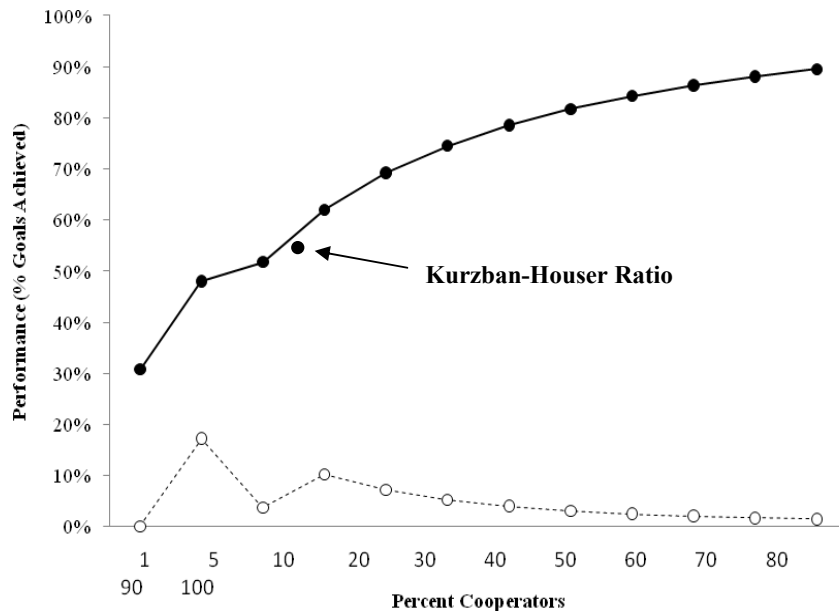

Fig. 3. Mean performance for cooperator ratio in the population (*solid line*), marginal improvement in performance (*dotted line*).

What accounts for the higher variation and unpredictability in performance with lower ratios of cooperators? To answer this question, we examined the sensitivity of performance to these specific mixes of reciprocators and free-riders. As can be seen in the figure, for each cooperator percentage, there is an embedded plot that shows performance decreasing from left to right, as less reciprocators (and more free-riders)

enter the population. We interpret this as a distinct sensitivity to the percentage of free-riders. For example, as indicated in the figure, populations with 5% cooperators exhibit a wide variation in performance depending on the particular mix of reciprocators and free-riders (compare performance in point A in the figure with performance in point B). Thus, variance in performance is driven by the ratio of reciprocators to free-riders.

As described in Figure 4, performance drops can be mitigated: for example, a change of 95% reciprocators to 55% reciprocators results in only a 3% average drop in performance, after which a significant decline performance ensues (C in the figure). A pattern of compensatory substitution between cooperators and reciprocators is revealed in the data: high levels of reciprocators can compensate for low levels of cooperators, and vice versa. For example, in the figure, the Φ-line depicts a performance level of 42%, which can be achieved with a population with 1% cooperators, but also with 5% cooperators or 10% cooperators, all under various mixes of Reciprocators and Free-riders. However, this flexibility of these types of tradeoffs decreases as the percent of Cooperators increase in a population. In part, this is due to the decrease in sensitivity to the number of Free-riders in the population and, consequently, the decline in performance variance.

Proposition 1d: Cooperator-Reciprocator Trade-off. Reciprocators improve performance and higher-ratios provide performance improvements. But the performance improvements decline as more Cooperators are present.
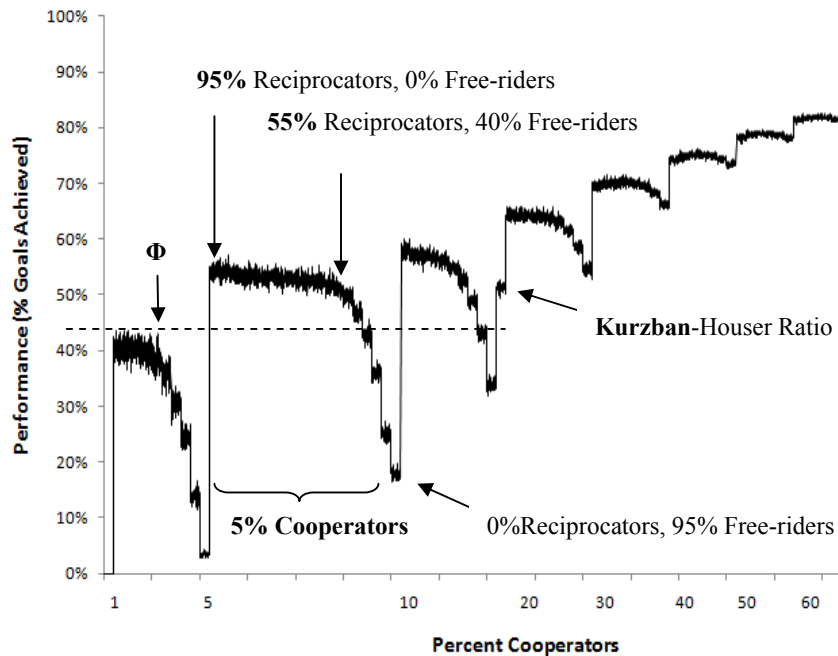
Fig. 4. Performance by varying types of agents. For each constant percent of cooperators (*x-axis*), the relative ratio of reciprocators to free-riders is varied (high to low reciprocators).

Intuitively, the finding is driven by the varying presence of *reciprocators* in the population, which is reversely correlated with the presence of free riders. Empirically, reciprocators, who adjust their behavior according to that of those around them, are the largest section of human population [67, 76]. At low levels of cooperators (left side of the X axis) it means that a large chunk of agents are willing to "change colors" according to the situation. Performance is then driven by the group (cooperators or free riders) that serves as the "role model" for reciprocators. In each step in the analysis, we held the ratio of Cooperators constant while varying the ratio of Reciprocators to Free-riders. As the ratio of Free-riders grew, they exert growing influence on the behavior of reciprocators pushing them to behave as free-riders. The conversion processes accelerates, which leads to rapid decrease, indeed a collapse, in performance. As one travels to higher ratios of cooperators (by moving towards the right side of the X axis), there is more influence on reciprocators to mimic a cooperator-like behavior. Reciprocators are less likely to be influenced by free-riders, the variance in performance decreases and the substitution effect is lessened.

4.2 The Impact of Rivalry and Demand

To simplify the initial analysis, we elected to explore the facets (the extreme values) of the parameter space. We examined three mixes of the population: 1) Cooperative, which we defined as 98% Cooperative, 1% Reciprocators, 1% Free-riders; 2) Reciprocators, which we defined as 1% Cooperative, 98% Reciprocators, 1% Free-riders;[8] and 3) KH ratio, defined as above with 13% cooperators, 53% reciprocators, 20% free-riders, and remaining 14% unclassified. Similarly, two levels of Rivalry were examined. In High Rivalry, goods were perfectly rival, such that they could be consumed by one individual at a time (e.g., could not be copied). In Low Rivalry, goods were perfectly non-rival, such that a good could be consumed by $N > 1$ individuals (e.g., could be copied). Also, Demand was defined in terms of the distribution of goals in the population. High Demand (or homogenous demand) was realized by a high replication of goals in the population (e.g., all desiring the same resource) while Low Demand (or heterogeneous demand) was realized as few goal replications in the population (e.g., all desiring different resources).[9] For each condition, 100 runs were conducted.

We found three statistically significant results. The first two are expected and

---

[8] One percent rather than zero was used in order to prevent conditions where no exchanges would occur in the case of cooperators, and therefore was incorporated in the other conditions for balance. Dominant free-riders (i.e., 98%) were eliminated from this analysis because of the extremely low performance.

[9] High Demand was determined by a random draw from a [90-100%] flat distribution of goal redundancy while Low Demand was determined by a random draw form a [0-10%] flat distribution.

intuitive. High Rivalry and high Demand *decreased* the performance of the model.[10] Expected as they are, these two main effects serve to substantiate the validly of the model by matching theoretical predictions made elsewhere (see Table 2).

Proposition 2a: Rivalry decreases the effectiveness of open source performance.

Proposition 2b: High (homogenous) demand decreases open source performance.

The third result is less expected. As visible in the figure, we found an interaction: under low Rivalry conditions performance is at its highest when Demand is high (50.9%); however, when Rivalry conditions are high, high Demand resources drive performance to its lowest level (21.1%). Recall that high demand conditions would be reflected in many agents seeking the same resource type and vice versa. The consequences for resource availability are specified by the Rivalry factor interacting with the Demand – high demand results high exchanges and resource redundancy when the Rivalry is low (e.g., exchanging music files). However, when both Demand and Rivalry is high, available resources existing in the network are quickly extracted and resistant to exchange. Therefore, available supply, and consequently performance, would vary across Demand conditions subject to resource availability within Rivalry constraints.
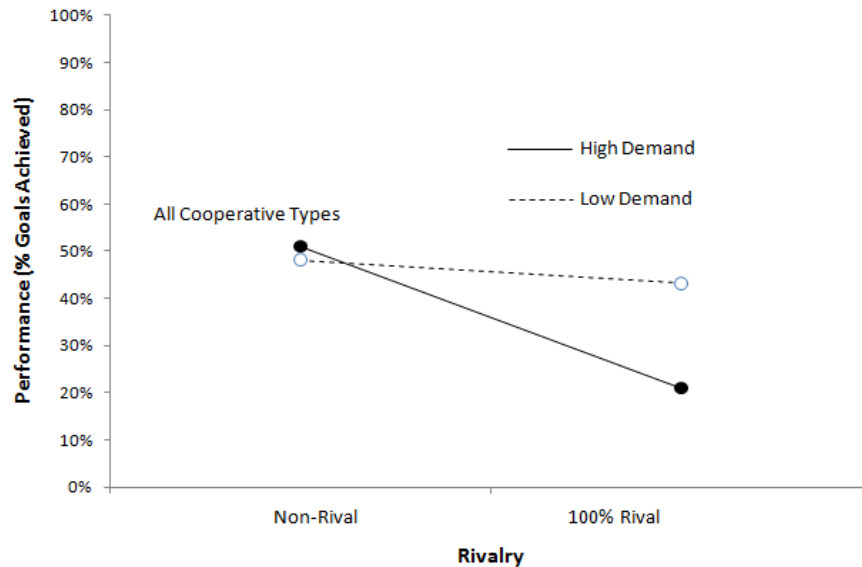
Fig. 5. Interaction between Rivalry levels and Demand levels, across the entire mix of cooperative ratios

---

[10] The overall main effects for Rivalry and Demand were $F(1, 1196) = 87.3$, $p < .001$) and $F(1, 1196) = 26.8$, $p < .001$) respectively. The interaction was $F(1, 1188) = 63824.0$, $p < .001$), and a post-hoc Tukey analysis indicated that all means differed significantly ($p < .001$).

But would such obvious constraints on exchange vary across agent types? To answer this, we analyzed the interaction between these factors and the population types. In this analysis, we also included the Kurzban-Houser ratio, the most empirically valid ratio. The results are shown in Figure 6. Populations with high numbers of cooperators under low demand are insensitive to rivalry, but populations with KH ratio compositions are sensitive, resulting in significant drops in performance, but not near those populations dominated by reciprocators. The former populations are willing and able to share resources, but the latter populations are suffering from the impact of free riders. In KH populations, there seem to be sufficient cooperators to sway the behavior reciprocators repeatedly toward higher donation behaviors. On the other hand, with high demand goods, cooperators are more sensitive to rivalry than KH ratio populations. Under high rivalry conditions, cooperating population performance declines greater than the KH decline, and the two converge to virtually the same performance levels. Despite well-intended cooperators, exchanging high rivalry goods essentially does not alter the performance as supplies are fixed within the population. Variation in performance (i.e., low reciprocators) is accounted for by a substantial drop in cooperation where the base level of resources in the population (from initial conditions) is not extracted. Therefore, KH populations function to distribute resources as efficiently as high ratios of cooperators.

KH ratio populations also exhibit an interaction between Demand and Rivalry levels, where under non-rival conditions low demand conditions outperform high demand, but this is reversed under high rivalry. The reason is that under low demand conditions with high rivalry, fewer exchanges are made (not all are seeking the same resource, and not all resources are associated with pure cooperators) which bias the reciprocators to act more like free-riders. Finally, and although substantially lower in performance, reciprocators also exhibit a strong interaction, but the opposite of the KH – under non-rival conditions, high demand performs better, while under rivalry low demand dominates.[11]

---

[11] Rivalry by Demand by Cooperator Type interaction: $(F(2,1188) = 21690.0, p < .001)$.
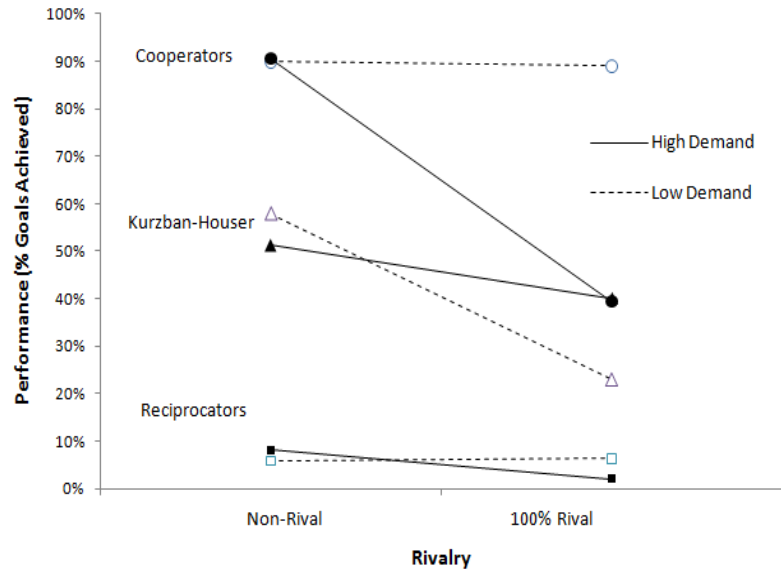
Fig. 6. Interaction between Rivalry levels, Demand levels, and three mixes of cooperator types.

Proposition 2c. Population types interact with rivalry and demand levels differently. The impact of Rivalry and Demand on performance varies with the cooperative mix of a population.

   Intuitively, lower rivalry alleges some of the damage done by free riding, because even massive free riding decrease only slightly the availability of goods to cooperators. In contrast, if the goods are highly rival, free riders will bear no cost but enjoy all benefits, while cooperators will still bear cost. Obviously, a system with more contributors will likely be better off, but the point here is that when goods are (perfectly) non-rival, the tragedy of the commons would be avoided even with free riding.

4.3 Typical Cooperation Type Ratios in the Population

With growing interest in the cooperative type distributions found in the human population [e.g., 67, 72-75], we elected to conduct further analysis on the KH ratio.[12] We examined multiple population compositions that meet KH ratio limits across three levels of Rivalry (0%, 50%, 100%) and three levels of Demand (1%: Low, 50%: Medium, 100%: High) with 100 runs per level, then plotted the surface mapped into

---

[12] Where 13% were classified as cooperators, 53% as reciprocators, and 20% as free-riders, with the remaining 14% behaving inconsistently.

performance levels. The result is shown in the figure.[13] Interestingly, the performance surface changes from a linear response (under High Rivalry, see α) to a non-linear response (under High Demand, see β). Thus,

Proposition 3a. Rivalry has Non-Linear Effect on Performance. Open source performance in Kurzban-Houser ratio responds linearly to Demand and non-linearly to Rivalry. The impact of Rivalry and Demand on performance varies with the levels of each.
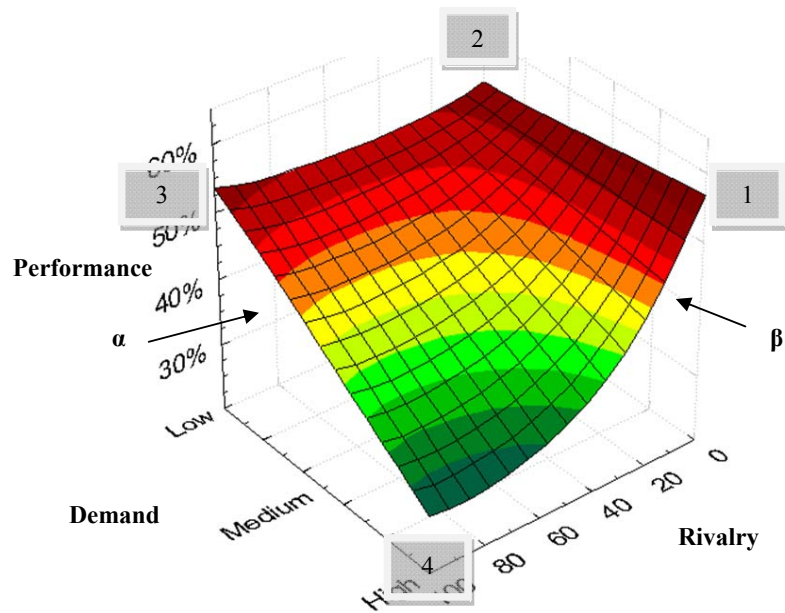
Fig. 7. Performance surface for Kurzban-Houser ratios, crossing Demand levels with Rivalry levels. *Colors* reflect bands of performance levels from maximum (*Dark Red*) to minimum (*Dark Green*).

This finding begins to explain why open source does not collapse due to free riding. First, the impact of free riders is not large because, as KH showed, they are a minority to begin with. It is apparent in the figure that even with few cooperators and many reciprocators an increase in the free rider population affects performance appreciably only when they become a large part of the population. Second, the effect is further weakened thanks to the low rivalry (e.g., in point β), where the presence of free riders does not lead to decreased availability of goods. Because the goods are non-rival, the consumption of each marginal unit is zero, much like a downloader of music or video file does not directly reduce the availability of the same file for other

---

[13] 900 points were plotted done using Statistica 8.0 (www.statsoft.com) a using distance-weighted least squares fit model.

users. The only effect of free-riders is in the opportunity cost of their time, which reduces the overall performance since they make no contributions. Overall then, the effect of free riders is muted.

The most interesting result comes from the interaction between the two variables. A combination of low rivalry (hypothesized to improve performance) and high demand (hypothesized to harm performance) (point 1) generates performance at level that is remarkably close to of the ideal combination of low rivalry and low demand (point 2). Similarly, the combination of high rivalry and low demand generates performance on par (point 3). Only the combination of high rivalry and high demand results in expectedly low performance (point 4). Hence,

Proposition 3b. Rivalry–Demand Compensatory Effect. Under realistic assumptions of cooperation, open source will exhibit close to perfect performance with either low rivalry *or* low demand.

The last finding is perhaps the most important of the three. The mechanism behind it may be as follows: when rivalry=100% and demand high, almost the entire population is composed of users only and the cost of contribution is expensive, equal to the cost of production. In these circumstances, decreasing rivalry leads to an instant performance boost, because although there are still few contributors, their cost of contributing goes down as fast as rivalry decreases (follow the curved line from point 4 to point 1, along the edge of the surface). When rivalry reaches zero, even few contributors are enough to provide the demands of a large user population. At this point, decreasing demand will not lead to much improvement in performance, as the model shows (follow the linear edge between point 1 and point 2, along the edge of the surface).

A similar process is apparent when one begins from the same point of departure and takes the second path for performance gain by reducing demand (follow the line from point 4 to point 3, along the edge of the surface). As demand shrinks, more and more users are also becoming contributors, catering to their own needs. While the cost of contributing remains high, users are providing their own needs, and the performance of the system improve. At the extreme (point 3*)*, all users are also self-sufficient as contributors. Then, decreasing rivalry leads to just small improvement in performance (follow the line connecting point 3 to point 2, along the edge of the surface).

5   Discussion and Conclusion

The results contribute to the emerging theory of open source. One, they validate the model and provide theoretical support of existing theory. Two, several propositions point out to novel effects, which we have not seen proposed hitherto. Importantly, we find decreasing marginal returns from cooperation (proposition 1b) and trade-off effect between cooperators and reciprocators in the population (proposition 1d). These findings may have significant implications for research and practice. For instance, they suggest that open source systems do *not* require a population of cooperators.

Open source can thrive even when cooperators are just a small minority in the population, implying that it can be expected to appear in more places. Not only that, but also the performance of the system depends critically on just a small core group of contributors. Increasing that group leads to performance improvement but in a decreasing manner. Of particular interest is the finding that under the realistic cooperation assumptions of the Kurzban-Houser ratio, the most empirically valid ratio, performance of 42.5% can be achieved with only 13% cooperators, While researchers have observed that contributions in open source setting are highly skewed, our propositions aid in making sense of that. Having a core group of cooperators is not a deficiency in the system, but a rather expected feature. Cooperators, which may be difficult to find as they are a minority in human population can be replaced by reciprocators, which are much more common. For practitioners who are interested in facilitating open source, these findings can offer relief: creating a viable system may be easier than previously presumed.

While many cooperators are not necessary for open source systems to perform well, their presence has another impact: they reduce variance in performance (proposition 1c). This can lead a step forward in designing more fertile grounds for open source. Some systems, such as research and development teams, may be geared more towards high performance than reliability [cf. 80]. In such environments, a small percentage of cooperators with a majority of reciprocators may be sufficient to achieve the breakthrough sought. Additional effort to attract cooperators will better performance but can be inefficient in terms of cost versus benefit. This can explain, for instance, the focus on open sources in innovation systems, as opposed to production systems. However, even when reliable performance are sought, for instance in when providing an on-going service (e.g., *Wikipedia*), open source can still thrive. There, the effort to attract more cooperators may be efficient as it leads to more reliable performance over time.

Finally, the results suggest interactions between conditions of rivalry and demand, on the one hand, and population composition, on the other hand. We find that the impact of rivalry and demand on performance varies with the cooperative mix of a population (proposition 2d). For instance, increasing rivalry leads to a dramatic drop in performance, but mostly in populations that are made of cooperators. In the more empirically likely case of KH ratio population, increasing rivalry leads to a gentler drop in performance in the case of high demand, which is important in applications of open source. Once again, this is good news for observers of open source as well as practitioners who would like to benefit from it.

Finally, two propositions carry particular importance for the performance of open source in harsh environments, such as with rival goods or high (homogenous) demand. The findings suggest that Rivalry has non-linear effect on performance in the likely case of a KH ratio population (proposition 3a). The implication is that even a slight reduction in rivalry can bring about a boost in performance. For practitioners this can be an effective tool in improving the performance of open source systems. Additionally, even systems with high rivalry can end up in close to perfect performance if the demand is low (proposition 3b). Thus, even products that are close to perfect rivalry (e.g., in producing food) can benefit from open source production, as long as demand is managed properly. This suggests that open source can be expected in greater variety of venues that we currently see.

Our analysis leaves much room for future work on the performance of open source. One future direction may include modeling of social institutions, such as those that regulate exchange. We know that individuals can exchange (or contribute) under multiple exchange configurations [81, 82]. In particular, two exchange configurations that were documented as legitimate in the institutional sense [83, 84]: *embedded exchange* [85, 86] and *generalized exchange* ("pay it forward", "gift economies"), where unacquainted participants help each other with the expectation that reciprocity will come from *any* other member, not necessarily the specific receiver [87, 88]. While such analysis is beyond the scope of this paper, we suspect that it may be fruitful in furthering our understanding of open source performance.

References

1. Raymond, E.S.: The Cathedral and the Bazaar. O'Reilly, Sebastopol, California (1999)
2. Henkel, J.: Champions of Revealing - The Role of Open Source Developers in Commercial Firms. Industrial and Corporate Change 18, 435-471 (2009)
3. von Hippel, E.: Open Source Software Projects as User Innovation Networks. In: Feller, J., Fitzgerald, B., Hissam, S.A., Lakhani, K.R. (eds.) Perspectives on Free and Open Source Software, pp. 267-278. The MIT Press, Cambridge, Massachusetts (2005)
4. Franke, N., Shah, S.: How Communities Support Innovative Activities: An Exploration of Assistance and Sharing Among End-Users. Research Policy 32, 157-178 (2003)
5. Lee, G.K., Cole, R.: From a Firm-Based to a Community-Based Model of Knowledge Creation: The Case of the Linux Kernel Development. Organization Science 14, 633-649 (2003)
6. Shah, S.: Community-Based Innovation & Product Development: Findings From Open Source Software and Consumer Sporting Goods. Massachusetts Institute of Technology, Cambridge, MA (2003)
7. Benkler, Y.: Commons-Based Strategies and the Problems of Patents. Sci. 305, 1110-1111 (2004)
8. Benkler, Y., Nissenbaum, H.: Commons-based Peer Production and Virtue. Journal of Political Philosophy 14, 394–419 (2006)
9. Berry, D.M.: The Contestation of Code: A Preliminary Investigation into the Discourse of the Free/Libre and Open Source Movements. Critical Discourse Studies 1, 65-89 (2004)
10. Maurer, S.M., Rai, A., Sali, A.: Finding Cures for Tropical Diseases: Is Open Source an Answer? PLOS Medicine 1, E56 (2004)

11. Bodie, M.T.: The Future of the Casebook: An Argument for an Open-Source Approach. Journal of Legal Studies 57, 10-35 (2007)
12. Guth, R.A.: Asia to Develop Software to Rely Less on Microsoft. The Wall Street Journal, pp. B4 (2003)
13. Lohr, S.: Novell to Buy SuSE Linux for $210 Million. The New York Times, pp. 6, New York (2003)
14. O'Mahony, S.: The Emergence of A New Commercial Actor: Community Managed Software Projects. Unpublished doctoral dissertation, Stanford University (2002)
15. Kogut, B., Metiu, A.: Open-source Software Development and Distributed Innovation. Oxford Review of Economic Policy 17, 248-264 (2001)
16. von Krogh, G., von Hippel, E.: The Promise of Research on Open Source Software. Management Science 52, 975-984 (2006)
17. von Krogh, G., von Hippel, E.: Special issue on open source software development. Research Policy 32, 1149-1157 (2003)
18. Shah, S.K.: Motivation, Governance and the Viability of Hybrid Forms in Open Source Software Development. Management Science 52, 1000-1014 (2006)
19. von Hippel, E., von Krogh, G.: Open source software and the private-collective innovation model: Issues for organization science. Organization Science 14, 209-223 (2003)
20. Roberts, J.A., Hann, I.-H., Slaughter, S.A.: Understanding the Motivations, Participation and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects. Management Science 52, 984-999 (2006)
21. MacCormack, A., Rusnak, J., Baldwin, C.Y.: Exploring the Structure of Complex Software Designs: An Empirical Study of Open Source and Proprietary Code. Management Science 52, 1015-1030 (2006)
22. O'Mahony, S., Ferraro, F.: The Emergence of Governance in an Open Source Community. Academy of Management Journal 50, 1079-1106 (2007)
23. Casadesus-Masanell, R., Ghemawat, P.: Dynamic Mixed Duopoly: A Model Motivated by Linux vs. Windows. Management Science 52, 1072-1084 (2006)
24. Economides, N., Katsamakas, E.: Two-Sided Competition of Proprietary vs. Open Source Technology Platforms and the Implications for the Software Industry. Management Science 52, 1057-1071 (2006)
25. Häfliger, S., von Krogh, G., Späth, S.: Code Reuse in Open Source Software. Mangement Science 54, 180-193 (2008)
26. Majchrzak, A.: The Effect of Expertise Sharing and Integrating Behaviors in Wiki-based Organizational Intranets. Working Paper (2009)
27. Jeppesen, L.B., Lakhani, K.R.: Marginality and Problem Solving Effectiveness in Broadcast Search. Organization Science (forthcoming)
28. Anthony, D., Smith, S.W., Williamson, T.: Explaining Quality in Internet Collective Goods: Zealots and Good Samaritans in the Case of Wikipedia. Rationality and Society 21, 283-306 (2009)
29. Bonaccorsi, A., Giannangeli, S., Rossi, C.: Entry Strategies Under Competing Standards: Hybrid Business Models in the Open Source Software Industry Management Science 52, 1085-1098 (2006)
30. von Hippel, E.: Democratizing Innovation. MIT Press, Cambridge, Massachusetts (2005)

31. West, J.: How open is open enough? Melding proprietary and open source platform strategies. Research Policy 32, 1259-1285 (2003)
32. Henkel, J.: Selective Revealing in Open Innovation Processes: The Case of Embedded Linux. Research Policy 35, 953-969 (2006)
33. Waguespack, D.M., Fleming, L.: Scanning the Commons? Evidence on the Benefits to Startups Participating in Open Standards Development. Management Science 55, 210-223 (2009)
34. Maurer, S.M., Scotchmer, S.: Open Source Software: The New Intellectual Property Paradigm. National Bureau for Economic Research Working Paper W12148, (2006)
35. Shah, S.K.: Open Beyond Software. In: Cooper, D., DiBona, C., Stone, M. (eds.) Open Sources 2.0: The Continuing Evolution, pp. 339-360. O'Reilly Media, Sebastopol, CA (2005)
36. Zittrain, J.: The Future of the Internet – and How to Stop It. Yale University Press (2008)
37. Uehara, E.: Dual Exchange Theory, Social Networks, and Informal Social Support. American Journal of Sociology 96, 521-557 (1990)
38. Hardin, G.: The Tragedy of the Commons. Sci. 1243-1248 (1968)
39. Olson, M.: The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard University Press, Cambridge, MA (1965)
40. Milinski, M., Semmann, D., Krambeck, H.-J.: Reputation helps solve the 'tragedy of the commons'. Nature 415, 424-426 (2002)
41. Cornes, R., Sandler, T.: The theory of externalities, public goods, and club goods. Cambridge University Press (1986)
42. Mockus, A., Fielding, R.T., Herbsleb, J.D.: Two Case Studies of Open Source Software Development: Apache and Mozilla. In: Feller, J., Fitzgerald, B., Hissam, S.A., Lakhani, K.R. (eds.) Perspectives on Free and Open Source Software, pp. 163-209. The MIT Press, Cambridge, Massachusetts (2005)
43. Lerner, J., Tirole, J.: Economic Perspectives on Open Source. In: Feller, J., Fitzgerald, B., Hissam, S.A., Lakhani, K.R. (eds.) Perspectives on Free and Open Source Software, pp. 47-78. The MIT Press, Cambridge, Massachusetts (2005)
44. Lakhani, K.R., von Hippel, E.: How Open Source Software Works: Free User to User Assistance. Research Policy 32, 923–943 (2003)
45. Lessig, L.: Open Code and Open Societies. In: Feller, J., Fitzgerald, B., Hissam, S.A., Lakhani, K.R. (eds.) Perspectives on Free and Open Source Software, pp. 349-360. The MIT Press, Cambridge, Massachusetts (2005)
46. Free Software Foundation, www.gnu.org/philosophy
47. Strauss, A., Corbin, J.: Basics of qualitative research. Sage Publications, Newbury Park, California (1990)
48. Glaser, B., Strauss, A.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine De Gruyter, New York (1967)
49. Maanen, J.V.: Different Strokes - Qualitative Research in the Administrative Science Quarterly from 1956-1996. In: Maanen, J.V. (ed.) Qualitative Studies of Organizations, vol. 1. Sage Publications, Thousand Oaks, CA (1998)
50. Perlow, L.A.: The Time Famine: Toward a Sociology of Work Time. Administrative Science Quarterly 44, 57 (1999)
51. Lewin, K.: Action Research and Minority Problems. Journal of Social Issues 2, 34-46 (1946)

52. Nelson, R.R., Winter, S.G.: An Evolutionary Theory of Economic Change. Harvard University Press, Cambridge, MA (1982)

53. Cyert, R., March, J.G.: A Behavioral Theory of the Firm. Prentice Hall, Englewood Cliffs, NJ (1963)

54. Black, L.J., Carlile, P.R., Repenning, N.P.: A Dynamic Theory of Expertise and Occupational Boundaries in New Technology Implementation: Building on Barley's Study of CT Scanning. Administrative Science Quarterly 49, 572-607 (2004)

55. Rudolph, J.W., Repenning, N.P.: Disaster Dynamics: Understanding the Role of Quantity in Organizational Collapse. Administrative Science Quarterly 47, 1-31 (2002)

56. Repenning, N.R., Sterman, J.D.: Capability Traps and Self-Confirming Attribution Errors in the Dynamics of Process Improvement. Administrative Science Quarterly 47, 265 (2002)

57. Bonaccorsi, A., Rossi, C.: Why Open Source software can succeed. Research Policy 32, 1243-1258 (2003)

58. Moss, S., Edmonds, B.: Sociology and Simulation: Statistical and Qualitative Cross-Validation. American Journal of Sociology 110, 1095-1131 (2005)

59. Zeitlyn, D.: Gift economies in the development of open source software: anthropological reflections. Research Policy 32, 1287-1291 (2003)

60. Kollock, P.: The Economics of online cooperation: Gifts and public goods in cyberspace. In: Smith, M.A., Kollock, P. (eds.) Communities in Cyberspace, pp. 220-239. Routledge, New York (1999)

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

62. von Krogh, G., Späth, S., Lakhani, K.R.: Community, joining, and specialization in open source software innovation: a case study. Research Policy 32, 1217-1241 (2003)

63. Koch, S., Schneider, G.: Effort, Cooperation and Coordination in an Open Source Software Project: GNOME. Information Systems Journal 12, 27-42 (2002)

64. Kuk, G.: Strategic Interaction and Knowledge Sharing in the KDE Developer Mailing List Management Science 52, 1031-1042 (2006)

65. Giles, J.: Internet encyclopaedias go head to head. Nature 438, 900-901 (2005)

66. Jeppesen, L.B., Laursen, K.: The role of lead users in knowledge sharing. Research Policy 38, 1582–1589 (2009)

67. Kurzban, R.O., Houser, D.: An experimental investigation of cooperative types in human groups: A complement to evolutionary theory and simulations. Proceedings of the National Academy of Sciences 102, 1803-1807 (2005)

68. Thomas-Hunt, M.C., Ogden, T.Y., Neale, M.A.: Who's Really Sharing? Effects of Social and Expert Status on Knowledge Exchange Within Groups. Management Science 49, (2003)

69. Dahlander, L., Mckelvey, M.: Who is not developing open source software? non-users, users, and developers Economics of Innovation and New Technology 14, 617-635 (2005)

70. Alexy, O., Henkel, J.: Promoting the Penguin: Who is Advocating Open Source Software in Commercial Settings? Working Paper (2007)

71. Leach, J.: A course in public economics. Cambridge University Press (2004)

72. Fischbacher, U., Gächter, S.: Heterogeneous Social Preferences And The Dynamics Of Free Riding In Public Good Experiments. Working Paper The Centre for Decision

Research and Experimental Economics, School of Economics, University of Nottingham, (2008)

73. Kim, J., Lee, S.M., Olson, D.L.: Knowledge Sharing: Effects of Cooperative Type and Reciprocity Level. International Journal of Knowledge Management 2, 1-16 (2006)

74. Simpson, B., Willer, R.: Altruism and indirect reciprocity: The interaction of person and situation in prosocial behavior. Social Psychology Quarterly 71, 37-52 (2008)

75. Fehr, E., Fischbacher, U., Gächter, S.: Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms. Human Nature 13, 1-25 (2002)

76. Ishii, K., Kurzban, R.: Public Goods Games in Japan: Cultural and Individual Differences in Reciprocity. Human Nature 19, 138-156 (2008)

77. Fehr, E., Fischbacher, U.: The Nature of Human Altruism. Nature 425, 785-791 (2003)

78. Fowler, J.H., Christakis, N.A.: Cooperative Behaviour Cascades in Human Social Networks. Working Paper (2009)

79. Lenth, R.V.: Some Practical Guidelines for Effective Sample-Size Determination. The American Statistician 55, 187-193 (2001)

80. Hannan, M.T., Freeman, J.: Structural Inertia and Organizational Change. American Sociological Review 49, 149-164 (1984)

81. Biggart, N.W., Delbridge, R.: Systems of Exchange. Academy of Management Review 29, 28-49 (2004)

82. Levine, S.S., Prietula, M.J.: Towards a Contingency Theory of Knowledge Exchange in Organizations. In: Weaver, K.M. (ed.) Best Paper Proceedings. Academy of Management, Atlanta (2006)

83. Thomas, G.M., Walker, H.A., Morris, Z.J.: Legitimacy and Collective Action. Social Forces 65, 378-404 (1986)

84. Meyer, J.W., Rowan, B.: Institutionalized Organizations: Formal Structure as Myth and Ceremony. American Journal of Sociology 83, 340-363 (1977)

85. Granovetter, M.: Economic Action and Social Structure: The Problem of Embeddedness. American Journal of Sociology 91, 481-510 (1985)

86. Uzzi, B.: Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. Administrative Science Quarterly 42, 35-67 (1997)

87. Ekeh, P.P.: Social Exchange Theory: The Two Traditions. Harvard University Press, Cambridge, MA (1974)

88. Baker, W.E., Levine, S.S.: Mechanisms of Generalized Exchange. Working Paper http://ssrn.com/abstract=1352101 (2009)