Fast Success and Slow Failure
The Process Speed of Dispersed Research Teams

Mors, Marie Louise; Waguespack, David M.

Document Version
Accepted author manuscript

Published in:
Research Policy

DOI:
10.1016/j.respol.2021.104222

Publication date:
2021

Fast Success and Slow Failure: The Process Speed of Dispersed Research Teams*

Marie Louise Mors
Copenhagen Business School
Kilevej 14A, DK - 2000 Frederiksberg
Denmark
lm.si@cbs.dk

David M. Waguespack
Robert H. Smith School of Business
University of Maryland
College Park, Maryland 20742
USA
dwaguesp@umd.edu

March 2021
Accepted for publication at Research Policy

*Caroline Duvier, Eucman Lee and Maya Cara provided research assistance at various stages of this project. We are grateful to Marcus Asplund, Johanna Birnir, Waverly Ding, Roberto Fernandez, Magnus Thor Torfason, and Tim Simcoe for discussion of the data and ideas in this paper. Bill McEvily and participants at the Imperial College Workshop on Networks and Formal Boundaries, our colleagues at the University of Maryland and Copenhagen Business School, Lars Frederiksen, Keld Laursen and Ram Mudambi provided helpful comments. Louise Mors acknowledges funding from the RAMD fund at the London Business School for this project. The authors are listed in alphabetical order. Both authors contributed equally.

Fast success and slow failure: The process speed of dispersed research teams

Research teams are often dispersed across geography and organizational boundaries. While prior work has recognized that dispersed teams face coordination challenges, it is not clear how dispersion affects team process efficiency outcomes, such as the speed with which team efforts come to a resolution. Process efficiency outcomes are important because, if the performance benefits associated with working in dispersed teams have a trade-off cost in terms of efficiency, then the net performance benefits from working in such teams may be seriously diminished or even reversed. Prior work has also tended to focus on successful outcomes, which may lead to deficits in our understanding of coordination challenges in dispersed teams. To better comprehend process efficiency, we examine 5,250 teams that work together in an open standard setting, and where the time to resolution of both successful and failed projects is observable. We find that teams that are organizationally or geographically dispersed are fast at reaching success, but fail more slowly than non-dispersed teams. Our interpretation of these disparate outcomes is that the endogenous processes that make dispersed teams more likely to on average select higher potential projects have a second order effect of making it harder for dispersed teams to abandon failing projects. We argue that slow failure has important implications for research & development efforts given the opportunity cost of tying up resources in research teams.

Keywords: research teams, dispersed teams, collaboration, process efficiency, coordination, open source communities

1. Introduction

Research teams are increasingly assembled across geographic and organizational boundaries to maximize heterogeneity in backgrounds, skills, functions and geography (Katz and Martin, 1997; Cummings and Kiesler, 2007; Hoekman et al., 2010; Bozeman and Boardman, 2014). It is well established that such teams benefit from complementary skills, knowledge or access to resources that allow for more successful outcomes (Bercovitz and Feldman, 2011; Vural et al., 2013; Tzabbar and Vestal, 2015; Banal-Estañol et al., 2019; Le Gallo and Plunket, 2020). However, collaborative research teams also face coordination challenges, and these challenges are particularly salient when team members are dispersed across geography or separated by organizational boundaries (Espinosa et al., 2007; Hoekman et al., 2010; Dahlander and Frederiksen, 2012; Tzabbar and Vestal, 2015).

Extant work has typically assumed that the performance benefits of dispersed teams outweigh any challenges associated with working across boundaries. Research teams may, in fact, only choose to work together across boundaries if they deem the benefits to outweigh the costs. For example, Katz and Martin (1997), argued that spatial proximity will encourage collaboration, but if necessary, scientists may seek out more distant experts. Likewise, Hoekman and colleagues (2010) found in a recent study of research teams in Europe that there was still a bias towards working with physically proximate partners. In addition, improvements in information technology allow dispersed teams to coordinate more easily and hence encourage collaboration across boundaries (Jarvenpaa and Leidner, 1999; Bozeman and Boardman, 2014). Even so, scheduling meetings across time zones, communication errors or misunderstandings, or failure to align incentives across organizational or geographic boundaries may still impede research collaboration (Espinosa and Carmel, 2003; Hoekman et al., 2010; Le Gallo and Plunket, 2020). As a result, dispersed teams may face process inefficiencies that lead to delayed project outcomes (Cummings and Kiesler, 2007).

Building on this prior work on dispersed teams, we seek to further the understanding of process efficiency in dispersed teams. We distinguish process efficiency from absolute efficiency, whereby the former measures the effort of undertaking a collaboration, such as the length of time it takes a team to deliver a finished product. In contrast, absolute efficiency is a measure related to performance, such as whether the project is delivered on time or if it is approved by others. Our core theoretical argument is that while team dispersion - i.e., the extent to which teams work across organization boundaries or geography (e.g., Sorenson et al., 2006; Cummings and Kiesler, 2007; Le Gallo and Plunket, 2020)[1] - is positively associated with performance, the same factors of dispersion may correlate negatively with process efficiency. If this argument is correct, then in practical terms the net performance benefits from working in dispersed teams may be seriously diminished or even reversed by delaying reallocation of research effort. While others have recognized this tension and tried to measure process variables related to teamwork quality (Hoegl and Gemuenden, 2001; Hoegl and Proserpio, 2004) or model coordination costs (Espinosa and Carmel, 2003), we are the first, to our knowledge, to simultaneously measure and analyze absolute and process efficiency outcomes.

To empirically examine the process efficiency of dispersed teams, we follow a line of recent work that has studied dispersed project teams in online and scientific communities (e.g., Hoegl and Proserpio, 2004; Comino et al., 2007; Cummings and Kiesler, 2007; Tzabbar and Vestal, 2015). Our study relies on unique data on project teams in an open source community that work together across different organizations and are geographically dispersed. We examine 5,250 teams composed of 17,818 authors in the Internet Engineering Task Force (IETF) from 1995 to 2003. This is an ideal setting for studying process efficiency as these self-selected teams vary in

[1] Sorenson, Rivkin and Fleming, in their study of collaborations on patent filings, control for geographic and organizational proximity. The authors argue that these boundaries are important determinants of social interaction (2006: 1005).

terms of organizational and geographic dispersion, and because we are able to observe the time to resolution of both successful and unsuccessful projects. We argue that studying both successful and failed projects is critical, as the endogenous selection processes that potentially lead dispersed teams to work together may become liabilities when those projects do not perform well. Specifically, as prior work has argued that dispersed teams face coordination challenges (Espinosa and Carmel, 2003; Cummings and Kiesler, 2007), we hypothesize that dispersed teams are likely to take longer than non-dispersed teams to reach resolution. Yet, there are also reasons to believe that dispersed teams may anticipate coordination costs (Kotha et al., 2013) and therefore select to work on projects that are more likely to succeed. Hence, we introduce a competing hypothesis that dispersed teams are faster than non-dispersed teams at reaching resolution. We further theorize that the endogenous selection by dispersed teams of projects that are more likely to succeed, means that process speed may differ depending on the outcome: Successful dispersed teams may reach the outcome faster than non-dispersed teams. In contrast, unsuccessful dispersed teams will take longer to fail than non-dispersed teams. We test and find support for the arguments that the selection effect determines process speed and that process speed therefore depends on whether projects are successful or fail. As such our key finding is that coordinating across organizational boundaries or physical distance decreases process efficiency – but only when it comes to abandoned projects. Our study thus contributes a more detailed understanding of the net benefits of dispersed teams working together, as consistent with work suggesting that team motivations are crucial to success and overcoming the challenges of coordinating across physical distance (as indeed suggested by Hoekman et al., 2010) or organizational boundaries.[2] Our work also has practical implications by revealing that dispersed

[2] It is important to acknowledge that while organizational variety and distance are commonly employed theoretical constructs, each is potentially conflated with other constructs such as time zone compatibility (e.g., Espinosa and Carmel, 2003) or variance in cross-national cultural/legal systems. In the supplementary data appendix, we explore how these measures are related in our data.

research teams may have a hidden cost by holding up resources that could have been reallocated to other projects.

2. Dispersed Team Work and the Coordination Challenge

In the context of scientific teams or open source communities, teams by design often work across geographic and organizational boundaries (Butcher and Jeffrey, 2007; Le Gallo and Plunket, 2020). Technological advances in communication technologies have resulted in lower barriers to communicating across geographic distance (Hoekman et al., 2010). Furthermore, in contexts such as open source communities that are voluntary and where participants come together to solve complex problems or generate new ideas, individuals often participate and work together in teams both within and across organizational affiliations. Prior work on dispersed research teams tends to consider both physical distance and organizational or institutional affiliation as dimensions of dispersion (e.g., Bercovitz and Feldman, 2011; Le Gallo and Plunket, 2020). We now go on to discuss these two dimensions of team dispersion and consider how they affect process speed.

Extant work would suggest that team dispersion may incur coordination challenges. We hypothesize that such coordination challenges will result in slower process speed. Yet, we also propose that there may be a selection effect where teams only choose to work together across boundaries if they believe the idea is worth the additional effort. This may result in selection of projects that are more likely to succeed and faster process speed. We therefore introduce competing hypotheses as to how team dispersion affects process speed. Finally, we hypothesize that the selection effect, whereby teams choose projects that are more likely to succeed, will lead process speed to differ depending on the outcome of the team.

2.1 Team collaboration across organizational boundaries

It is well established that when teams in open innovation or scientific contexts are composed of individuals with different organizational affiliations; their differences will lead them to better outcomes and they are more likely to be successful (Butcher and Jeffrey, 2007; Bercovitz and Feldman, 2011). There are a number of reasons why teams benefit from collaborations across organizational boundaries. Team members that are part of different organizations will be exposed to different information and knowledge than that circulated within a particular firm or research institution (Butcher and Jeffrey, 2007; Bercovitz and Feldman, 2011). It has also been argued that collaborations dispersed across different organizations allow for team members with complementary and/or unique skills to work together. Team members can therefore assist and compensate for each other's weaknesses and solve complex problems (Giuri et al., 2010; Vural et al., 2013). Finally, when teams come together across formal organization boundaries they will likely bring access to diverse resources (Butcher and Jeffrey, 2007; Faraj et al., 2011). Despite the benefits associated with collaborations dispersed across organizations; prior work has also recognized that dispersion of research teams leads to coordination challenges.

When collaborators work together in dispersed teams, different mechanisms may be at play that affect the ability of the team to coordinate. In particular, diverging backgrounds and experiences may lead to communication failure as team members lack common knowledge and shared experiences (Cramton 2001; Kotha et al., 2013). Differences in organizational knowledge backgrounds may also create difficulties integrating knowledge and thereby affect the ability to apply it in new contexts (Espinosa et al. 2007; Majchrzak et al. 2012; Pershina et al., 2019). Coordination challenges that "require shared understanding and common knowledge" (Kretschmer and Puranam, 2008: 862; see also Cramton 2001) are often viewed as a cognitive challenge (Pershina et al., 2019). When crossing organization boundaries there may also be differences in organizational incentives or monitoring that create further barriers to successful collaboration and therefore lead to coordination difficulties (Bercovitz and Feldman, 2011). When teams are spread across distinct organizations they may also face difficulties coordinating their joint production efforts and recent work has shown that even the anticipation of such coordination challenges may impact the success of the team (Kotha et al. 2013).

2.2 Team collaboration across geographic distance

Research teams dispersed across geography have also been shown to have advantages in terms of access to diverse knowledge domains or resources (Katz and Martin, 1997; Gibson and Gibbs, 2006; Espinosa et al., 2007), or bringing together unique and/or complementary skills or perspectives (Daniel et al., 2013) that allow them to solve new or complex research problems.

Team coordination is also challenged when collaborators are located in different geographies (Tzabbar and Vestal 2015). Not only do teams that operate across geography have to deal with interaction across physical distance, but potential interaction across time zones, cultures or language make it harder to communicate and coordinate tasks (Montoya-Weiss et al., 2001; Espinosa and Carmel, 2003; Espinosa et al., 2007). Teams that are separated by geographic distance may not be able to meet in person as easily as teams that are in close proximity. Physical proximity may in turn facilitate the quality of the teamwork and ultimately affect the efficiency of the team, for example, in terms of delivering on budget and schedule (Hoegl and Gemuenden, 2001; Hoegl and Proserpio, 2004; Staats et al., 2012). Teams that are unable to meet and work in person and therefore have to work virtually may also develop temporal trust that is more fragile (Jarvenpaa and Leidner, 1999). Levels of trust have been shown to affect team success on the one hand and coordination costs on the other (Gibson and Gibbs, 2006; Peters and Karren, 2009). The experience of the team may also affect their ability to work together across geography. Prior research has shown that when teams are inexperienced in working together this may hamper their success (Bercovitz and Feldman, 2011; Dahlander and McFarland, 2013; Majchrzak et al., 2012). Finally, language and time zones may play a role in coordinating across geography. Espinosa and Carmel (2003) argued that in the context of software development, separation across distance is symmetric and normally teams work using the same technical language. Therefore, it is relatively easy to coordinate and dispersed teams can work well. However, separation of teams across time zones is asymmetric and may hinder the ability to work effectively and it therefore becomes more important to coordinate activities (Espinosa and Carmel, 2003).

In sum, prior work on dispersed research teams has argued that coordination across organizational boundaries or geographic distance becomes increasingly challenging, the more dispersed the team. In line with this, we argue that the coordination challenges will increase with dispersion and therefore lead to lower process efficiency that delay project outcomes. We therefore expect that teams dispersed across organizational boundaries or geographic distance will have slower process speed before reaching a project outcome. In formal terms,

Hypothesis 1a: The greater the dispersion of the team in terms of a) organizational affiliation or b) geographic distance; the slower the process speed before the team comes to resolution, relative to non-dispersed teams.

2.3 Selection Effects and Process Speed of Dispersed Teams

While hypothesis 1a posits that dispersed teams - relative to non-dispersed teams - are likely to be hampered by coordination costs that lead to lower process efficiency, there are also arguments to be made for why dispersed teams might have greater process efficiency. Given that dispersion will lead to various coordination challenges, it is reasonable to expect that team members might be hesitant about working in dispersed teams as they likely will know that it may require some form of coordination effort. It has been argued that research teams may anticipate barriers to coordination and that this may impact team outcomes. Kotha et al (2013), in their study of university inventions, argued that scientific teams may in fact anticipate coordination challenges and such anticipation may impact the success of the team. O'Leary and colleagues in their paper on team productivity also posited that when team members know "that they have smaller fractions of each other's time, and knowing that the coordination of that time will be challenging, team members develop ways to accomplish more in less time" (2011: 467). Hence, teams likely make calculated decisions when they collaborate with others that work for different organizations or are geographically remote from themselves. More concretely, given that dispersion imposes known difficulties in aligning tasks, we would not expect individuals to voluntarily undertake projects that involve dispersed team members unless if they believe the probability of success is high. For performance efficiency outcomes, such as publication of a research article or completion of a research project, we therefore have a classic omitted variable endogeneity problem: dispersion may simply be a proxy for unobserved project quality, which is the real factor driving performance.

To be clear, we expect that dispersed teams only choose to work together if they, for example, have complementary skills or good ideas and that this selection will help them overcome the challenges related to coordination. Specifically, and in line with prior work, dispersed teams anticipate that the benefits of working together outweigh the costs of coordination. Hence, we expect that dispersed teams will have higher ex post observed process efficiency as their complementary skills lead them to be successful or the quality of the project idea is such that they easily resolve the project. In formal terms,

Hypothesis 1b: The greater the dispersion of the team in terms of a) organizational affiliation or b) geographic distance; the faster the process speed before the team comes to resolution, relative to non-dispersed teams.

In hypotheses 1a and 1b, we consider the process speed of dispersed relative to non-dispersed teams before a project comes to resolution: regardless of whether or not the outcome is successful. When turning to behavioral outcomes such as time allocated to the project we argue, however, that ex ante beliefs about quality of a project may produce distinct patterns for successful and failed projects. Put succinctly, the characteristics that make a dispersed team attractive may shift from positive when things are going well, to negative when things are going poorly. In the next two subsections, we therefore separately consider and hypothesize about successful and failed projects.

2.4 Successful Dispersed Team Efforts and Process Speed

As argued in the previous section, team members are unlikely to be ignorant of the challenges associated with dispersed coordination. In fact, much prior work on dispersed teams has tended to assume that dispersed teams come together when the incentives are such that the benefits are likely to outweigh any costs of coordination (Kretschmer and Puranam 2008). As Vural and co-authors observe; "successful collaborations occur when the benefits from collaboration outweigh the costs of coordination that team members face" (2013: 13). Hence, we expect that in voluntary entities such as an open source or scientific research community where individuals can choose with whom they collaborate (Faraj et al., 2011), individuals will be unwilling to enter dispersed collaborations unless they expect that there is an upside to the collaboration. The expected upside may come from the belief that they are working on a superior idea or a project that is likely to be successful. Alternatively, individuals may decide to work together because they have complementary skills or bring unique resources to the table that will allow them to be successful or provide opportunities for learning (Giuri et al., 2010; Faraj et al., 2011; Mindruta, 2013). Regardless of the motivation; their expectations of success are likely to lead them to invest additional effort in aligning the tasks needed to complete the project. This in turn will lead to greater process efficiency. In other words, the expected upside means that once individuals enter such a collaboration they are likely to be efficient in completing the project.

In contrast, non-dispersed teams that by definition are in close proximity to each other or work within the same organization will likely be aware that little effort is needed to align activities or tasks to complete the project. Hence, they may expend less effort to complete the project than dispersed teams, which may translate into lower process efficiency. Moreover, non-dispersed teams may come together simply by the fact that they are in close proximity to each other (Hoekman et al., 2010). Hence, they may not be as selective on the composition of skills of the team, as dispersed teams. Consequently it may take non-dispersed teams longer to reach project resolution. We therefore predict that,

Hypothesis 2: The greater the dispersion of the team in terms of (a) organizational affiliation or (b) geographic distance; the faster the process speed of teams that succeed, relative to non-dispersed teams.

2.5 Failed Dispersed Team Efforts and Process Speed

Prior work on research collaborations has tended to focus on successful research efforts. Nevertheless, some collaborations fail (Butcher and Jeffrey, 2007). Research teams may end up pursuing ideas or problems that are unconventional or alternatively too difficult for the team to solve. Regardless of why a team fails; some teams are simply not able to succeed with the task they have set out to solve. A failing project will hold up resources including team members' time and effort, and these resources could be invested in new research projects. In research on innovation, failing is often encouraged as it is a process through which learning occurs. When failure happens quickly, resources are freed up to engage in new projects (e.g., Khanna et al., 2016; Guler, 2018).

As already argued, engaging in dispersed teamwork may require the participants to invest more effort than non-dispersed teams to overcome any challenges related to coordination of their work. This additional effort may result in sunk costs that the participants will be reluctant to write off (Gunia et al., 2009). Hence, they may be unwilling to abandon their efforts, even if it turns out the project is not going so well. The continued investment in the project may also lead to escalation of commitment over time (Staw, 1976). As teams become more involved in the project and have resources tied up in the effort, they may be less willing to abandon the project, even if it means freeing up resources that could have been invested elsewhere (Guler, 2018). We therefore argue that dispersed teams will be less willing than non-dispersed teams to abandon their failing projects. This in turn will result in them taking longer to reach the failed outcome, that is, reduced process efficiency.

If there is indeed a selection effect, where dispersed teams only choose to work together on what they believe to be successful projects, then there may also be an issue with coordination of the failure itself. Espinosa and Carmel (2003) argue that coordination involves communication, which becomes more difficult when faced with non-routine issues or working asynchronously. As long as a project is progressing according to plan, it should be easy to communicate and therefore to coordinate tasks. Yet, if a project starts to fail, this may require additional communication in order to correct course or reach consensus on quitting. Such communication may be relatively easy in proximate teams that can work synchronously, but prove more difficult when teams are dispersed across organization boundaries or geographic distance. Hence, we expect that dispersed teams working on failing projects are likely to face lower process efficiency and that this will lead them to take longer than non-dispersed teams to abandon the failing project.

On the contrary, we expect that non-dispersed teams quickly will be able to recognize when projects should be abandoned. This is because they face few or no barriers to coordinating tasks or meeting in person; nor face the communication challenges associated with dispersed teams. Hence, we expect that:

Hypothesis 3: The greater the dispersion of the team in terms of (a) organizational affiliation or (b) geographic distance; the slower the process speed of teams that fail relative to non-dispersed teams.

3. Context and Data

We study collaborations in the Internet Engineering Task Force (IETF). In this open innovation community, participants work in teams to develop and maintain Internet standards. Similar to other open innovation communities such as open software development, any interested individual can participate and documents are freely available on the Internet. The ethos of the IETF is perhaps best captured by its unofficial slogan: "We reject kings, presidents and voting. We believe in rough consensus and running code" (Hoffman and Harris, 2006: 5). Our interpretation of the slogan, bolstered by our own observation of IETF meetings and archives, is that the IETF functions very much like a scholarly community. While there is an organizational structure and a hierarchy, problem identification and solution occur from the bottom up and relies on persuasion rather than fiat.

The main product of the IETF is Request-for-Comments (RFC) publications.[3] As of January 2017, the IETF had published over eight thousand RFCs, each of which either describes a technological standard, or has informational content regarding networking practices or technology. All RFCs are authored by individuals and enter the publication process as "Internet Drafts", i.e., documents submitted to the organization as candidates for publication. There are two paths to publication. First, individuals can attempt to submit a document via one of the IETF's ~150 chartered working groups. If the working group Chairperson(s) consents, such documents are labeled "Working Group Drafts" and affiliated with the group. Ultimately, working group drafts must be approved by a group known as the Internet Engineering Steering Group (IESG) in order to be published. Note that according to IETF by-laws all standards documents must go through the working group submission process. The second path to RFC publication is via the "Individual Draft" process. Documents submitted to the IEFT with topics that do not fit within the domain of a working group are routed to the RFC Editorial Board, which consults with the IESG, but makes the final decision on publication. As an example of the type of data we are working with, Figure 1 shows the header section of a draft: 'draft-worster-mpls-in-ip-00.txt', that appears in our data.

---------- Insert Figure 1 about here ----------

All Internet Drafts go through an open review process – there is no official handling editor or blinded review, and community members at large can choose to offer feedback or ignore these submissions. Draft submissions typically go through multiple rounds of revisions as the authors refine their ideas or respond to feedback. Documents are not officially rejected, as authors retain the right to continue revising and resubmitting documents not yet accepted for publication. As we will discuss below, when measuring project duration, we will examine the date from the

[3] RFC is a historically deprecated term. In current parlance, it simply means an IETF publication.

initial submission, to the date of publication or the expiration date of the final non-updated revision.

While we pool the working group and individual submissions in our analysis, there are differences. Generally, the individual drafts go through fewer revisions, but are less likely to get published. The collaborative working group drafts have a mean revision rate of 79.7% and a mean publication rate of 40.3%. For the collaborative independent drafts, the mean revision rate is 45.4% and the mean publication rate is 7.7%. Working group drafts are on average considered to be "more serious", whereas the independent drafts have variable quality and maybe even variable goals – i.e., the authors may not care that much about publication. As such they are often ideas of less significance or relevance[4].

Some of the documents are ultimately published, while others are abandoned. Hence, we can document whether a project ends up being successful or not. Moreover, we can observe the number of drafts, that is, the number of iterations that a document goes through, as well as the amount of time elapsed at each step in the process, for both the successful and the failed projects. Hence, we are able to observe how long it takes before a project comes to resolution, but also to separate whether projects come to a successful resolution or are abandoned. Another feature of the data is that we can observe individual authors collaborating across firm boundaries, as well as across geographic distance as authors typically report both organizational affiliations and locations. For example, employees of a firm like Microsoft could team up with other employees in Microsoft or with employees in a different firm. They could work with employees in a proximate or a distant location.

Our final sample consist of drafts that were first submitted from 1995 to 2003, that have 2+ authors and no missing data on geographic or organizational affiliation data and that were not

[4] Of course, this may be true within the working group documents as well, but we are unable to address that here.

published on the first version.[5] These sampling criteria result in the cross-section of 5,250 internet drafts summarized in Table 1. Before turning to the elements in that table, we will provide some more description of the context. To begin, figure 2 shows the new project submissions by date for solo-authored projects, and those that include 2+ authors. The IETF has grown steadily over time, has periodic bursts in activity related to meetings, and drafts with multiple collaborators are the modal type.

---------- Insert Figure 2 and Table 1 about here ----------

Figure 3 shows a histogram of final revision number for every draft version in our data from 1995 to 2003. A plurality of projects is never revised. The figure also shows that the publication rate is much lower at version 1 and increases as the revision number increases. Finally, figure 4 shows a histogram of total project duration in elapsed days for projects that have at least one revision, where final termination accords with the IETF rule of 180 days without activity.

---------- Insert Figures 3 and 4 about here ----------

3.1 Outcome Variables

We are primarily interested in observed process speed, i.e., how long it takes a dispersed relative to a non-dispersed team to reach an outcome. In addition, we examine the difference between successful outcomes, i.e., those drafts that get published, and failed outcomes, i.e., those drafts that are abandoned, for dispersed, relative to non-dispersed teams. Turning to table 1, we therefore include three outcome variables: published, versions and duration. Note at the outset that while we select collaborations from the period 1995 to 2003, all outcomes are observed through 2005. The dependent variable *Published* is a dummy coded as 1 for published drafts. Recall that Internet Draft submissions are never formally rejected for publication and the working group decides if it wants to abandon the project.[6] *Published* is our successful outcome measure,

[5] In our study filtering out collaborative drafts published on the first version results in dropping 50 cases.
[6] Although there is no official rejection process, authors are sometimes informally advised to quit. Unfortunately, we cannot observe such informal communication in our data.

as it indicates that the project has successfully achieved the status of an approved IETF document. While there are distinct types of IETF documents, consistent with extant work (Fleming and Waguespack, 2007) we treat all positive outcomes as equivalent.

To understand *process speed*, we look at the time it takes for a team to reach a successful outcome. *Versions* is the total number of revised documents, i.e., the iterations of the document that are submitted for the focal draft. For instance, for the example draft in figure 1, the authorship team subsequently submitted four revised documents, bringing the total number of *Versions* to five.[7] Similarly, *Duration* measures the total number of days elapsed between first submission and publication, or first submission and draft expiration for projects that are unsuccessful. As a precise termination date for abandoned drafts is not recorded, we follow the IETF convention of treating drafts as expired after 180 days. For example, and again referring to figure 1, the clock for the example starts on 31 August 2001, and expired for the fifth version on 04 November 2002, giving a total *Duration* of 430 days. While our main analysis is conducted at the draft level of analysis, note that in some of the supporting analysis in the supplementary Data Appendix, we examine within draft variables, such as the duration from one version to the next.

3.2 Main Independent Variables

The first main independent variable is an *Author Remoteness Index* indicating the geographic distance between the members of the team. We look at whether author pairs within a particular team are remote from one another in terms of geographic distance. Figure 5 shows an example of the author information that we extracted from each internet draft. For authors based in the USA, we determine latitude and longitude based on zip code, city and state, or phone area code, whichever is most precise. For authors outside the US, we use city name, or the capital of the

[7] We track individual drafts through multiple versions primarily via file naming conventions. For example, draft-worster-mpls-in-ip-00.txt is the first version of a single draft, while draft-worster-mpls-in-ip-04.txt is the fifth version. There are some cases where file name changes, either due to an individual draft changing owner or shifting to working group status. In those cases, we rely on the RFC editor tracking system to identify related drafts.

country if more precise information is not listed. Once we have determined the city location of each author, we code the longitude and latitude of that location. We then calculate the distance in miles between each pair of authors. We generate an *Author Remoteness Index*, which is a count of the number of author pairs within a draft that are further than 2,500 miles apart, divided by the total number of author pairs within a draft. As shown in Figure 6, IETF contributors come from all over the world, with 41 unique countries represented in our sample.

Calculating a parsimonious geographic dispersion for teams with more than two members is not straightforward. Our reading of the extant literature on geographic remoteness is that measures generally fall into two types. The first type are pairwise measures of physical distance in miles (e.g., Singh, 2008; Hoekman et al., 2010; Catalini, 2017). Second, Herfindahl like indices of the extent to which participants operate in different regions, regardless of physical distance are used when three or more underlying elements are potentially involved (e.g., Tzabbar and Vestal, 2015; Kafouros et al., 2018). Note that the triplet+ indices are typically intended to operationalize differences in background or shared environment rather than increasing anticipated coordination difficulty per se. For a concise measure of physical distance in triplets+, O'Leary and Cummings (2007) propose utilizing mean distance among pairs, but we were unable to find a similar paper employing mean team distance. The closest match to our work is Gittelman (2007) where mean distance is calculated among team members, but then converted to a binary measure of average distance greater than 50 miles, 50-800 miles, and so forth. We elected to deploy a remoteness index of 2500+ miles because such a measure, at least in our setting, better captures the construct of coordination difficulty in teams greater than two. In the supplementary online Data Appendix, we discuss and systematically compare 600 alternative distance thresholds, all of which produce consistent results.

---------- Insert Figures 5 and 6 about here ----------

The second main independent variable is *Author Different Organization Index.* In order to explore the effects of teams working together across organizational boundaries, we create a different organization index, calculated as the number of author pairs within a draft that have different organizational affiliations, divided by the total number of author pairs within a draft. Figure 5 again shows an example of the author affiliation information we extracted. We coded organizational affiliation data from root email addresses. In total the 17,818 authors in our sample represent 1,715 different organizations.[8] [9]

3.3 Control Variables

We also include a number of control variables in our analyses. First, we consider whether the team primarily consists of *US Authors*. US based authors represent 65.7% of the sample, and moreover arguably have an advantage in terms of publication as the IETF is a predominantly US-based institution. Second, as prior work has shown that the experience of the research team is important for success (Bercovitz and Feldman, 2011); we consider the experience of the authors along different dimensions. First, we include a control for whether the authors have prior experience working together. *Author Prior Index* is measured in the same manner as our main independent variables – the percentages of author pairs on the focal draft that have worked together in the past. The authors in a team might also have prior experience publishing in the IETF, which may give them an experience advantage in terms of publication. We therefore also include a control for the number of logged *Prior Publications* with the IETF by the authors on the team. We also control for *Organizational Presence* by estimating the number of times an organization is observed in the data, as it might be argued that teams with authors from larger or more powerful organizations are more likely to get published (Comino et al., 2007). We measure

[8] Some authors may use personal email addresses. At the project level, 42 cases in our sample include a co-author who lists a "yahoo.com", "hotmail.com", or "gmail.com" address. Excluding those cases has no bearing on the results reported below.

[9] While we consider geographic distance and organizational boundaries as two different dimensions of dispersion; in supplementary analyses we consider how the mean values of organizational dispersion vary together with distance (see section 2.1 in the online appendix). We see that the propensity for the author pair to work for the same organization is greatest when the authors are 0-100 miles apart. Beyond the 10-mile threshold the propensity for cross-organizational collaboration increases rapidly, but beyond the 0-100-mile bucket there is no obvious relationship between distance and organizational dispersion.

this variable as a count of the number of author credits for the organization for drafts submitted in the prior trimester. We also include the logged *file size* of the draft. We consider this as a rough indicator of the amount of effort the authors put into each draft, where larger file size indicates that a draft is more complete.

As the process efficiency of a team may be affected by the complexity or novelty of the project, we also control for observable project content. Each draft has keywords in its name that describe the technology category or subject addressed, such as "dhc" (dynamic host configuration) or "security". We employ these keywords in several ways as proxies for both the complexity and the novelty of the project. First, we control for the complexity of the project by including a count measure of the number of keywords; *Keyword Count*. Second, we include two different measures for the novelty of a project, both in which we observe the temporal order in which keywords appear at the IETF. The first, *Keyword Novelty,* records the minimum order value for all keywords associated with a draft. A value of one, for instance, means that this is the first time this keyword has appeared in IETF submission. The second variable, *Combination Novelty*, is computed as the minimum order value for all keyword pairs on the draft.

Finally, we address several additional structural aspects of the project team. Logged *simultaneous drafts* is intended to capture the extent to which team members' current efforts are split in different directions, and counts the number of new drafts on different topics submitted by co-authors in the next six months. Working group submissions have a higher rate of publication and so we also include a control for *Working Group* submissions. Prior work has argued that larger teams are more likely to solve problems in open source communities (Comino et al., 2007). Table 1 also includes a count of total *Co-Authors* on the draft, although this is included in the regression analysis as a fixed effect, ranging from teams of 2 people to those with 10+[10].

[10] In further analyses we explore the potential correlation between team size and our dispersion measures and find no evidence that dispersion factors function substantially differently across teams of different sizes (see the online Data Appendix).

3.4 Analyses of the Data

To avoid bias in subsequent selection of teams over the period of collaboration of the team, we run cross-sectional analyses on our sample of data based on the first submission of a draft in each project. As prior work has argued that dispersed teams are more likely to be successful (Bercovitz and Feldman, 2011; Le Gallo and Plunket, 2020), we first run a base model testing whether dispersed teams relative to non-dispersed teams are more likely to publish. The first dependent variable is a binary outcome variable for publication (i.e., success of the collaboration), and we therefore run a logit model testing the likelihood of publication. The second stage of our study aims to observe the process efficiency of cross boundary collaboration and is based on time or the speed with which projects reach an outcome. First, we observe the number of revision versions that a document goes through, measured as project version. Second, we observe the number of days until publication (or failure measured as abandonment) of the draft. For project version we run a robust Poisson model as this is a count variable, and because the distribution of the data is skewed (non-normal) as shown in Figure 3. Note that our estimates for versions remain consistent when using negative binomial or ordinary least square (OLS) regression. For project duration we run an OLS regression model. The results of the analyses are discussed in detail below.

All regressions include the full suite of control variables, as well as fixed effects for time period, as it could be argued that some periods might be better for publication than others. We also incorporate co-author number fixed effects as the number of co-authors on the team might affect the ability of the team to coordinate effectively. In all models we deploy robust standard errors.

4. Results

The main results are reported in table 2. Table 2, model 1 reports the likelihood of publication, a measure of success in the IETF setting. To reiterate, the data are a cross-section consisting of the first version of the draft that was submitted for projects with 2+ authors in the period 1995-2003, with outcomes measured through 2005. The results show that more dispersed projects (relative to non-dispersed projects), both in terms of geographic and organizational diversity are more likely to get published. This finding is in line with prior work that has argued that dispersed teams are likely to benefit from bringing together team members across boundaries (e.g., Le Gallo and Plunket, 2020). In order to determine effect sizes, we calculated the marginal effects, based on the model reported in table 2 model 1. We find that when the author remoteness index increases from zero to one, the publication likelihood increases by 3.9% from 18.2% to 22.1%, and a unit change in the different organization index increases the publication rate 5.6% from 15.3% for the same organizational affiliation to 20.9% for different organizational affiliations. The effects are statistically significant.[11]

Looking at the control variables, we see that those project teams that have US authors are more likely to get published. We do not find that teams with higher organizational presence are more likely to get published. However, the results do show, in line with extant work, that previous experience working together as a team or prior publications within the IETF improves the chances of publication (Bercovitz and Feldman, 2011). The file size of the documents is also positively associated with publication, perhaps indicating that completeness matters for initial odds. Among our keyword-based measures, the simple keyword count is negative and significant,

---

[11] In alternative models we also included an additive dispersion effect including both geographic distance and the organizational diversity index. This index is created by taking the number of author pairs within the team that are both remote and belong to different organizations and dividing by the total number of author pairs in a team. This term has a negative and significant effect on publication, which we interpret as indicating that if there is too much dispersion in the team then it has a negative effect on publication. In other words, the coordination costs are so high that the likelihood of publication decreases. This is consistent with prior work, which has shown that boundary spanning may be good for innovation, but if there is too much diversity in the team this may become detrimental as it becomes difficult to allocate attention and integrate heterogeneous information (e.g., Dahlander and Frederiksen, 2012).

perhaps indicating that projects with broader scope are inherently more difficult to publish. Lastly, author teams that introduce more new projects in the near future decrease their odds of getting the focal draft published.

---------- Insert Table 2 about here ----------

Having established in a baseline model that team dispersion is, as expected, positively associated with outcome efficiency, we next turn to the question of process efficiency. Models 2a, 2b, 3a, and 3b in Table 2 show the estimates for project version and duration. Focusing first on control variables, we find that author teams with more experience working together are faster at publishing. By contrast, teams that have prior experience publishing, as well as those who are part of more powerful organization, take longer to publish. We speculate that these results may reflect a tendency for such teams to take on more technically or politically complicated projects. The file size of the projects also matters for process efficiency: Projects with larger file size go through more revisions and take longer to publish. For the project content measures, the simple keyword count is negative and significant, indicating that projects with broader scope go through fewer versions and take less time to publish. The novelty of the projects does not affect the number of revisions, but novel projects take longer to publish. Finally, author teams that introduce more new projects in the near future are significantly faster at publishing.

We now turn to the testing of the hypotheses. First, we test the competing hypotheses 1a and 1b to understand the main effects of team dispersion on project efficiency. The results are reported in Table 2 models 2a and 3a. Table 2 model 2a shows the estimates for project version and model 3a the estimates for project duration. Neither of the coefficients for the dispersion measures are significant and we therefore find no support for the main effects of team dispersion on project efficiency. Hence, at least in our setting, dispersed teams on average seem to be no more or less efficient in general at reaching an outcome than non-dispersed teams. Moving on to hypotheses 2 and 3, and as explained in more detail below, this result is likely explained by the selection effect related to dispersed teams.

In Table 2 models 2b and 3b, we report the results testing hypotheses 2 and 3. Table 2 model 2b shows the estimates for project version. For easier interpretation of those results we report marginal effects calculations for the changes in the key independent variables of interest. Among published papers, we find that organizationally diverse (Auth Diff Org Index = 1) project teams go through 4.84 versions, while non-diverse teams (Auth Diff Org Index = 0) increase to 5.58 versions. This relationship flips for unpublished manuscripts, with organizationally diverse manuscripts taking longer to fail at 2.65 versions relative to 2.43 versions. A similar pattern is present for geographic distance: for published manuscripts distance decreases the number of expected versions from 5.23 to 4.69, while in unpublished manuscripts geographic distance increases the number of expected versions from 2.53 to 2.71.

Table 2 model 3b shows the estimates for duration. We again report marginal effects here. Generally, published drafts persist longer than abandoned drafts. Even so, published drafts by an organizationally diverse team generally take 110 *fewer* days to publication than publications by non-diverse teams (561.00 vs. 671.29 days). Similarly, geographically remote teams take 52 *fewer* days with 563.62 days to publication than non-remote teams with 615.72 days to publication. For the abandoned projects, we find that organizationally diverse projects take 33 *more* days, at 422.76 days total, before they are abandoned than the non-diverse projects (389.91 days). Similarly, the geographically remote projects take 25 *more* days to abandonment than the non-remote projects (429.25 vs. 404.23).

These differences are also illustrated in figure 7, which shows the predictive marginal effects for published versus not published drafts. We see that those drafts that are finally published have fewer versions and are published faster, the higher dispersion of the project team. This seems to indicate that successful dispersed team projects are more efficient, presumably reflecting an ability to overcome coordination challenges. On the contrary, we see that unsuccessful projects (those that are abandoned) take longer to fail and go through more versions, the more dispersed the project team. In sum, our results show support for both hypotheses 2 and

3. The fact that we find support for hypotheses 2 and 3, but not for hypotheses 1a and 1b suggests that in line with our theory there is a selection effect where team members choose to work on dispersed projects if they deem that they are likely to be successful. Hence, we are only able to observe the differential effects of dispersion once we separate projects into successful and failed projects.

---------- Insert Figure 7 about here ----------

4.1 Within Project Patterns of Process Efficiency for Dispersed Teams

The preceding analyses are cross-sectional and examine the relationship between dispersed teams as measured by organizational diversity or geographic distance and project success, the number of revisions, and total days in process. For additional insight on coordination mechanisms, in this section we turn to analysis at the revision level. Figures 8, 9, and 10, present bin-scatter results for file sizes (figure 8), time between revisions (figure 9), and propensity to abandon the project (figure 10) by draft version number and collaboration type. For these figures we restrict the analysis to the first six versions of each draft, which excludes the 5% of cases in version seven or greater, and treat collaboration types as binary variables, distinguishing zero and any organizational diversity (*single organization* or *multiple organizations*), and zero and any remote team members (*co-located* or *distant*). For example, the left panel of figure 8 shows that *single organization* drafts submit documents with a larger *filesize* at each *version* level.

We draw two substantive conclusions from considering figures 8-10 together. First, at each step in the process, dispersed teams on average produce more complete or complicated work, as indicated by the increase in file size, and they also revise their efforts more rapidly, as indicated by the days to the next revision. Second, in figure 10, we see that dispersed teams are substantially less likely to quit, particularly earlier in the process. Taken together, the figures suggest a particular type of coordination problem that is hinted at in table 2 but is more explicitly displayed here: dispersed teams hold on to failing projects for longer than non-dispersed teams. There could be one or several underlying mechanisms that might explain this process outcome for dispersed teams. Perhaps dispersed teams are more likely to make early investments in the projects that they are unwilling to forgo, i.e., sunk costs. Dispersed teams are perhaps also more likely than non-dispersed teams to suffer from escalation of commitment, as the dispersed team invests more effort in coordination as the project progresses. Finally, perhaps the outcome we are observing is really an indicator of the coordination challenges that arise in dispersed teams as they are unable to reach agreement on when it is time to abandon a failing project.

5. Discussion

In this paper, we examined the process efficiency of dispersed teams that work together across organizational boundaries and geographic distance. We first considered whether dispersed teams are more likely than non-dispersed teams to have slower process speed due to the coordination challenges associated with dispersion (hypothesis 1a). Alternatively, we proposed that dispersed teams are likely to be aware of the potential coordination challenges related to dispersion and they may therefore invest effort to overcome such challenges. In turn, this will lead dispersed teams, relative to non-dispersed teams, to have faster process speed (hypothesis 1b). We find no support for hypotheses 1a and 1b, a result we attribute to the fact that the authors self-select into dispersed team collaborations (Shaver, 1998). Hence it is likely that they are aware that such collaborations potentially entail challenges related to coordination, and therefore tend to select into dispersed teams only when the idea has greater ex ante quality or entails lower ex ante risk (Laursen et al., 2011).

With selection effects in mind, we go on to distinguish between successful and failed projects. Hypothesis 2 related to process efficiency for successful dispersed teams and our findings show that among successful projects, dispersed teams reach success with fewer iterations and in less time than non-dispersed teams. This finding indicates that - at least in our setting – successful dispersed teams are more efficient than successful non-dispersed teams. For hypothesis

3, we find that dispersed teams that fail take longer to abandon their project relative to non-dispersed teams.

We next examine the behavior of the teams within successive iterations of projects (figures 8 to 10) and see evidence that dispersed collaborative teams are unwilling to let go of their idea or to forgo any upfront investments in the project. If the authors only engage in dispersed project with more certain outcomes because of the challenges related to dispersion; this also means they are more likely to keep engaging in the project even when it starts to become apparent that it is failing. In other words, the continued engagement in the project is likely to lead to escalation of commitment (Staw, 1976). This escalation of commitment also leads them to cling on to failing projects.

In addition, the concern related to coordination challenges will likely lead authors involved in dispersed teams to exert additional effort – early on in the project – that may be considered as sunk costs (Arkes and Blumer, 1985). Sunk costs lead to reluctance by the dispersed team to abandon the project – even if it becomes clear that the project has a low chance of a payoff (Gunia et al., 2009). To be sure, and as we saw in figure 8, at each stage of the project, the dispersed teams submitted more detailed drafts than the non-dispersed teams. While this is by no means conclusive evidence of the sunk cost mechanism, this does seem to indicate that perhaps the authors do anticipate more exertion related to coordination when they collaborate in dispersed teams (cf. Kotha et al., 2013). As a result, dispersed teams invest more effort early on in the project in order to try and minimize coordination challenges. Once they have made such an investment, they become more reluctant to abandon the project.

5.1 The Price of Slow Failure

Our finding that failing dispersed teams take longer to fail than non-dispersed teams sheds an interesting light on the perspective that team dispersion should be encouraged for successful outcomes (Gibson and Gibbs, 2006; Giuri et al., 2010). Our findings (Table 2, model 1) confirm that dispersed teams are more likely than non-dispersed teams to be successful. Yet, the fact that dispersed teamwork both increases the likelihood of success and the duration to failure has interesting implications: For organizations that wish to undertake multiple research projects, slower failure means that they are delayed in re-allocating resources, and therefore have fewer opportunities to initiate new projects (Guler, 2018). Thus, depending on the marginal effect of dispersed team projects on success probability and the cost of failure, an organization may reap more value from a portfolio of research projects with fewer coordination challenges related to dispersion simply because there are more opportunities to achieve success. In other words, over some fixed period of time, the performance benefits associated with dispersed teams may be offset by having fewer opportunities to reallocate resources.

In our results, some simple scenarios illustrate the net effect on cumulative outcome efficiency of dispersed teams where individual projects have greater expected success but are slower to fail. Imagine a situation where a team dispersed across organizations (team A) and a non-dispersed team (team B) each work on one project at a time for 10,000 days. Based on table 2, over the observation period we would expect team A to produce 4.82 positive results, and team B to produce 4.11 positive results.[12] In other words, team A produces 1.17 (=4.82/4.11) more cumulative positive results, which is good but lower than the 1.22 ratio (=.221/.181) suggested by the single project performance analysis. This reduction in total outcome efficiency comes about from team A being slower to reallocate resources after a failure, resulting in fewer attempted projects over the observation period. Moreover, if we tweak the parameters to decrease team A's single project boost to19.0% or failure duration to 529 days, the expected long run positive results ratios between teams A and B become identical.

[12] Positive results = (10,000 days / ((success rate* success duration) + ((1-success rate)*failure duration))) * success rate. Based on table 2, for team A the success rate = .221, success duration = 563 days, and failure duration = 429 days. For team B the success rate = .181, success duration = 615 days, and failure duration = 404 days.

The point of this exercise is to show that the observed performance boost associated with any research strategy for a single project does not tell the whole story. When it is expected that resources are reallocated to other projects, challenges related to coordination in dispersed teams that delay failure may become significant, sharply reducing or even reversing the net benefit associated with a research strategy encouraging work in dispersed teams. Moreover, while in our setting the observed result from pursuing dispersed team projects is in net superior, the performance benefits and failure penalties associated may vary in other research contexts. Regardless, our study suggests that the net downside of working on dispersed teams resides in those teams that are unsuccessful: When they fail to abandon an unsuccessful project, they also hold up resources that could have been reinvested elsewhere.

5.2 Limitations

One concern with the literature relating dispersed teamwork to performance is that, in observational settings, dispersed teams only form when ex ante quality is high. This is also an obvious limitation in this paper when we look at the publication outcomes in table 2, model 1, but in table 2, models 2a, 2b and 3a, 3b; our outcome variables related to process efficiency are behavioral and we are therefore less concerned that omitted project quality factors are driving the results. Specifically, we are reasonably confident that endogeneity issues do not undermine the analysis of project termination decisions. The higher selection bar is theoretically probable if participants have expectations about dispersion increasing the challenges associated with coordination, and means that the correlation between project dispersion and performance is potentially spurious. By contrast, our results on variance on the ex post termination decision are explained by unobserved ex ante quality only if dispersed team projects have a bimodal distribution (i.e., if some proportion of dispersed projects have greater average quality, while the rest have lower average quality). We posit that dispersion leading to self-selection of low quality projects is unlikely. Or in other words, the finding that there are variations in the rate of success and failure of dispersed collaborations, suggests that the outcomes are related to the challenges related to coordination and not the quality of the projects per se. Future studies may be able to disentangle the deeper mechanisms at work by analyzing the quality of the ideas, as well as the quality of the authors. Unfortunately, we are unable to examine this further in our context.

Another concern might be the extent to which our findings are generalizable to other contexts than open source communities. For example, one could argue that organizational boundaries have little meaning in this context, as the IETF has developed its own community where the participants understand the terms of interaction and so may not suffer the usual constraints of participation across formal organizations. Yet, while we certainly find evidence that teams that have more experience with publishing in the IETF are more likely to be successful, we find that they are actually less efficient, as those teams have projects that go through more versions and take longer before they reach an outcome. This suggests that organizational affiliation does matter – even in this context. This is likely because the authors that participate in the IETF still have constraints in terms of different organizational requirements, norms and incentives and they therefore still suffer from process inefficiency related to coordinating across organizational boundaries.

5.3 Contributions and Conclusion

One of the main contributions of this paper comes from our examination of process efficiency in dispersed teams. While prior work on dispersed research teams has tended to focus on absolute outcomes such as whether or not a project is successful or innovative, our understanding of the behavioral decisions related to the process of completing a project are still limited. We suggest that in order to better understand how dispersed teams work together, it is also important to understand process related outcomes such as speed to project completion. We show that there may be a tradeoff cost of working in dispersed research teams due to the slower process efficiency of failing projects. Future work on dispersed research teams could look at other process related outcomes that may impede or facilitate process efficiency and thereby generate further insight into when it makes sense to pursue dispersed research efforts.

The fact that we find that process speed differs for dispersed versus non-dispersed teams when we consider the outcome, i.e., failure vs. success, suggests that endogenous selection effects are at play. In other words, the authors in our setting apparently select into dispersed teams only when the idea has greater ex ante quality or entails lower ex ante risk (Laursen et al., 2011). While scholars have tended to treat selection effects as hindrances to causal inference related to performance, we believe our work shows that selection processes also have implications for subsequent team behavioral choices. Future work on dispersed research teams that does more to unpack selection mechanisms may lead to additional findings.

Recent work has made a call for understanding not only successful collaboration efforts, but also those collaborations that fail (Butcher and Jeffrey, 2007). The finding that dispersed project teams may incur additional costs when they fail has implications for how we think about dispersed teams and suggests that in some instances the costs of dispersion may in fact outweigh the benefits. Hence, encouraging dispersion for successful outcomes (e.g., Gibson and Gibbs, 2006; Giuri et al., 2010) may have an unexpectedly high price. Managers working with research teams may therefore need to consider when it might be better to encourage non-dispersed collaboration or carefully design collaborations in dispersed teams. Klingebiel and Rammer (2014) in their study of resource allocation for product innovation suggest that in order to avoid escalating costs of investment, resource commitments need to be limited in order to be able to select out failing projects. Our work shows that investments in dispersed projects may not only escalate, but also endure, as dispersed teams are reluctant to let their projects fail. Hence, managers working with dispersed research teams need to consider carefully how resources are allocated to projects and at what stage each project's resources should be cut off. That dispersed research teams hold on to failing projects also suggests a boundary condition to work on so-called fast failures, which suggests that projects should fail quickly to encourage learning and to free up resources to engage in new projects (e.g., Khanna et al., 2016; Guler, 2018). Yet, this may be hard to achieve for dispersed teams if they are reluctant to abandon their failing projects.

While we cannot pinpoint the exact mechanisms at work here, the longer total time to failure may be an indication that the coordination barriers related to dispersion are particularly salient when teams are faced with failure. In our setting, geographic distance and nation effects like time zones, legal systems, institutional structures, cultural norms, and language may affect the ability of teams to coordinate. In this study, we are unable to disentangle which of these effects play a smaller or larger role for coordination of tasks[13]. Prior work has also established that these mechanisms are likely to be highly correlated and vary together as dispersion increases, hence making it difficult to tease apart the effects (Espinosa and Carmel, 2003). Yet, a significant proportion of the co-author pairs in our setting work across time zones, which means that authors likely will be working on projects at different times[14]. In this setting, coordination across time zones could therefore play a greater part in the coordination challenge than other factors such as language[15]. Espinosa and Carmel (2003) in their conceptual piece on coordination in software teams suggest that coordinating tasks across time zones is particularly complicated, as it leads to asymmetric work on tasks and therefore requires formalization of coordination as well as additional communication. This is particularly true when dispersed teams are faced with non-routine issues. These arguments may help shed light on our findings: When projects are going well; the complexity involved in alignment of tasks in dispersed teams may be easy to facilitate. But when projects start to fail; communication difficulties may be exacerbated by the asymmetric task alignment associated with working across time zones (Espinosa et al., 2007). Consequently, problems may arise in deciding when to terminate the project. In other words, the authors in dispersed teams are simply unable to reach consensus on when it makes sense to abandon a failing project.

[13] We discuss the author pair dispersion in more detail in the supplementary Data Appendix.
[14] Figure 7 in the supplementary Data Appendix illustrates author-pairs collaborating across time zones.
[15] The primary working language on the standards is English.

Finally, our work has important practical implications for the organization and management of dispersed research teams. To the extent that we know more about the process efficiency of dispersed teams it may be possible to design interactions to help increase efficiency. Prior work has shown that putting in place activities to help facilitate coordination in research teams may indeed help reduce coordination costs (Cummings and Kiesler, 2007). Changing tasks to reduce interdependencies or creating multiple opportunities for communication are two ways in which coordination of tasks across organization boundaries can be facilitated (Srikanth and Puranam, 2011). Traversing between different knowledge domains may also help overcome knowledge differences (Majchrzak et al., 2012). Finally, if crossing time zones is a significant barrier to coordination, then meeting in person, i.e., in the same time zone, may still be important. The implication is that when coordinating across time zones, teams may need to consider carefully how work is organized and to what extent there is overlap in time as to when tasks are completed (Espinosa and Carmel, 2003). Future work should explore further how to optimally design and support dispersed research teams to facilitate seamless coordination and increase process efficiency, not only when projects are successful, but also when they should be abandoned for other research efforts.

6. References

Arkes, H. R., C. Blumer. 1985. The psychology of sunk costs. *Organizational Behavior and Human Decision Processes* 35 124-140.

Banal-Estañol, A., I. Macho-Stadler, D Pérez-Castrillo. 2019. Evaluation in research funding agencies: Are structurally diverse teams biased against? *Research Policy* 48 1823-1840.

Bercovitz J, Feldman M. 2011. The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy* 40(10) 81-93.

Bozeman, B., C. Boardman. 2014. *Research Collaboration and Team Science. A State-of-the-Art Review and Agenda.* Springer.

Butcher, J., P. Jeffrey. 2007. A view from the coal face: UK research student perceptions of successful and unsuccessful collaborative projects. *Research Policy* 36 1239-1250.

Catalini, C. 2017. Microgeography and the direction of inventive activity. *Management Science* 64 4348–4364.

Cramton, C. D. 2001. The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science* 12(3) 346-371.

Comino, S., F. M. Manenti, M. L. Parisi. 2007. From planning to mature: On the success of open source projects. *Research Policy* 36 1575-1586.

Cummings J. N., S. Kiesler. 2007. Coordination costs and project outcomes in multi-university collaborations. *Research Policy* 36 1620-1634.

Dahlander, L., L. Frederiksen. 2012. The core and cosmopolitans: A relational view of innovation in user communities. *Organization Science* 23(4) 988-1007.

Dahlander, L., D. A. McFarland. 2013. Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly* 58(1) 69-110.

Daniel., S., R. Argawal, K. J. Stewart. 2013. The effects of diversity in global, distributed collectives: A study of open source project success. *Information Systems Research* 24(2) 312-333.

Espinosa, J. A., E. Carmel. 2003. The impact of time separation on coordination in global software teams: a conceptual foundation. *Software Process Improvement and Practice* 8 249-266.

Espinosa, J. A., S. A. Slaughter, R. E. Kraut, J. D. Herbsleb. 2007. Familiarity, complexity, and team performance in geographically distributed software development. *Organization Science* 18(4) 613-630.

Faraj, S., S. L. Jarvenpaa, A. Majchrzak. 2011. Knowledge collaboration in online communities. *Organization Science* 22(5) 1244-1239.

Fleming, L., D. Waguespack. 2007. Brokerage, boundary spanning, and leadership in open innovation communities. *Organization Science* 18(2) 165-180.

Gibson, C., J. Gibbs. 2006. Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure and national diversity on team innovation. *Admininistrative Science Quarterly* 51(3) 451-495.

Gittelman, M. 2007. Does Geography Matter for Science-Based Firms? Epistemic Communities and the Geography of Research and Patenting in Biotechnology. *Organization Science* 18 724–741.

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

Guler, I. 2018. Pulling the plug: The capability to terminate unsuccessful projects and firm performance. *Strategy Science* 3(3) 481-553.

Gunia, B. C., N. Sivanathan, A. D. Galinsky. 2009. Vicarious entrapment: Your sunk costs, my escalation of commitment. *Journal of Experimental Psychology* 45 1238-1244.

Hoffman, P., and S. Harris. 2006. The Tao of IETF: A Novice's Guide to the Internet Engineering Task Force. *IETF Request for Comments*. 4677.

Hoekman, J., K. Frenken, R. J. Tijssen. 2010. Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy* 39 662–673.

Hoegl, M., H. G. Gemuenden. 2001. Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science* 12(4) 435-449.

Hoegl, M., L. Proserpio. 2004. Team member proximity and teamwork in innovative projects. *Research Policy* 33 1153-1165.

Jarvenpaa, S. L., D. E. Leidner. 1999. Communication and trust in global virtual teams. *Organization Science* 10(6) 791-815.

Kafouros, M., C. Wang, E. Mavroudi, J. Hong, C. S. Katsikeas. 2018. Geographic dispersion and co-location in global R&D portfolios: Consequences for firm performance. *Research Policy* 47 1243–1255.

Katz, J. S., B. R. Martin. 1997. What is research collaboration? *Research Policy* 26 1-18.

Khanna, R., I. Guler, A. Nerkar. 2016. Fail often, fail big, and fail fast? Learning from small failures and R&D performance in the pharmaceutical industry. *Academy of Management Journal* 59(2) 436-459.

Klingebiel, R., C. Rammer. 2014. Resource allocation strategy for innovation portfolio management. *Strategic Management Journal* 35 246-268.

Kotha, R., G. George, K. Srikanth. 2013. Bridging the mutual knowledge gap: Coordination and the commercialization of university science. *Academy of Management Journal* 56(2) 498-524.

Kretschmer, T., P. Puranam. 2008. Integration through incentives within differentiated organizations. *Organization Science* 19 (6) 860-875.

Laursen, K., T. Reichstein, A. Salter. 2011. Exploring the effect of geographic proximity and university quality on university / industry collaboration in the United Kingdom. *Regional Studies* 45(4) 507-523.

Le Gallo, J., A. Plunket. 2020. Regional gatekeepers, inventor networks and inventive performance: Spatial and organizational channels. *Research Policy* 49.

Majchrzak, A., P. H. B. More, S. Faraj. 2012. Transcending knowledge differences in cross-functional teams. *Organization Science* 23(4) 951-970.

Mindruta, D. 2013. Value creation in university-firm research collaborations: A matching approach. *Strategic Management Journal* 34 644–665.

Montoya-Weiss, M. M., A. P. Massey, M. J. Song. 2001. Getting it together: Temporal coordination and conflict management in global virtual teams. *Academy of Management Journal* 44(6) 1251-1262.

O'Leary, M.B., J. N. Cummings. 2007. The Spatial, Temporal, and Configurational Characteristics of Geographic Dispersion in Teams. *MIS Quarterly* 31 433–452.

Pershina, R., B. Soppe, T. M. Thune. 2019. Bridging analog and digital expertise: Cross-domain collaboration and boundary-spanning tools in the creation of digital innovation. *Research Policy* 48(9).

Peters, L., R. J. Karren. 2009. An examination of the roles of trust and functional diversity on virtual team performance ratings. *Group and Organization Management* 34(4) 479-504.

Shaver, J. M. 1998. Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science* 44(4) 571-585.

Singh, J. 2008. Distributed R&D, cross-regional knowledge integration and quality of innovative output. *Research Policy* 37 77–96.

Sorenson, O., J. W. Rivkin, L. Fleming. 2006. Complexity, networks and knowledge flow. *Research Policy* 35 994-1017.

Srikanth, K., P. Puranam. 2011. Integrating distributed work: Comparing task design, communication, and tacit coordination mechanisms. *Strategic Management Journal* 32 849-875.

Staats, B. R., K. L. Milkman, C. R. Fox. 2012. The team scaling fallacy: Underestimating the declining efficiency of larger teams. *Organizational Behavior and Human Decision Processes* 118 132-142.

Staw, B. M. 1976. Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance* 16 27–44.

Tzabbar, D., A. Vestal. 2015. Bridging the social chasm in geographically distributed R&D teams: The moderating effects of relational strength and status asymmetry on the novelty of team innovation. *Organization Science* 26 811-829.

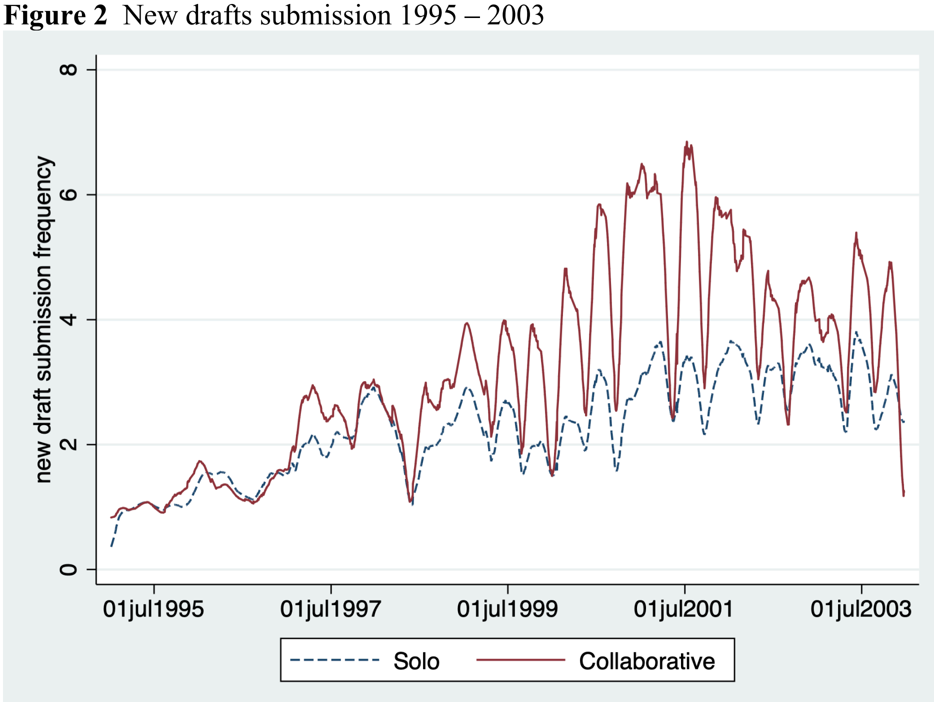
Vural, M. O., L. Dahlander, G. George. 2013. Collaborative benefits and coordination costs: Learning and capability development in science. *Strategic Entrepreneurship Journal* 7 122-137.

Figure 1 Example of Internet Draft Header for draft-worster-mpls-in-ip-00.txt*

```
INTERNET DRAFT                              T Worster, Ennovate Networks
Network Working Group                           A Malis, Vivace Networks
                                                 Y Katsube, Toshiba Corp

                                             Expires December 26th 2000


                      MPLS Label Stack Encapsulation in IP

                       <draft-worster-mpls-in-ip-00.txt>


     This document is an Internet-Draft and is in full conformance with
     all provisions of Section 10 of RFC2026.
```

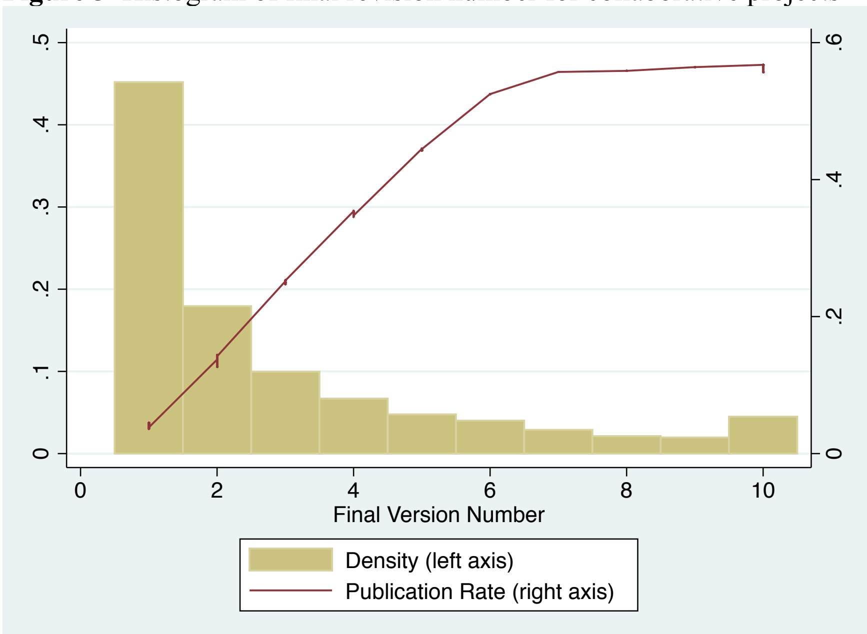
* This draft was initially submitted on June 28, 2000, and went through 4 subsequent revisions. The final revision had an expiration date of August 31, 2001 and was not published, which corresponds to a total duration of 430 days. Each co-author pair works for different organizations and are more than 2500 miles apart.

Figure 2 New drafts submission 1995 – 2003



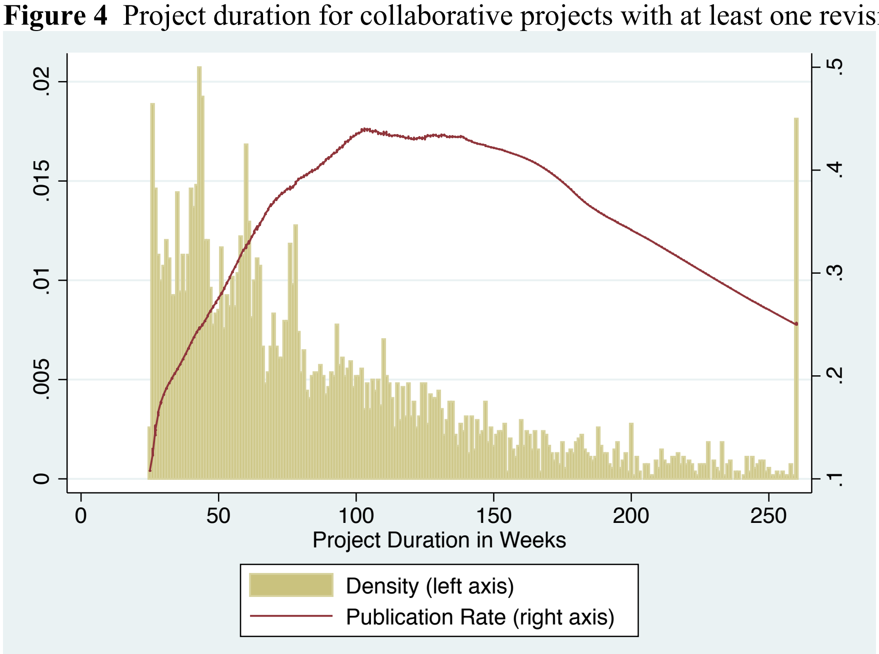
*This figure shows the new project submissions by date for 3714 Solo authored drafts and 5250 collaboratively authored (2+ authors) projects.

Figure 3 Histogram of final revision number for collaborative projects



*The figure shows the final revision count and publication rate for 5250 Collaborative projects. 2127 collaborative projects, 42.23% of the total, were never revised. The publication rate is much lower at version 1, and increases as the revision number increases.

Figure 4 Project duration for collaborative projects with at least one revision

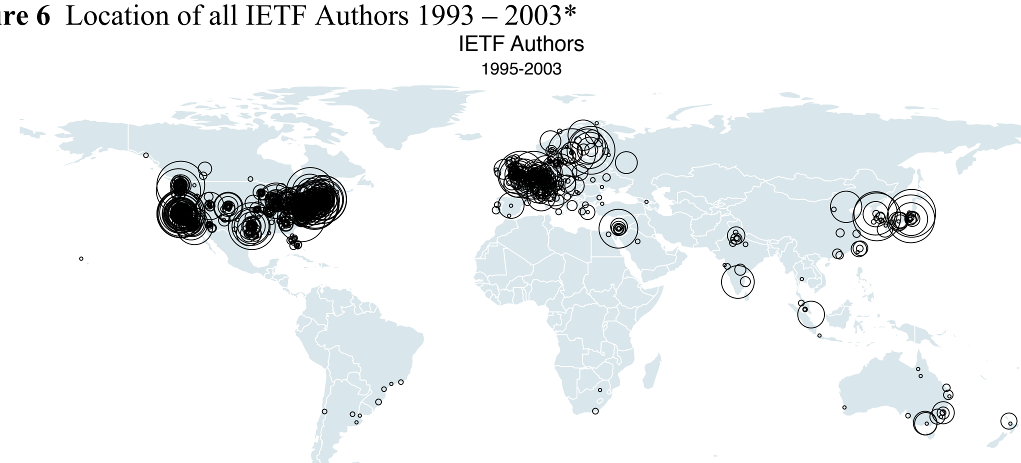


*The figure shows the histogram of total project duration for 3033 collaborative projects with at least one revision for the period 1995–2003. We don't observe precise termination date for the last version of failed projects, but by IETF rules projects are considered to "expire" at 180 days (they may be revived later, but are removed from the roster of active projects). For this figure, we capped the value of cumulative project duration at 260 weeks.

Figure 5 Example of Author Information for Draft draft-worster-mpls-in-ip-00.txt*

```
Authors' Addresses

   Tom Wortster  (contact for comments)
   Ennovate Networks, Inc.
   60 Codman Hill Road
   Boxborough, Mass, 01719
   Email: tom@ennovatenetworks.com
   Tel: +1 978 206 0490
   Fax: +1 978 263 1099




Worster, et. al.           Expires Dec 24th 2000           [Page 5]

                                 Text
Internet Draft    MPLS Label Stack Encapsulation in IP      June 2000

   Yasuhiro Katsube
   Toshiba Corporation
   1, Toshiba-cho,
   Fuchu, Tokyo 183-8511
   Email: yasuhiro.katsube@toshiba.co.jp
   Tel: +81 42 333 2884
   Fax: +81 42 340 8059

   Andrew G. Malis
   Vivace Networks
   2730 Orchard Parkway
   San Jose, CA 95134
   Email: Andy.Malis@vivacenetworks.com
   Tel: +1 408 383 7223
   Fax: +1 408 904 4748
```

Figure 6 Location of all IETF Authors 1993 – 2003*



*Each circle shows the location of authors by latitude and longitude. Each latitude and longitude is rounded to the first decimal place. The size of the circles represents, on a logarithmic scale, the number of times that location is present in the authorship table.

Table 1 Descriptive Statistics and Correlations for Internet Drafts First Submitted 1995-2003 (N=5,250)

		mean	s.d.	1	2	3	4	5	6
1	Published	0.19	0.39	1.00					
2	Versions	3.13	3.15	0.46	1.00				
3	Duration (days)	442.18	382.35	0.31	0.80	1.00			
4	Auth Remote Index	0.23	0.35	0.08	0.08	0.06	1.00		
5	Auth Diff Org Index	0.61	0.43	0.17	0.15	0.13	0.36	1.00	
6	US author(s)	0.77	0.42	0.16	0.13	0.11	0.23	0.23	1.00
7	Auth Prior Index	0.43	0.45	0.09	0.12	0.08	0.08	0.10	0.10
8	Org Presence (100s)	0.54	0.81	-0.05	0.03	0.03	0.09	-0.12	0.19
9	ln(prior pubs)	2.40	1.46	0.06	0.15	0.10	0.22	0.33	0.29
10	ln(file size)	10.31	0.76	0.06	0.16	0.11	0.03	0.05	0.03
11	Keyword Count	2.55	0.85	-0.03	-0.01	0.00	0.02	-0.01	-0.02
12	Keyword Novelty	3.99	7.00	-0.01	0.04	0.04	0.07	0.04	-0.00
13	Comb. novelty	1.08	0.60	0.06	0.11	0.09	0.08	0.06	0.06
14	ln(simultaneous drafts)	1.47	1.06	-0.04	0.03	-0.01	0.14	0.19	0.19
15	Co-Authors	3.32	1.78	-0.02	0.06	0.04	0.09	0.17	0.09
16	WG submission	0.36	0.48	0.40	0.41	0.38	0.08	0.23	0.21

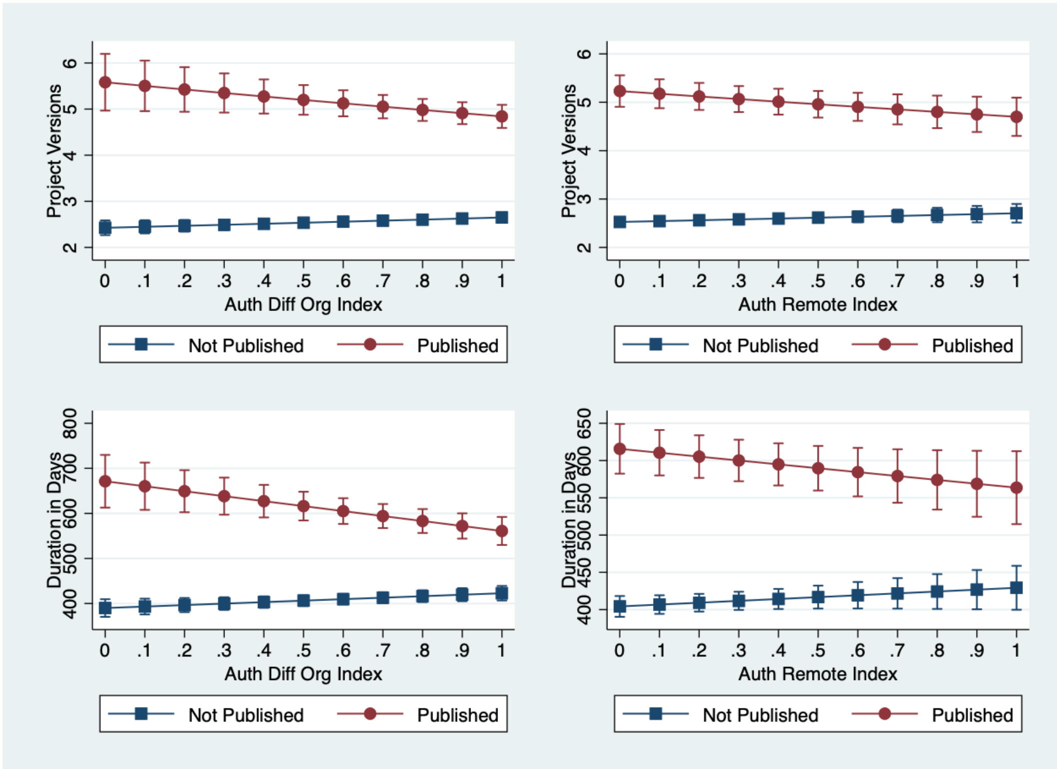
		7	8	9	10	11	12	13	14	15	16
7	Auth Prior Index	1.00									
8	Org Presence (100s)	0.13	1.00								
9	ln(prior pubs)	0.48	0.25	1.00							
10	ln(file size)	0.06	-0.03	0.06	1.00						
11	Keyword Count	0.11	0.11	0.10	-0.05	1.00					
12	Keyword Novelty	0.09	0.10	0.14	0.04	0.06	1.00				
13	Comb. novelty	0.21	0.08	0.18	0.09	0.21	0.27	1.00			
14	ln(simultaneous drafts)	0.25	0.19	0.62	0.03	0.13	0.09	0.10	1.00		
15	Co-Authors	0.07	0.02	0.36	0.25	0.06	0.05	0.08	0.42	1.00	
16	WG submission	0.23	0.01	0.17	0.10	0.15	0.03	0.22	0.02	0.02	1.00

Table 2 Project Publication, Versions and Duration

	(1) Published Logit	(2a) Versions Poisson	(2b) Versions Poisson	(3a) Duration OLS	(3b) Duration OLS
Auth Remote Index	0.331**	-0.003	0.069	6.836	25.024
	(2.75)	(-0.09)	(1.56)	(0.44)	(1.43)
Auth Diff Org Index	0.493**	0.024	0.088*	13.544	32.853*
	(4.12)	(0.66)	(2.09)	(1.04)	(2.43)
Published		0.657**	0.877**	171.929**	299.497**
		(19.48)	(12.80)	(10.56)	(9.45)
Published X Auth Remote			-0.176**		-77.120*
			(-2.70)		(-2.27)
Published X Auth Diff Org			-0.231**		-143.153**
			(-3.18)		(-3.91)
US author(s)	0.326*	0.013	0.004	-8.686	-11.791
	(2.31)	(0.36)	(0.12)	(-0.72)	(-0.97)
Auth Prior Index	0.152	-0.047	-0.048	-27.232*	-27.758*
	(1.40)	(-1.47)	(-1.50)	(-2.18)	(-2.22)
Org Presence (100s)	0.032	0.035*	0.032*	17.240**	16.356**
	(0.46)	(2.24)	(2.05)	(2.84)	(2.70)
ln(prior pubs)	0.280**	0.051**	0.051**	20.096**	20.150**
	(5.66)	(3.53)	(3.50)	(4.02)	(4.05)
ln(file size)	0.122*	0.123**	0.125**	34.737**	35.523**
	(2.27)	(8.01)	(8.14)	(5.01)	(5.13)
Keyword Count	-0.210**	-0.068**	-0.068**	-16.355**	-16.414**
	(-3.77)	(-4.62)	(-4.64)	(-2.87)	(-2.89)
Keyword Novelty	0.002	0.002	0.001	1.386+	1.277+
	(0.34)	(1.04)	(0.79)	(1.83)	(1.68)
Comb. novelty	-0.036	0.019	0.022	3.988	5.120
	(-0.53)	(0.96)	(1.08)	(0.37)	(0.47)
ln(simultaneous drafts)	-0.348**	-0.020	-0.020	-27.584**	-27.724**
	(-6.55)	(-1.32)	(-1.34)	(-4.66)	(-4.71)
WG submission	1.783**	0.562**	0.559**	245.474**	246.544**
	(19.16)	(17.09)	(17.11)	(17.32)	(17.45)
Constant	-1.943**	-0.417+	-0.491*	45.943	22.112
	(-2.64)	(-1.68)	(-1.96)	(0.35)	(0.17)
Co-author number FEs	Y	Y	Y	Y	Y
Time Period FEs	Y	Y	Y	Y	Y
Observations	5250	5250	5250	5250	5250
R^2				0.205	0.210
Pseudo R^2	0.243	0.171	0.173		

t statistics in parentheses
Time Period and Co-author number Fixed Effect coefficients not reported
+ p<0.10, * p<0.05, ** p<0.01

Figure 7 Predictive Margins for Project version and Duration*



*The graphs are based on the predictive margins and 95% confidence intervals for the results presented in table 2.

Figure 8 Internet Draft File Size (kilobytes) by version number and collaboration type

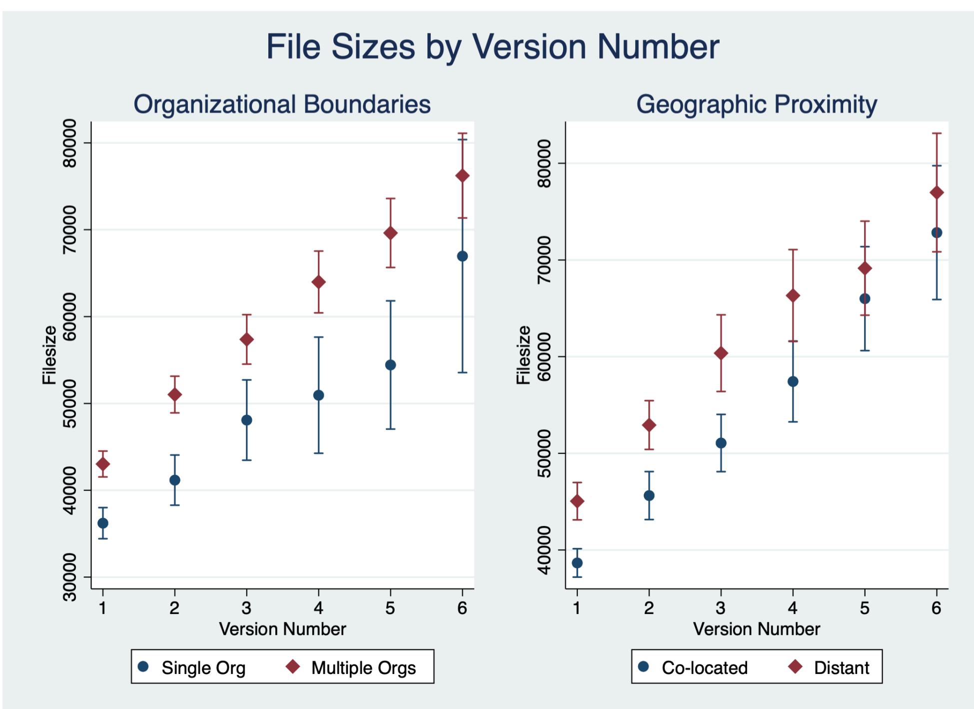


Figure 9 Work Duration (Days) by version number and collaboration type

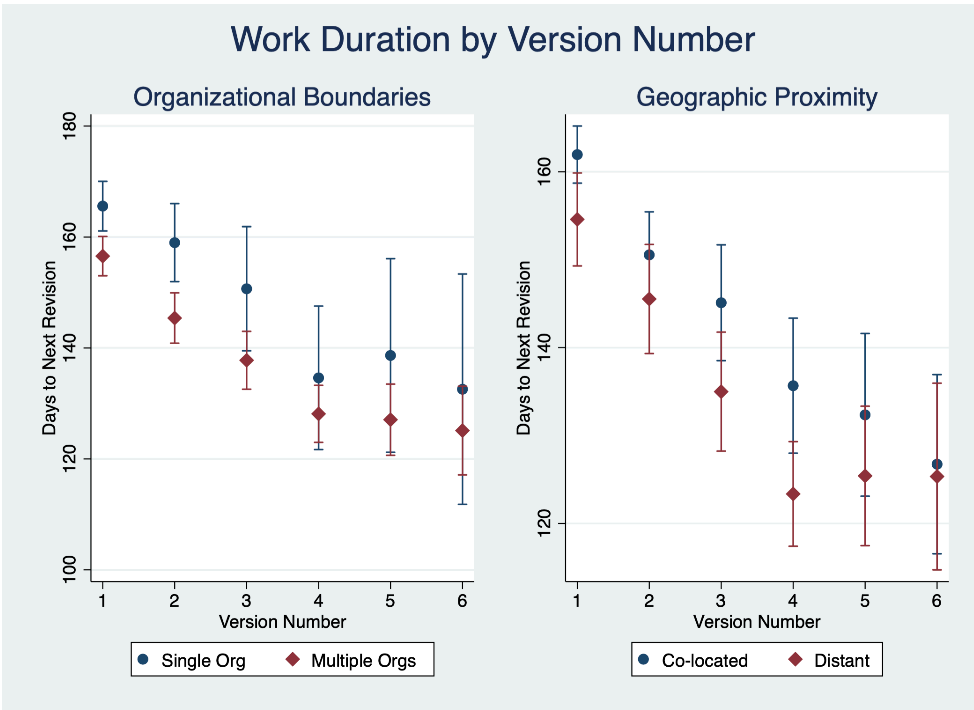


Figure 10 Project Abandonment (Final Version) by version number and collaboration type