How does Division of Labor Affect Team Productivity? Evidence from GitHub

Jinci Liu *

February 27, 2025

Preliminary Draft –please do not circulate or cite without permission of the author

Abstract

Does division of labor increase team productivity? This paper investigates this question using novel data from GitHub, which covers 35 million task allocations across 262,000 software development teams over seven years. It reveals a surprising negative relationship between organizational specialization and team productivity, challenging classical views on the returns to specialization. Although specialized teams have higher code acceptance rates indicating gains in task-specific human capital, they take significantly longer to resolve user-reported issues, suggesting a decrease in general human capital. To establish causality, I exploit the adoption of an automatic task assignment feature on GitHub that evenly distributes tasks among some members. Using a matching design, I find that after this feature adoption, teams become less specialized and show improvements in both output quality (measured by platform users' evaluations) and quantity (measured by lines of code), along with a 10% increase in team communication. These findings highlight a fundamental trade-off: while specialization increases task-specific human capital, excessive specialization impedes knowledge spillovers that are essential for innovation.

*IIES, Stockholm University. Email: jinci.liu@iies.su.se.

I thank David Card, Sahiba Chopra, Mitch Downey,Patrick Kline, Yu Ching Lam, Danielle Li, Zhongyi Liu, Arash Nekoei, David Schönholzer, David Seim, Jösef Sigurdsson, Carolyn Stein, David Strömberg, Hui Sun, Dominik Werh, Horng Chern Wong and seminar participants at IIES Stockholm University and UC Berkeley. I thank Yundi Xiao and Zhehao Zhu for their excellent research assistance. I acknowledge support from the Google Cloud Research Credits program (Credit Name: 317889488 and 216835289) and I thank the Mannerfelt Foundation and Institute for Evaluation of Labour Market and Education Policy (IFAU) for their generous financial support.All errors are my own.

1 Introduction

Significant differences in productivity and organizational structure exist both between and within firms. A central question in economics is how to optimally allocate tasks among workers. The seminal insight from Smith (1776), illustrated through the pin factory example, suggests that the division of labor increases productivity by allowing workers to specialize. Subsequent theoretical work has built on this idea, showing that specialization enhances task-specific human capital accumulation [1].

While extensive literature builds on the theory that specialization improves productivity (Deming, 2017; Bassi et al., 2023), direct empirical evidence on specialization remains limited. One reason is that most datasets lack the detailed information needed to track task allocation and the evolution of specialization. Additionally, measuring specialization and productivity consistently poses a challenge.

This paper addresses these gaps by leveraging a unique dataset from GitHub—the world's largest software collaboration platform—to analyze how the specialization of task allocation affects team productivity. GitHub provides an extensive dataset of 35 million task assignments across 262,000 teams over seven years. This level of detail enables the construction of novel measures of team specialization by comparing actual task allocations to hypothetical benchmarks of "fully specialized" and "fully generalized" structures, taking into account team size and task types. These measures are then linked to various productivity indicators, including output quality, output quantity, problem-solving time, and code acceptance rates.

Software development presents an ideal setting for studying specialization in innovative-intensive work. Unlike routine production, which involves predictable and standardized tasks, software development involves high uncertainty, frequent cross-task coordination, and continuous feedback loops. These conditions create a tension in the returns to specialization: on one hand, focusing on specific tasks allows team members to deepen their expertise; on the other, excessive specialization limits general human capital development and reduces knowledge sharing (Levinthal and March, 1993). While task-specific human capital increases productivity, workers overly focused on a narrow set of tasks may struggle to provide feedback for other members. This trade-off is particularly relevant for creative problem-solving, where effective collaboration often requires integrating knowledge from multiple domains. As Hayek (1945) observed, organizations primarily exist to coordinate and combine diverse information—an essential function in complex, uncertain environments.

To examine the impact of specialization on productivity, I measure within-team specialization and key productivity indicators. Two main findings stand out. First, teams with higher specialization produce lower output quality and quantity: a one-unit increase in the specializa-

[1] For example: Rosen (1983); Baumgardner (1988); Becker and Murphy (1992)

tion index is associated with a 13.8% decline in output quality (measured by peer evaluations from platform users) and a 56.7% decrease in output quantity (measured by lines of code). I have controlled for task type distribution, team age and both team fixed effect and member fixed effect to account for unobserved team-level heterogeneity. Second, while specialized teams experience a 1.82% increase in code acceptance rates—indicating improved task-specific human capital—they take 8.53% longer to resolve user-reported issues, highlighting a trade-off between specialization and general human capital.

The negative relationship between specialization and productivity within team may not solely stem from specialization itself, as task allocation is often endogenous. Managers may adopt specialized structures in response to poor performance. To address this concern, I exploit GitHub's auto-assignment feature, which evenly distributes tasks among some team members. Using a matched difference-in-differences design, I compare teams that adopt automatic task assignment (treatment) to those that do not (control), ensuring both groups share similar characteristics. This matching approach mitigates selection bias and pre-existing differences in specialization. I find that automatic task assignments lead to a 2% decrease in specialization, a 15% increase in output quality, a 20% increase in output quantity, a 2.4 % decline in the code acceptance rate (from a baseline of 0.83), and a 10% increase in team communication.

This paper makes three key contributions to our understanding of how organizational structure influences productivity. First, this paper provides the first large-scale empirical evidence on specialization patterns in software development. While previous studies often focus on controlled experiments, such as Gong and Png (2024), which find that cashiers specializing in scanning tasks perform faster than those handling multiple functions, these findings primarily apply to routine work. In contrast, this paper investigates specialization in software development—an innovative sector where the trade-offs of specialization are more nuanced. By analyzing detailed task-level data, this study offers a comprehensive examination of how specialization evolves, its impact on productivity, and its broader implications for the future of work.

Second, this paper provides empirical evidence for fundamental theories of the division of labor. While classical research emphasizes the productivity gains from specialization (Rosen, 1983; Baumgardner, 1988; Becker and Murphy, 1992), empirical validation has long been hampered by measurement challenges and limited data. To overcome these constraints, I exploit a platform feature designed to evenly assign tasks among team members and apply a matched difference-in-differences approach to evaluate how specialization influences team productivity.

Third, this paper identifies a new cost of specialization—reduced cross-task human capital—which has critical implications for innovative work. While specialization is often associated with productivity gains, excessive task focus can hinder broader learning, adaptability, and the

ability to integrate knowledge across domains. This trade-off is particularly relevant in creative environments, where collaboration and knowledge spillovers play a crucial role in productivity. While Garicano and Rossi-Hansberg (2006) and Garicano and Hubbard (2009) argue that organizations should structure themselves as knowledge hierarchies with specialized agents, this framework does not fully capture the needs of innovation-driven work. Unlike settings where specialists can function independently, software development requires continuous interaction and feedback across domains. This insight helps explain my finding that higher specialization correlates with slower problem-solving, even as it improves code acceptance rates. As described by Levinthal and March (1993), excessive specialization can lead to organizational myopia, where teams prioritize short-term efficiency at the expense of long-term adaptability.

The remainder of this paper is structured as follows. Section 2 describes the data and sample construction. Section 3 details the measurements. Section 4 presents the stylized facts. Section 5 reports the empirical strategy, and Section 6 concludes.

2 Data

This section describes the data used in this paper and the different sources from which we retrieve and combine data to create the final panel. The first source is data about developer activities from the GitHub Archive, which includes timeline and event log. The second source is public profiles of GitHub users, which include their profile pictures, names, locations, email addresses, and biographies.

2.1 GitHub: World's Largest Web Platform for Software Developers

GitHub is a cloud-based platform for software developers to store, share, and collaborate on code projects. As of April 2024, GitHub reports having over 129 million users and more than 205 public million repositories [2]. Repositories (projects) [3], where source code and related documents are stored, can be owned by users or organizations. As user accounts may be either a person or an automated process (a "bot"), this study will focus on the former. Each project has only one owner, but the owner can invite other users to collaborate by offering them to become members and granting them specific permissions. As Figure 1a shows, some teams operate under organizations like Microsoft and DeepSeek, while others are independent, such as users' personal websites. This analysis focuses on organizational projects, as these are more likely to be driven by a firm's profit motives than individual hobbies.

[2]See https://github.com/search?q=type:user and https://github.com/search?q=is:public, accessed 2024-04-30

[3]In this paper, I use "projects" and "repositories" interchangeably, though I recognize that on GitHub, "projects" might have varied meanings.

[Figure 1 about here.]

To start with GitHub, each user must create a unique login and can provide personal information that will be displayed on their profile page. This allows users to showcase their programming skills and enhance their visibility on the platform, which can be beneficial for their careers. Many developers are motivated to share their real information on GitHub; some even provide additional information, such as links to their LinkedIn profiles and personal websites.

GitHub allows developers to save snapshots of files, thereby effectively recording the history of changes and facilitating collaborative development. Furthermore, GitHub activities are available from GHArchive, which records the public GitHub timeline and event log.

2.2 Data Construction

To construct my dataset, I combine the GitHub Archive database with developer profile information via the GitHub API. The GHArchive provides time-variant information on the events for all public projects, with a comprehensive record of the GitHub timeline since February 2011. For instance, the system records events related to modifying and suggesting code files, such as pull request. It also provides the number of lines of code contributed by each user, the time required for a pull request to be opened and closed, and the number of stars each project received from users of the universe platform.

To gather demographic information, I use the GitHub API to access each developer's profile page, which includes their profile image, location, name, email, and biography. Career factors significantly motivate developers to contribute to online platforms such as GitHub and Stack Exchange (El-Komboz and Goldbeck, 2024; Forderer and Burtch, 2024). This suggests that developers have an incentive to disclose accurate personal information that may enhance their career prospects. I focus on developers who are affiliated with repositories and have permission to access repository files, defining these individuals as team members. For these developers, GitHub functions more as a workspace than a hobby space, providing a greater incentive to disclose accurate personal information that may benefit their careers.

2.3 GitHub Team Members

A team is defined as a project-based entity. Ideally, I prefer to focus on full-time employees within each team. However, GitHub lacks a specific indicator for such categorization. Consequently, I rely on another category related to author association: collaborator, contributor, mannequin, member, none, and owner [4]. This category has been available since 2017 and is tied to access permissions.

[4]Detailed information from GitHub Docs: author association, accessed 2024-04-26

A team member is defined as anyone classified as a member, owner, or collaborator in the author association[5]. Since author association information is only recorded when a developer participates in specific events, months without participation result in missing author association information. To address this issue of missing values, I document the earliest and latest instances when each developer is recognized as a team member, assuming they were part of the team during the time. It is rare for a member to leave and rejoin a team midway through the observation period.

2.4 Mapping Code Files to Tasks

I link every code file with team members who have worked on it, whether by writing, providing feedback, or evaluating its quality. I do not focus on the code's original authors, as receiving feedback and requesting reviews from other members is also crucial. Such collaborative efforts reduce the bugs and conflicts when merging code files into the main project.

The data contains 34,762,813 code filenames from 2017 to 2023. I use the Latent Dirichlet Allocation (LDA) topic modeling method to categorize these files into 10 task categories (Blei et al. (2003)). Figure 2a illustrates an example of code file names used as input for LDA. Having too many categories can lead to sparse observations, which leads to measurement problems. Conversely, too few categories will obscure important variations, making it difficult to capture differences in task composition across teams. Therefore, I choose 10 task categories for the whole analysis. Table 1 presents a list of the 10 task groups and the results of the LDA model keywords and 2b shows a member task snapshot in the real dataset. The assumption is that code files within the same category have similar skill requirements and lower coordination costs. The appendix section A provides additional details about the process and examples.

[Table 1 about here.]

[Figure 2 about here.]

[Table 2 about here.]

2.5 Project Type

Given that each team may work on different projects, I collect information from the descriptions and labels of each team. For example, the Visual Studio Code team has the description "Visual Studio Code" and labels including "electron," "microsoft," "editor," "typescript," and "visual-studio-code". Using LDA, I categorized the team into 6 distinct project types. The appendix section B provides additional details about the process and word cloud.

[5]I don't use MemberEvent because not every user becomes a member through the MemberEvent. Additionally, some users added via MemberEvent do not work after their addition.

2.6 Gender Prediction

To predict the gender of developers based on their avatars, I select a state-of-the-art model that closely matches the resolution and cropping methods of the training images. This chosen model[6] trains vision transformers on male and female portraits and develops a corresponding image classifier. The model predicts the probability that an image represents a female or a male. To find the optimal cutoff for gender classification, I use a sample labeled by some annotators and find that the highest accuracy is achieved with a cutoff of 0.85. If the predicted probability is above 0.85, the image will be classified as male or female; otherwise, it will be labeled as "others".

2.7 Race Prediction

I predict developers race and ethnicity based on their name and I choose Ethnicolr, as developed by Laohaprapanon et al. (2017). Ethnicolr uses both first and last names to classify individuals into four combined race and ethnicity categories: Hispanic (of any race), and non-Hispanic Asian, Black, and White. The algorithm is trained on data from the US Census, Florida voter registration, and Wikipedia. Each name is assigned probabilities for belonging to each of the four categories, with the category of highest probability being designated as the predicted race or ethnicity.

3 Measuring Task Assignment and Team Productivity

This section outlines the measurement approach for task specialization and team productivity. I introduce a novel specialization metric that quantifies how a team's task allocation deviates from theoretical benchmarks based on team size and task distribution. Specifically, I compare each team's actual task assignment to two counterfactual benchmarks: a generalized benchmark, where tasks are evenly distributed among members, and a specialized benchmark, where each member focuses on a single task[7]. This relative metric enables comparison of specialization across teams with varying sizes and project types.

For team productivity, I develop four distinct metrics that capture various dimensions of performance in software development. These metrics address both the quantity and quality of team outputs, making it easier to distinguish between the effects of specialization on task-specific human capital and general problem-solving capability. While these measures are designed for software teams, they can be applied to other contexts.

[6]https://huggingface.co/rizvandwiki/gender-classification, accessed 2024-04-23

[7]Some teams could not assign one task to one member due to tasks requiring multiple units of labor

3.1 Measure Team Specialization

Denote $\mathbb{K}$ the set of all task types. Each team $m = 1, \ldots, M$ consists of $N_m \in \mathbb{N}$ members and handles a set of task $K_m \subseteq \mathbb{K}$. Denote $A_m \in \mathbb{R}_+^{N \times |K_m|}$ is the task assignment for team m, where $A_m(i,j)$ represents worker i's labor input on task j.

$$A_m = \begin{pmatrix} A_m(1,1) & A_m(1,2) & \cdots & A_m(1,|K_m|) \\ A_m(2,1) & A_m(2,2) & \cdots & A_m(2,|K_m|) \\ \vdots & \vdots & \ddots & \vdots \\ A_m(N_m,1) & A_m(N_m,2) & \cdots & A_m(N_m,|K_m|) \end{pmatrix}$$

Denote $l_m(i) := \sum_j A_m(i,j)$ to be worker i's labor share and $\alpha_m(j) := \sum_i A_m(i,j)$ to be the task share of task j. Assume each member has one unit of inelastic labor supply.

Definition 1 *Define G_m to be the most generalized task allocation matrix of team m, where the task assignment of member i to task j is $G_m(i,j) = \frac{\alpha_m(j)}{N_m}$.*

Definition 2 *Define S_m to be the most specialized task allocation matrix of team m.*

Definition 3 *Define SPE_m to be the specialization index where $SPE_m = \frac{d(A_m, G_m)}{d(S_m, G_m)}$*

Consider three task assignments in Table 3. There are three members (A, B, C) and three task types (1, 2, 3), with task shares $\alpha_m = \begin{bmatrix} 3/2 & 1/2 & 1 \end{bmatrix}$. Suppose the actual assignment(A_m) shows how each member allocates their labor supply across tasks. The most generalized (least specialized) assignment is when all members evenly split their labor across all tasks (G_m) as defined in the Definition 1. In contrast, the most specialized assignment aims to assign each member to a single task, with any residual workload allocated to another member due to unequal task shares (S_m). Since members are homogeneous, it doesn't matter who handles the leftover tasks. One interpretation of the member handling the leftover tasks (e.g., member B) could act as a manager overseeing various tasks. To measure the degree of specialization in A_m, we calculate the distance between A_m and G_m, normalized by the distance between S_m and G_m as defined in Definition 3. The specialization index $\text{SPE}_m \in [0,1]$ and is higher if A_m is farther from the counterfactual generalized allocation G_m.

[Table 3 about here.]

There are many ways to measure the distance between two matrices. In this paper, I use Euclidean distance[8] and

$$d(A_m - G_m) = \sqrt{(A_m(i,j) - G_m(i,j))^2}$$

[8]I only focus on teams with more than 2 members and 2 tasks. Otherwise, the team does not require allocation decisions, as the solo member must perform all tasks, or all members can only complete one task. Mathematically, when $|K_m|$=1 or N_m=1,SPE_m is not well-defined

$$d(S_m - G_m) = \sqrt{\sum_j \underbrace{\lfloor \alpha_m(j) \rfloor}_{\text{\# of 1s}} \cdot \left(1 - \frac{\alpha_m(j)}{N_m}\right)^2} \tag{1}$$

$$+ \underbrace{(N_m - \lceil \alpha_m(j) \rceil)}_{\text{\# of 0s}} \cdot \left(\frac{\alpha_m(j)}{N_m}\right)^2 \tag{2}$$

$$+ (\lceil \alpha_m(j) \rceil - \lfloor \alpha_m(j) \rfloor) \cdot \left(\underbrace{\alpha_m(j) - \lfloor \alpha_m(j) \rfloor}_{\text{residual workload}} - \frac{\alpha_m(j)}{N_m}\right)^2 \tag{3}$$

To apply the specialization index SPE_m to empirical data, each code file is categorized into one of the ten task types outlined in Table 1. Using this classification, I compute the fraction of code files associated with each task type that each team member handles on a monthly basis, which forms the matrix $A_m(i, j)$. [9] For example, consider a team with two tasks: Front-end Development and Back-end Development. If a member exclusively works on files related to Front-end Development, her task assignment would be represented as $A_m(i) = \begin{bmatrix} 1 & 0 \end{bmatrix}$.

Figure 3a presents the distribution of the Team Specialization Index for teams with more than three members, illustrating both the overall distribution and its variation by team size. Many teams exhibit highly specialized structures (index values near the upper bound), while others opt for no specialization (index value of 0). The majority of teams concentrate around the mid-range values (0.4–0.7), indicating a balanced degree of specialization. Figure 3b further decomposes the specialization distribution by team size. Each subplot represents a different team size, ranging from 4 to 15 members. While the overall distribution remains similar across team sizes, larger teams tend to adopt more specialized structures. Figure C1 provides additional details on other dimensions.

[Figure 3 about here.]

3.2 Measure Team Productivity

Measuring productivity is inherently challenging. I develop 4 different measures that capture different dimensions of team-level productivity, drawing from the existing literature and the context of software engineering. These measures include output quality and quantity, submission acceptance rate, and problem-solving speed.

[9] An alternative approach would be to measure task assignment based on the fraction of time each member allocates to different tasks. However, I rely on the number of code files due to the challenge of precisely observing the time spent on each file.

Output Quality. - GitHub users can give a project a star to express their interest and appreciation. Each user can only give one star to a project, and this action is not anonymous. GitHub uses these stars to recommend content to users on their dashboards, and they also serve as an indicator of community satisfaction. Studies show that 75% of developers check a project's star count before deciding to use it (Borges and Valente, 2018). Additionally, stars have a monetary value, with each star estimated to be worth approximately 0.88 USD (Eldeeb and Sikora, 2023). To measure a team output quality, I track the number of stars it receives each month. Figure D1 provides an example of the cumulative stars received by a team over time.

Output Quantity. - I measure output quantity using the number of lines of code produced by team members each month. The relationship between quantity and productivity is difficult to determine, as coding techniques like loops can reduce the number of lines without affecting functionality. Despite the limitation, quantity metrics still offer valuable insights into the volume of code generated, with the number of lines serving as a general indicator of a team's overall output (Vasilescu et al., 2015; Casalnuovo et al., 2015; Wagner and Ruhe, 2018).

Submission Acceptance Rate. - This metric evaluates the success rate of code submissions through pull request (PR), similar to measuring product passing rate in manufacturing. It is calculated as the ratio of successfully merged PRs to the total number of PR submitted in a given month. Developers create branches to isolate development work without affecting others and then use PRs to merge those branches once complete. PR can be rejected when conflicts arise, such as when both the PRs and target branches modify the same part of a file differently, GitHub is unable to automatically decide which changes to keep. Additionally, PRs can fail if developed do not adhere to branch protection rules or if they do not pass required checks. Reviewers may also reject PRs due to issues with code quality, design, or functionality issues. This requires developers to address feedback and update their PRs accordingly. A low PR success rate indicates teams frequently encounter conflicts or require significant rework before merging code. Figure D5 provides examples of the success and failure of PRs.

Problem-Solving Speed. - This metric captures the duration between the creation of an issue on GitHub and its initial closure. Users create issues to raise questions or report problems, which are then addressed and closed by team members. Once resolved, the issue is closed, indicating successful problem resolution. This metric is constructed by tracking issues opened within a given month and recording the time to their first closure. It serves as an indicator of the team's efficiency in handling issues. Figure D6 provides an example of one issue from opening to closing.

Descriptive statistics are provided in Table 4.

[Table 4 about here.]

4 Stylized Facts

This section presents stylized facts about team specialization and productivity in software development:

Fact 1 *Higher team specialization is associated with lower output quality.*

Figure 4 displays a binned scatter plot illustrating the relationship between organizational specialization and output quality, measured by the log of monthly stars received from platform users. To mitigate omitted variable bias, I include several controls and fixed effects. First, I control for task share distribution to account for variations in project complexity and scope. Second, I control for team age, as more established teams may systematically differ in both specialization and output quality. Third, I include fixed effects for teams, members, and team size to account for time-invariant heterogeneity in team composition, member characteristics, and project attributes. These controls help mitigate potential confounding factors, such as differences in team structure, project characteristics, and lifecycle effects.

After accounting for these factors, the results indicate a negative relationship between team specialization and project quality: as the team structure becomes more specialized, project quality declines.

[Figure 4 about here.]

Table 5 presents the results from OLS and Poisson pseudo-maximum likelihood (PPML) regressions. In the OLS models, where the dependent variable is the log of monthly stars, team specialization (SPE) is negatively associated with output quality across all specifications. Column (1) shows that a one-unit increase in the specialization index is associated with a 32.1% decrease, implying a substantial reduction in output quality. The magnitude of this negative effect decreases to 8.7% but remains statistically significant in Column (4) when controlling for team and member fixed effects. Team age is positively associated with the output quality in the simpler specifications, but the coefficient becomes negative when additional controls are introduced. This suggests that while older teams may initially benefit from accumulated experience and recognition, they may also experience diminishing quality over time, potentially due to reduced adaptability or shifts in user interest.

In the Poisson models, where the dependent variable is the count of monthly stars, the negative relationship between specialization and quality is even clearer. The coefficient in

Column (1) indicates that a one-unit increase in the specialization index is associated with a 35% decrease in the expected value of monthly stars. When team and member fixed effects are included, the magnitude of the effect decreases but remains statistically significant: a one-unit increase in specialization is still associated with a 13.8% decrease in output quality.

[Table 5 about here.]

Fact 2 *Higher team specialization corresponds to lower output quantity.*

Figure 5 illustrates the relationship between team specialization and the number of lines of code produced by team members each month. The findings align with the specification in Figure 4, revealing that more specialized teams tend to produce less coding output.

[Figure 5 about here.]

Table 6 reports the regression results. In the OLS models, where the dependent variable is the log lines of code, specialization (SPE) shows a negative association with output quantity across all specifications. Column (4) indicates that a one-unit increase in the specialization index corresponds to a 94.2% decrease in lines of code. Similarly, in the Poisson models, which measure the count of lines of code, Column (4) shows that a one-unit increase in the specialization index results in a 56.7% decline in the expected value of lines of code.

[Table 6 about here.]

Fact 3 *Specialized teams achieve higher acceptance rates but require more time to resolve user problems.*

Figures 6a and 6b illustrate the trade-off between task-specific and general human capital. Unlike Figure 4, these analyses additionally include fixed effects for the number of code submissions and the number of questions, respectively. This ensures that the observed relationships account for variations in coding activity and inquiry frequency. Figure 6a shows that code acceptance rates increase with increased team specialization, suggesting that developers who focus on specific tasks produce higher-quality code. However, Figure 6b reveals that specialized teams take longer to resolve platform users' problems compared to less specialized teams. This delay likely arises because effective problem-solving often requires integrating knowledge across multiple domains, such as code architecture, user interface design, and database management.

Additionally, Figure 6c examines the relationship between team specialization and the number of comments. The results indicate that highly specialized teams tend to have fewer comments and less communication, suggesting lower levels of knowledge exchange within the team. This finding supports the hypothesis that while specialization increases task-specific human capital, it also leads to a decrease in general human capital and reduced knowledge diffusion.

[Table 7 about here.]

[Table 8 about here.]

Table 7 and Table 8 present the regression results. After controlling for team task distribution and team age, and adding team size fixed effects, team fixed effects, and member fixed effects, a one-unit increase in the team specialization index is associated with a 1.82% increase in the code acceptance rate and a 8.53% increase in the time required to solve a problem. These findings suggest that while specialization may improve certain aspects of code quality, it can compromise problem - solving efficiency.

[Figure 6 about here.]

Fact 4 *Team members have similar demographic compositions regardless of team specialization.*

To measure diversity, I follow the literature on ecological diversity and use the Shannon - Wiener Index to quantify team composition diversity. For example, the Gender Diversity Index is calculated as:

$$H' = -\sum_{i=1}^{S}(p_i \ln p_i)$$

where H' represents gender diversity index, S is the total number of different categories (e.g., male and female), and p_i is the proportion of individuals in the i-th gender category. This index measures both the abundance and evenness of gender within the team. A higher gender diversity index indicates greater diversity, meaning that the team has a more balanced gender distribution (Krebs, 1989).

Figure 7 illustrates the relationship between organizational specialization and diversity across two dimensions: gender and race. There is little difference in gender and race diversity among teams with different levels of specialization. Teams with high specialization and those with low specialization tend to have a similar demographic composition of people. This result suggests teams coordinate members from similar backgrounds which could avoid communication and coordination costs generated by diversity (Lazear, 1999; Prat, 2002; Hong and Page, 2004; Kahane et al., 2013).

[Figure 7 about here.]

5 Empirical Strategies

The descriptive evidence shows a negative correlation between specialization and team productivity. The choice of task allocation structure is often endogenously determined. Managers

might adopt specialized structures in response to poor performance or workforce disruptions. To mitigate these concerns, I exploit GitHub's auto-assignment feature, which evenly distributes tasks among team members. Using a matched difference-in-differences design, I compare teams that adopt auto-assignment (treatment) with those that do not (control) but share similar characteristics. This matching approach mitigates selection bias and pre-existing differences in specialization, allowing for a more credible assessment of how specialization affects team productivity.

5.1 GitHub Automatic Assignment Feature

I exploit GitHub's auto-assignment feature, which introduces exogenous variation in team specialization by evenly distributing tasks. This system automatically assigns team members to Pull Requests (code submissions requiring review) and Issues (bug reports or feature requests). Unlike traditional workflows where certain members specialize in specific tasks, auto-assignment randomly selects reviewers or assignees. Teams enable this feature via configuration files[10], ensuring a balanced workload and reducing biases in task allocation. By making all members equally likely to engage in different tasks, this mechanism shifts teams toward a more generalized structure. Figure D7 provides an example of such a configuration file.

This setting offers a unique opportunity to study how decreased specialization affects team performance, as adopting teams effectively shifts toward a more generalized task allocation structure. To identify the causal effect, I employ a matched difference-in-differences design comparing teams that adopt automatic assignment (treatment) with those that do not (control). Following the approach from the first identification strategy, I implement a 1:1 matching procedure based on pre-treatment characteristics measured from $t-5$ to $t-1$: deciles of average team size, task types, and output levels.

I successfully matched 3,864 teams in the treatment and control groups, achieving a 98.2% success rate, which confirms that treated teams have closely comparable counterparts. Table 9 presents the summary statistics of both groups at the time of adopting auto assignment. The treatment and control teams are closely aligned in terms of team size, task types, lines of code, and specialization index, consistent with the matching procedure. According to the table, at the time of adoption, treatment teams had an average team size of 7.86, compared to 6.93 for control teams, and engaged in 5.84 task types, slightly higher than the 4.79 task types in the control group. Despite not being directly matched on all characteristics, both groups remain balanced in key productivity measures. The specialization index (0.53–0.56) and code acceptance rate (0.83) are nearly identical across groups. Overall, the table highlights that treatment and control groups are well-balanced across core characteristics related to team composition, specialization,

[10]For example, .github/workflows/ or .github/auto_assign.yml

and performance. This strengthens the credibility of the matching process, providing a robust empirical foundation for assessing the impact of even task assignment adoption.

[Table 9 about here.]

5.2 Dynamic Effects

Building on the identification strategy in Jaravel et al. (2018); Sharoni (2024), I use OLS regressions with a full set of leads and lags around the adoption of even task assignments. This approach allows me to examine how adoption affects team specialization and productivity while also validating the research design by testing for pre-existing trends before adoption, addressing concerns about potential common shocks. Using the constructed treatment and control groups, I align adoption events so that the time index k represents time relative to adoption, with $k = 0$ marking the adoption event. This stacking method follows Roth (2022) to ensure a robust estimation of pre-trends. I then estimate equation (4),

$$Y_{mk} = \gamma_m + \beta_k + \psi_{j(m)} + \sum_{k=-5, k\neq -1}^{6} \alpha_k \times \mathbb{I}\left(\text{period}_k\right) \times \text{Treated}_m + \epsilon_{m,k} \tag{4}$$

where Y_{mk} denotes the outcome for team m in month $k = t - d$ (e.g., monthly stars received from platform users, code acceptance rate, problem-solving time) in the treatment group and are normalized to zero in $k = -1$. The specification includes team size fixed effects (γ_m), time fixed effects (β_k), and firm fixed effects ($\psi_{j(m)}$). Treated_m indicates teams that adopt auto-assignment. I cluster the standard errors at the matching level.

If the adoption of even task assignment serves as an exogenous variation, the coefficients α_k estimate its causal effect on even task assignment and productivity, measured k periods relative to adoption. A positive α_k suggests that even task assignment improves productivity, while a negative coefficient indicates a decline. If there is no effect, the coefficient would be zero. However, a key identification concern is selection bias—if teams that adopt even task assignment systematically differ from those that do not, the estimated effect may not reflect true causality. To mitigate this issue, I construct a matched control group of teams with similar pre-treatment characteristics but no adoption. In the analysis below, I demonstrate that pre-trends are not statistically significant, addressing concerns about potential biases. Specifically, the lag coefficients α_{-5} to α_{-1} show no significant differences before adoption, suggesting that prior trends in specialization and productivity are not driving the observed effects. This finding supports the validity of the estimates and reinforces the causal interpretation of the results.

Figure 8 presents estimates from regression equation (4) that show no significant pre-trends. As shown in Figure 8a, adopting auto task assignment reduces team specialization by about

2%, as team members—especially reviewers and assignees—are randomly allocated to tasks, resulting in a more generalized task allocation structure.

The impact on team performance varies across different outcomes. Figure 8b shows that, relative to the control group, teams with reduced specialization experience a 15% statistically significant increase in output quality —measured by monthly stars—approximately three months after adoption. In terms of output quantity as shown in Figure8c, teams adopting auto-assignment increase output production by 20% (measured by the number of lines of code written), as measured by the number of lines of code written. However, despite this increase in coding activity, the code acceptance rate declines immediately by roughly 2.4% with a baseline of 0.83, suggesting the loss of task-specific human capital. There is no significant effect on problem-solving speed(e.g., time required to resolve issues). Finally, Figure 8f indicates a 10% increase in team communication following adoption, providing evidence that reduced specialization promotes greater knowledge sharing and cross-task feedback.

[Figure 8 about here.]

6 Conclusion

In this paper, I investigate how specialization in team task allocation affects productivity and knowledge sharing in software development. Although the traditional view, dating back to Smith (1776), suggests that specialization increases productivity, my findings reveal a negative relationship between team specialization and productivity. They highlight that excessive specialization can inhibit the general human capital necessary for innovative work where the next steps are often uncertain and cross-task feedback is essential.

To address potential endogeneity in task allocation, I use GitHub's automatic task assignment feature, which distributes tasks evenly among team members and reduces specialization. A matched difference-in-differences analysis shows that when teams become less specialized, they experience productivity gains.

Another significant contribution of this study is the development of new empirical indicators for both specialization and productivity. I create a novel dataset measuring team task allocations based on the code developers produce, alongside various productivity metrics. My specialization measure captures deviations from fully generalized or fully specialized benchmarks, allowing for comparisons within teams over time. For productivity, I establish metrics that encompass both output quantity (lines of code) and quality (project popularity), along with finer indicators such as code acceptance rates and problem-solving speed. Together, these metrics provide a comprehensive view of how specialization affects innovation processes.

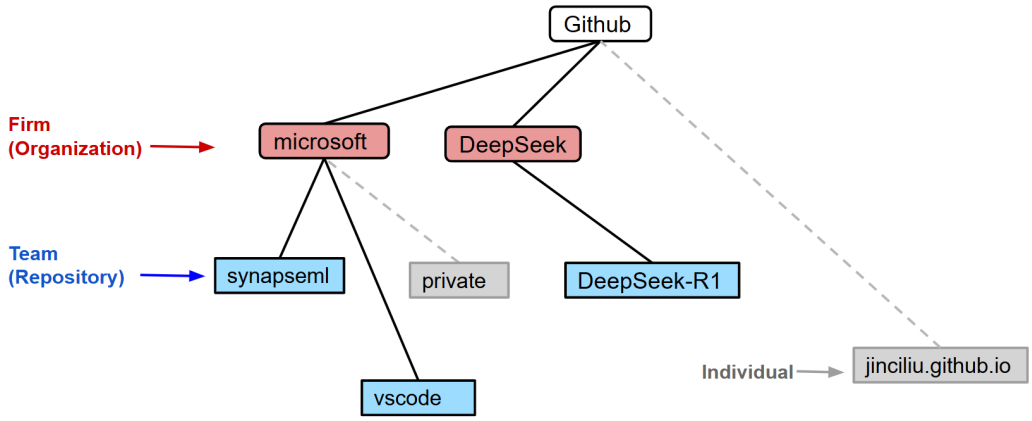
These findings enhance our understanding of organizational design in innovative industries.

While previous research has emphasized the benefits of specialization through task-specific human capital accumulation, I illustrate an important trade-off: excessive specialization can hinder the development of general human capital essential for complex problem-solving and knowledge integration. This distinction is especially relevant as economies increasingly shift toward knowledge work, where solutions often require integration across multiple domains.

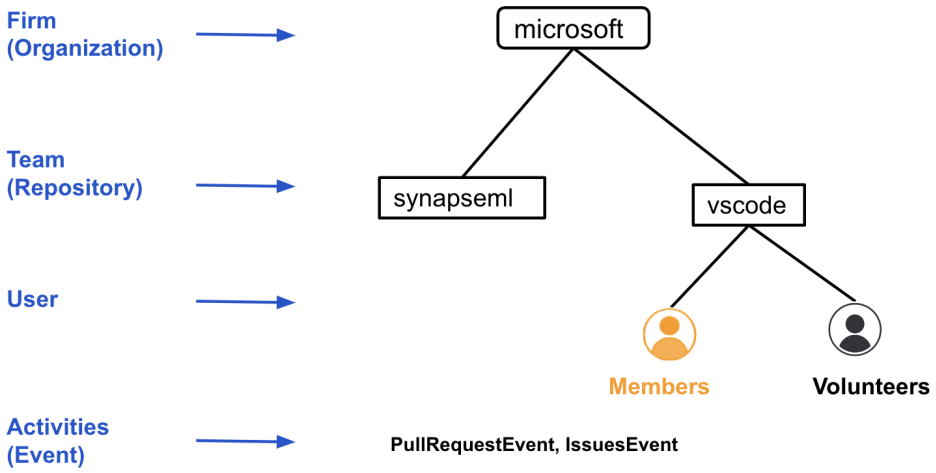
Moreover, the results contribute to the broader literature on firm organization and productivity variation. The persistent differences in organizational structures among firms may reflect not only coordination costs or information frictions but also different optimal positions along the specialization spectrum, depending on the nature of production processes. Future research could explore how these trade-offs vary across industries and types of knowledge work, potentially using the measurement framework developed in this paper.

Figure 1: Schematic Representation of GitHub Structure

(a) Organization and Team



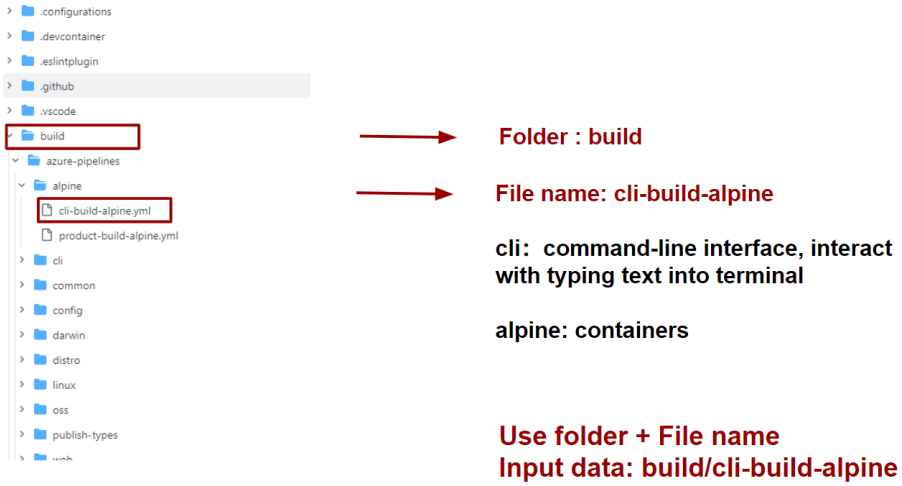
(b) Team members and Volunteers



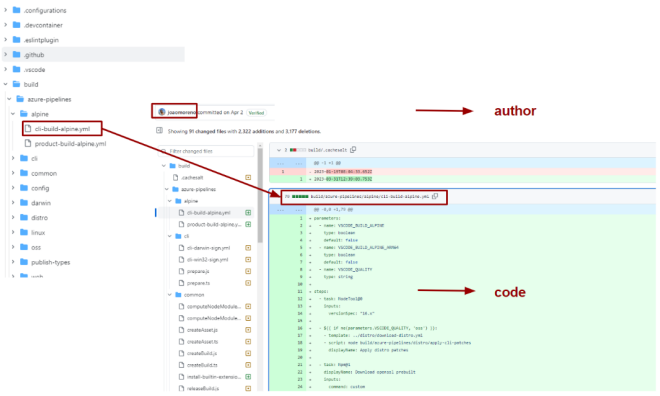
Notes: As shown in Figure 1a, some repositories belong to the organization, while others do not. In this project, I include only repositories affiliated with the organization to exclude those intended for individual purposes or hobbies, such as personal websites. Figure 1b illustrates the structure within a single organization. For instance, within Microsoft, there are multiple teams, each with various users. Some users are official team members (see Section 2.3), while others are volunteers who also participate and contribute. This project focuses solely on the activities of team members.

Figure 2: Code File and Task-Member Snapshot

(a) Code File Example



(b) Task-Member Example

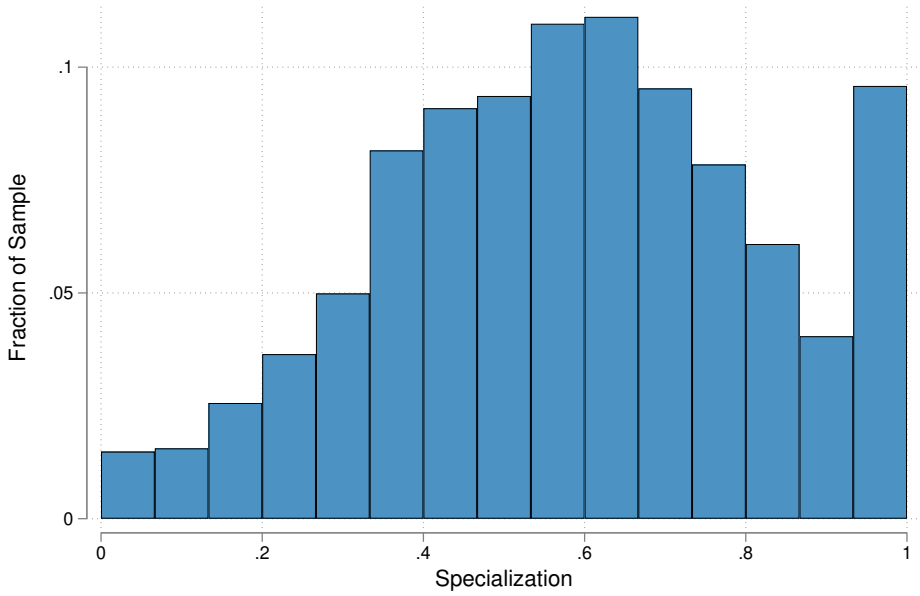


Team	Month	Member	Task Type	Labor Share
vscode	2021-05	joaomoreno	1	0.714286
vscode	2021-05	joaomoreno	4	0.285714
vscode	2021-05	isidorn	4	0.018809
vscode	2021-05	...	...	...

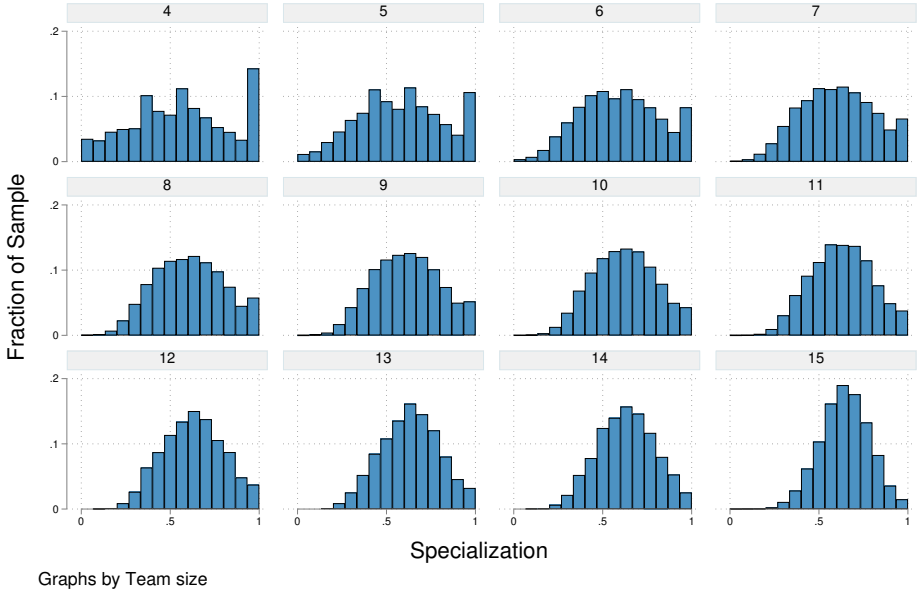
Notes: Figure 2a illustrates an example of code file names used for task type classification. The input data consists of the folder name combined with the file name, such as "build/cli-build-alpine" in the example above. Figure 2b presents a data snapshot of Member "joaomoreno" from Team "vscode" highlighting their labor share in task types 4 and 1.

Figure 3: Distribution of Team Specialization Index

(a) Specialization

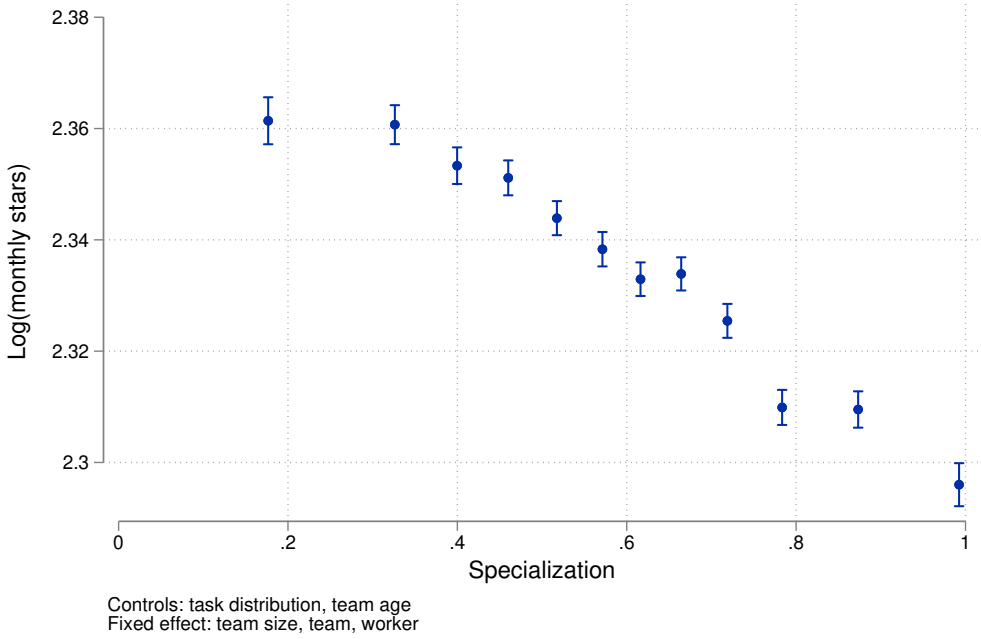


(b) By Team Size



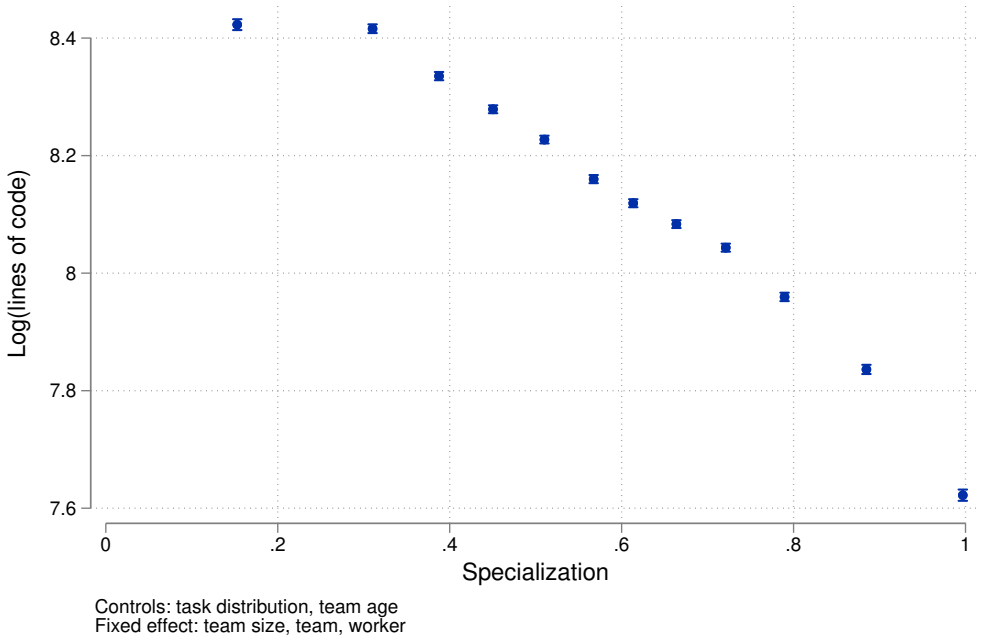
Notes: All team-month observations are included, except for teams with fewer than four workers (team size). A higher index value reflects a greater degree of specialization. FigureC1 provides more details about the distribution of the team specialization index.

Figure 4: Organizational Specialization and Output Quality



Notes: This figure presents a binned scatter plot, accounting for task distribution and team age. The analysis also includes team size, worker fixed effects, and team fixed effects. These fixed effects help control for differences in worker composition, team size, and the impact of team project types on team productivity. Observations are at the individual level and are weighted by 1/team size. The graph highlights a stronger negative relationship between specialization and output quality.

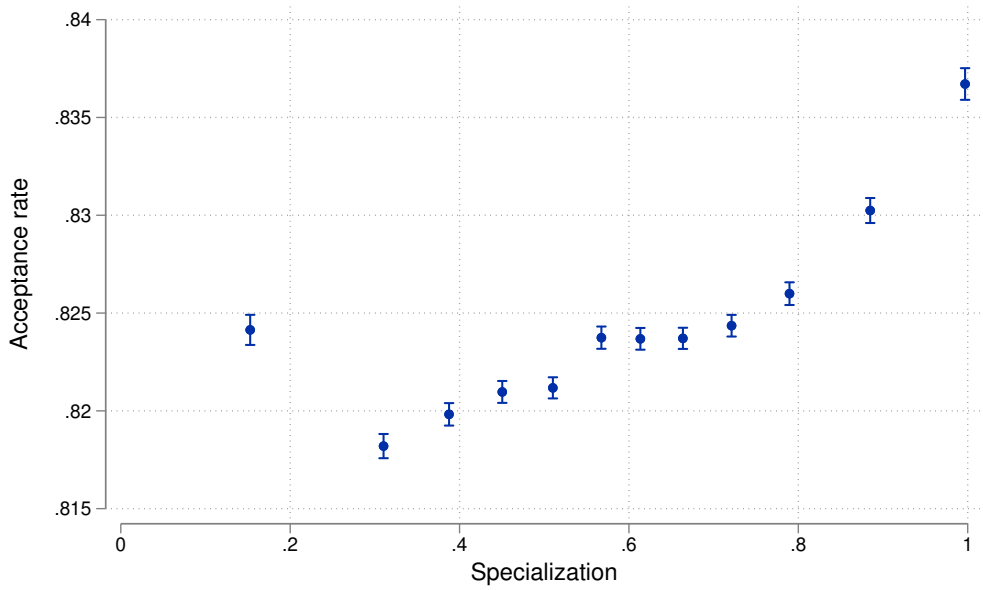
Figure 5: Organizational Specialization and Output Quantity



Notes: This figure presents a binned scatter plot, accounting for task distribution and team age. The analysis also includes team size, worker fixed effects, and team fixed effects. These fixed effects help control for differences in worker composition, team size, and the impact of team project types on team productivity. Observations are at the individual level and are weighted by 1/team size. The graph highlights a stronger negative relationship between specialization and output quantity.

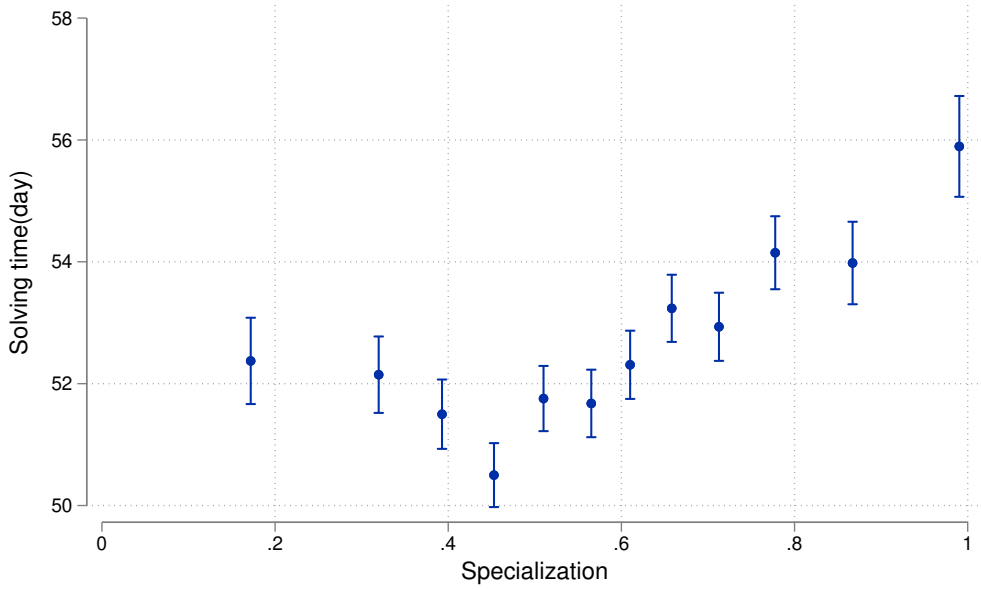
Figure 6: Organizational Specialization Trade-off

(a) Code Acceptance Rate



Controls: task distribution, team age
Fixed effect: team size, #submissions, team, worker

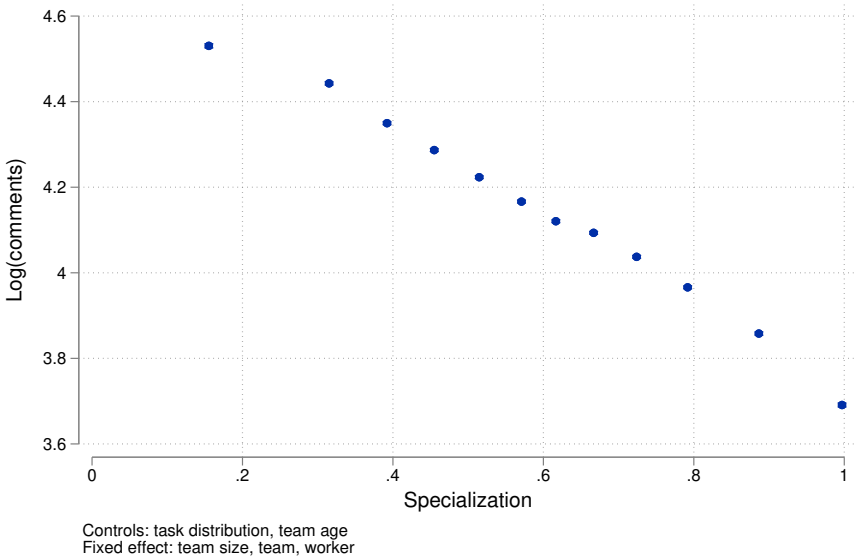
(b) Problem Solving Speed



Controls: task distribution, team age
Fixed effect: team size, team, #problem, worker

Figure 6: Organizational Specialization Trade-off

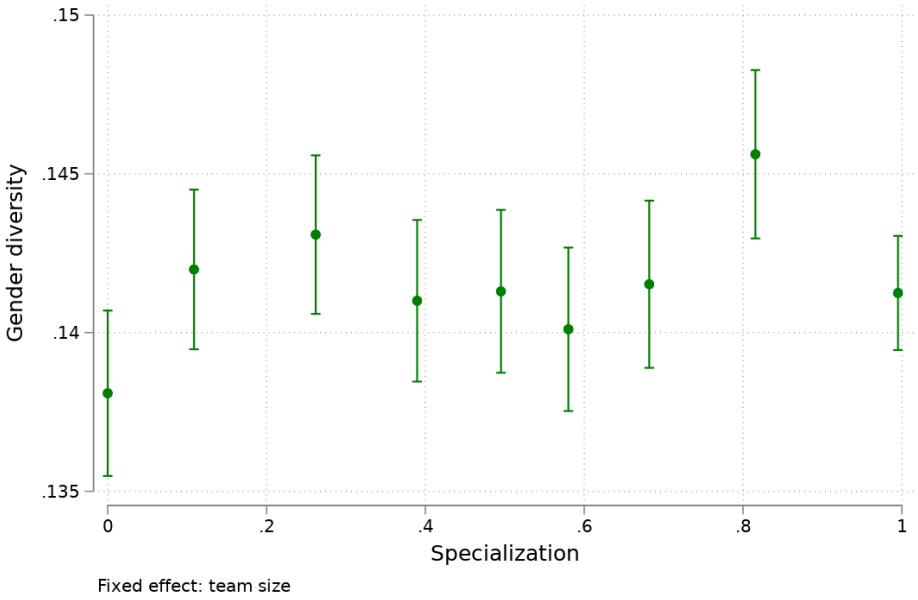
(c) Discussion



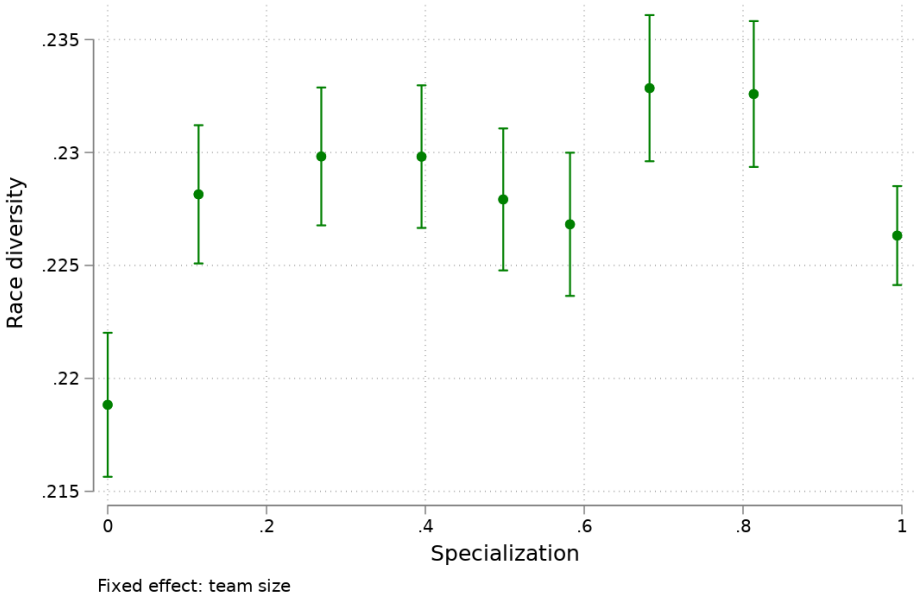
Notes: This figure shows the correlation between team specialization and code acceptance rate (Panel 6a), problem-solving speed (Panel 6b), average comments per question (Panel 6c) after controlling for task share distribution and team age as well as team size, team, member fixed effects. The observations are at the individual level and weighted by 1/team size.

Figure 7: Team specialization and diversified composition

(a) Gender



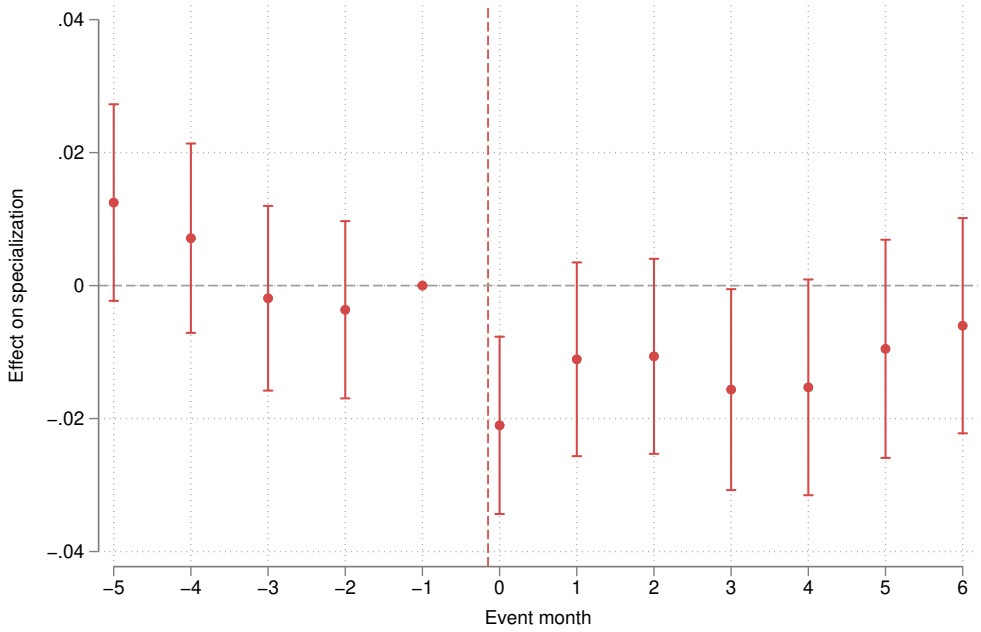
(b) Race



Notes: The positive relationship between organizational specialization and diversity in gender (Panel 7a), and race (Panel 7b) is presented in a binned scatter plot. I calculate each diversity index separately and exclude teams with only one member. Gender is predicted based on developers' profile images and names. Race is predicted using the name. Teams with a single task are also excluded. Section 2 provides more details.

Figure 8: Dynamic Effects for Task Assignment Adoption

(a) Team Specialization Index



(b) Output Quality

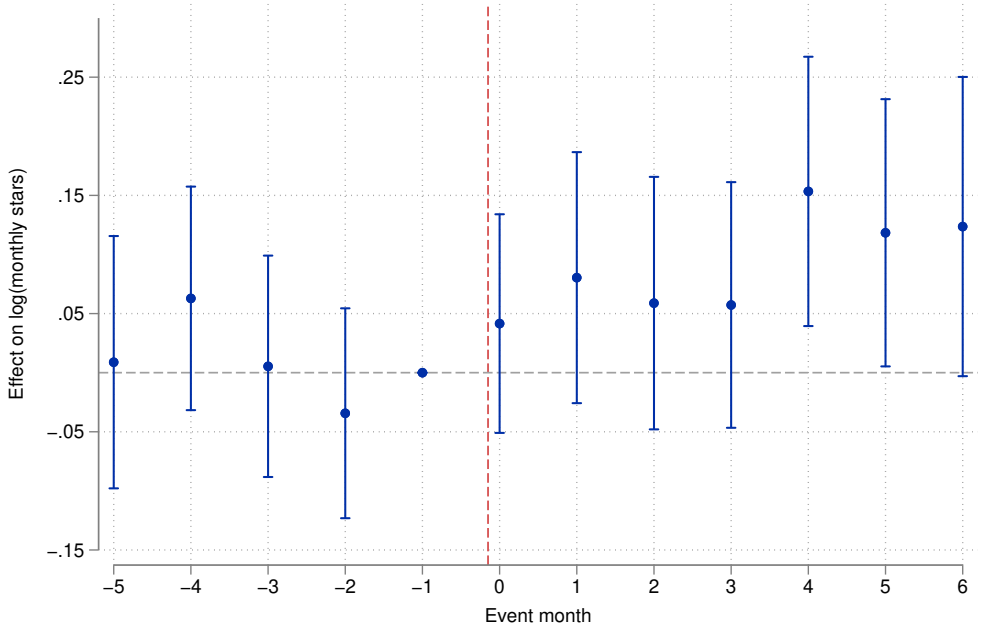
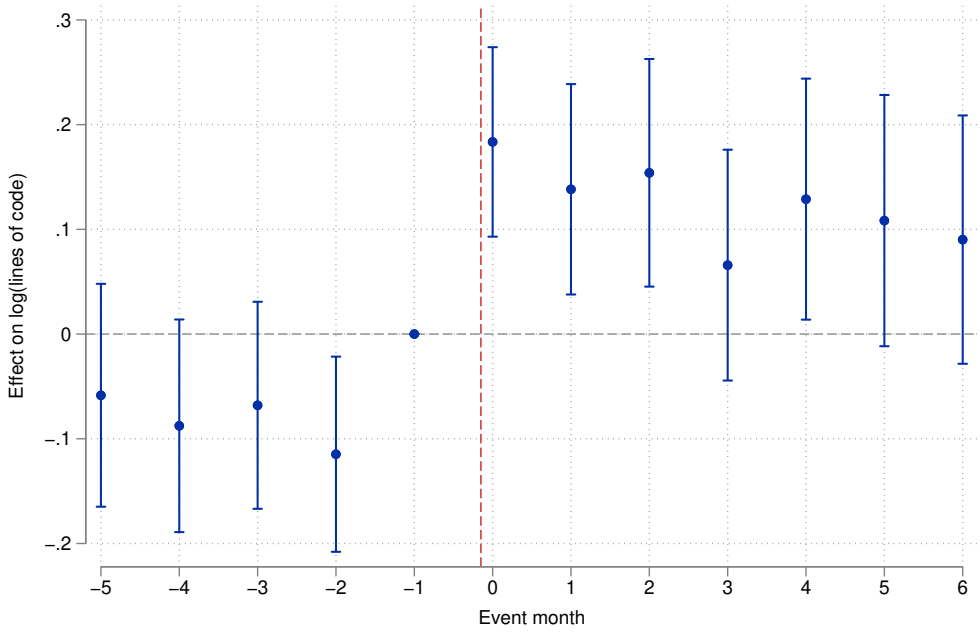


Figure 8: Dynamic Effects for Task Assignment Adoption

(c) Output Quantity



(d) Code Acceptance Rate

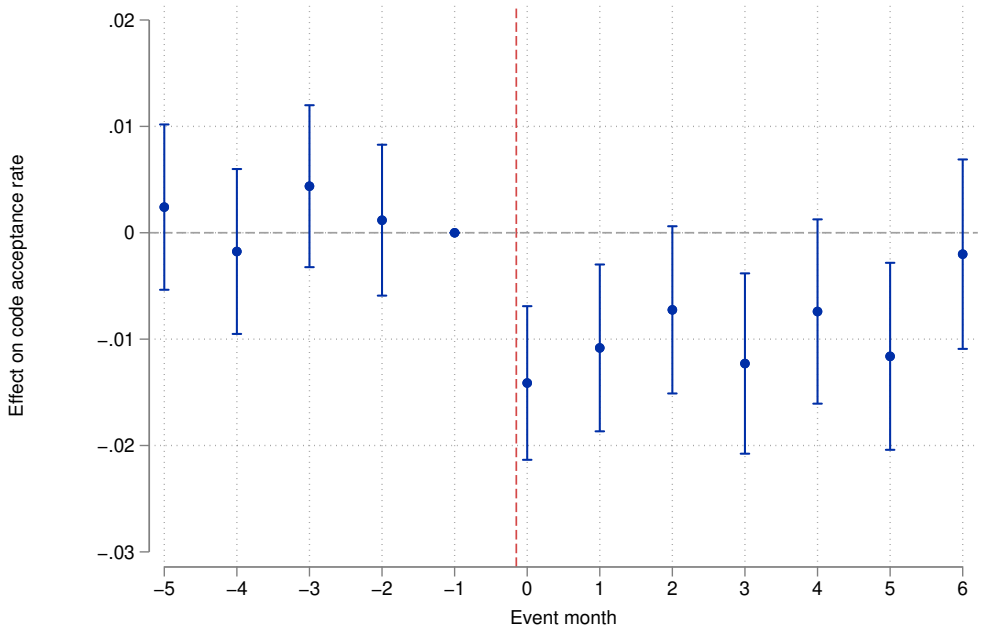
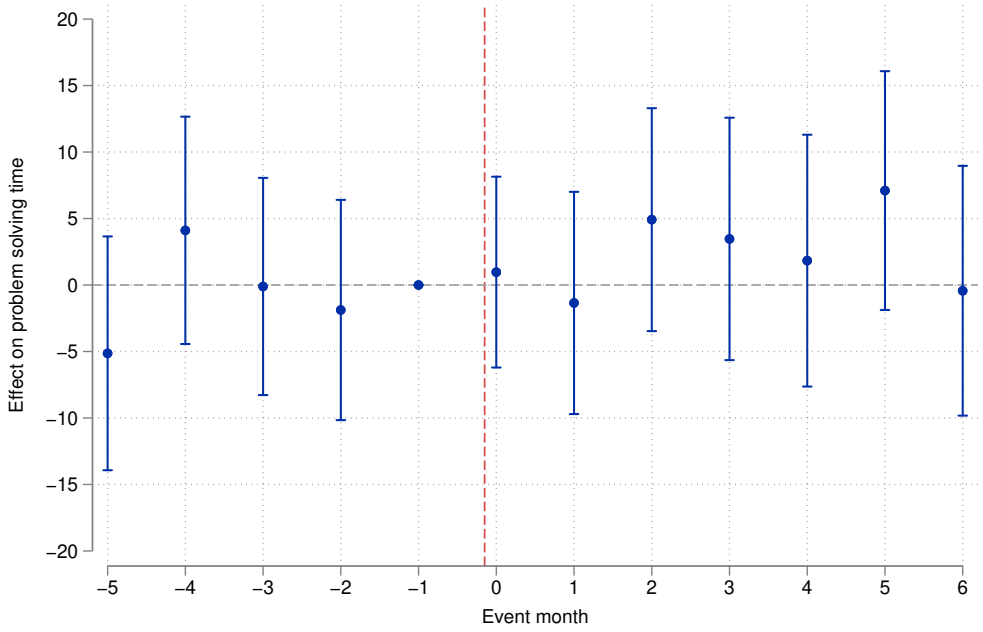
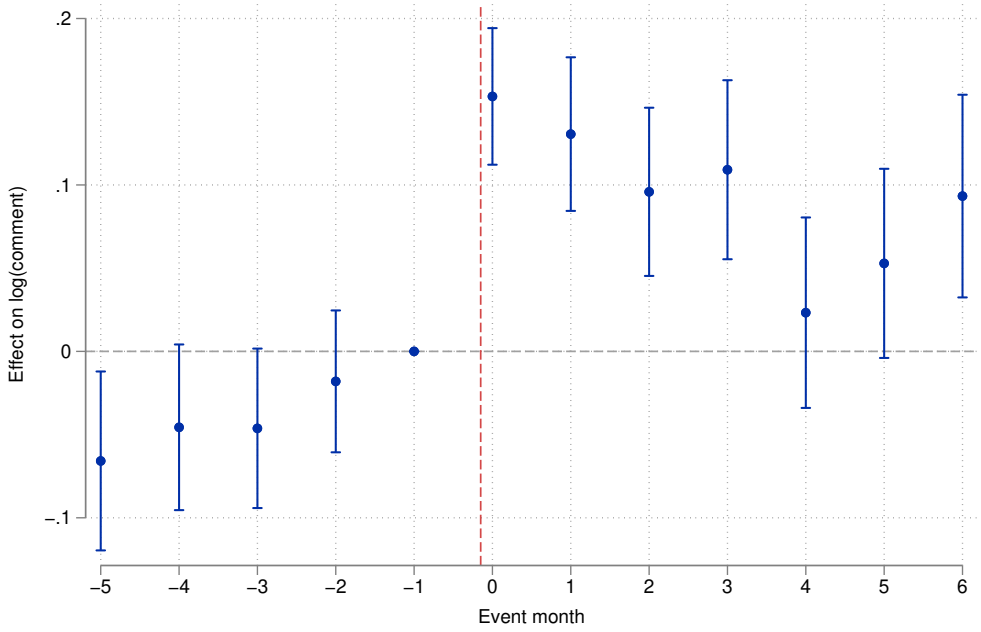


Figure 8: Dynamic Effects for Task Assignment Adoption

(e) Problem Solving Time



(f) Communication



Notes: This plot presents the effect of auto-assignment feature adoption on team organizational specialization and productivity. The vertical lines represent a 95% confidence interval with standard errors clustered at the matching level. The dependent variables—team specialization index (SPE), team output quality, quantity, problem-solving time, code acceptance rate, and communication are defined in Section 3.

Table 1: Task Classification

	Task Type	Keywords (LDA result)	Occupational Role
1	Frontend development	Frontend, UI	Front-end engineer
2	Server management, Platform migration	Client, server	DevOps engineer
3	Android mobile development	Kotlin, runtime	Mobile engineer
4	Cloud feature implementation	Feature, sdk	Cloud engineer
5	Data management	Data, web	Data engineer
6	Internal system management	Internal, apache	System Administrator
7	CLI Development and Framework	User, cli	Technical Writer
8	API and Backend Services	API, controller	Back-end engineer
9	Integration system	Integration, function	System Integrator
10	App Development and UI Design	App, style	App Developer

Notes: This table shows the results of task classification using LDA on 34,762,813 code filenames from 2017 to 2023. Keywords are from the LDA model. Task summary and job role are based on keywords.

Table 2: Task Share Statistics

Task Type	**N**	**Mean**	**St. Dev.**	**Min**	**P75**	**Max**
Frontend development	1054274	0.461	0.291	0	0.711	0.998
Server management	1054274	0.056	0.143	0	0.019	0.994
Android mobile development	1054274	0.056	0.138	0	0.033	0.994
Cloud feature implementation	1054274	0.046	0.126	0	0.010	0.993
Data management	1054274	0.059	0.143	0	0.032	0.997
Internal system management	1054274	0.047	0.128	0	0.008	0.998
CLI Development and Framework	1054274	0.051	0.129	0	0.028	0.993
API and Backend Services	1054274	0.078	0.164	0	0.069	0.998
Integration system	1054274	0.055	0.135	0	0.030	0.993
App Development and UI Design	1054274	0.091	0.171	0	0.103	0.997

Notes: This table provides summary statistics for the share of labor required by each task across all team-months from 2017-01 to 2023-12. Sample includes only teams with valid specialization index (Team size > 1 & Task types > 1).

Table 3: Task Assignment Example

Actual (A_m)

	Task 1	Task 2	Task 3	
A	1/2	1/6	1/3	1
B	1/2	1/6	1/3	1
C	1/2	1/6	1/3	1
	3/2	1/2	1	

Generalized (G_m)

	Task 1	Task 2	Task 3	
A	1/2	1/6	1/3	1
B	1/2	1/6	1/3	1
C	1/2	1/6	1/3	1
	3/2	1/2	1	

Specialized (S_m)

	Task 1	Task 2	Task 3	
A	1	0	0	1
B	1/2	1/2	0	1
C	0	0	1	1
	3/2	1/2	1	

Notes: Each row sum in the matrix represents a member's task assignment, $l_m(i)$ and each column sum is the task share $\alpha_m(j)$. Given three members (A, B, C) and three task types (1, 2, 3), Actual (A_m) shows an example of how each member allocates their labor supply across the tasks. The most generalized (G_m) assignment is when all members evenly split their labor across all tasks as defined in Definition 1, and the most specialized assignment aims to assign each member to a single task, with any remaining task share allocated to another member due to unequal task shares (S_m) as defined in Definition 2.

Table 4: Descriptive Statistics

	N	Mean	St. Dev.	Min	Median	Pctl(75)	Pctl(95)
SPE	1054274	0.48	0.35	0.00	0.50	0.77	1.00
Team size	1054274	4.48	6.18	2.00	3.00	5.00	11.00
Task type	1054274	3.69	1.94	2.00	3.00	5.00	8.00
Lines of code	1054274	26774.80	298346.68	0.00	1409.00	6364.00	63727.05
Activities	1054274	422.36	1900.69	0.00	185.00	417.00	1420.00
Monthly Stars	1054274	28.06	183.38	0.00	1.00	9.00	119.00
Communication	1054274	77.22	244.55	0.00	22.00	67.00	300.00
Solving time	635223	54.73	130.85	0.00	13.05	43.32	252.57
Edited files	1054274	331.41	2270.44	0.00	56.00	185.00	1063.00
Code acceptance	999456	0.83	0.20	0.00	0.89	1.00	1.00
Create year	1054274	2018.42	2.76	2011	2019	2020	2022

Notes: This table provides the summary statistics for the main variables of interest at the team-month level. Data is from 2017-01 to 2023-12. Only include teams with valid SPE index (team size>1 & task type>1). Code acceptance rate, activities, edited files, lines of code, and communication metrics are measured for team members only.

Table 5: Specialization and Team Popularity

	(1)	(2)	(3)	(4)
OLS: dependent var- log(monthly star)				
SPE	-0.321***	-0.308***	-0.080***	-0.087***
	(0.021)	(0.021)	(0.007)	(0.007)
Team age	0.066***	0.060***	-0.046***	-0.038***
	(0.005)	(0.005)	(0.002)	(0.002)
Dependent mean	2.13	2.13	2.13	2.13
R-squared	0.648	0.650	0.903	0.910
Observations	1,823,750	1,823,750	1,823,750	1,770,310
Poisson: dependent var- monthly star				
SPE	-0.431***	-0.376***	-0.146***	-0.149***
	(0.061)	(0.059)	(0.043)	(0.035)
Team age	0.004	-0.003	-0.089***	-0.078***
	(0.009)	(0.009)	(0.008)	(0.006)
Dependent mean	45.738	49.877	48.522	49.234
Observations	4,162,380	3,827,976	2,918,500	2,755,378
4,611,188	4,206,778	4,325,677	4,205,368	
Task type share control	Y	Y	Y	Y
Team size fixed effect	Y	Y	Y	Y
Firm fixed effect	Y	Y		
Project type fixed effect		Y		
Team fixed effect			Y	Y
Member fixed effect				Y

Notes: This table presents OLS and Poisson pseudo maximum likelihood (PPML) regression results on the impact of team specialization. The outcome variable for OLS is the log of monthly stars received by each team, while the outcome variable for Poisson is the monthly star count, which captures the extensive margin. Team age is calculated as the difference between the calendar year and the year the team was created. Observations are at the individual level and weighted by 1/team size. Standard errors are clustered at the team level. $p < .10$, ** $p < .05$, *** $p < .01$.

Table 6: Specialization and Team Output

	(1)	(2)	(3)	(4)
OLS: Dependent var - log (lines of code)				
SPE	-1.239***	-1.275***	-0.954***	-0.942***
	(0.021)	(0.022)	(0.015)	(0.015)
Team age	-0.055***	-0.056***	-0.116***	-0.121***
	(0.004)	(0.004)	(0.004)	(0.004)
Dependent mean	8.124	8.128	8.124	8.118
R-squared	0.398	0.403	0.597	0.616
Observations	3,212,871	2,982,107	3,212,871	3,124,751
Poisson: Dependent var - lines of code				
SPE	-0.382***	-0.422***	-0.588***	-0.567***
	(0.110)	(0.119)	(0.072)	(0.073)
Team age	0.000	-0.005	-0.063***	-0.065***
	(0.020)	(0.021)	(0.016)	(0.015)
Dependent mean	4.4e+04	4.4e+04	4.4e+04	4.3e+04
Observations	3,243,291	3,000,021	3,235,747	3,147,523
Task type share control	Y	Y	Y	Y
Team size fixed effect	Y	Y	Y	Y
Firm fixed effect	Y	Y		
Project type fixed effect		Y		
Team fixed effect			Y	Y
Member fixed effect				Y

Notes: This table presents OLS and Poisson pseudo-maximum likelihood (PPML) regression results on the impact of team specialization on code quantity. The outcome variable for OLS is the log of lines of code additions, while the outcome variable for Poisson models is the raw count of lines of code, which captures the extensive margin. Team age is calculated as the difference between the calendar year and the year the team was created. Observations are at the individual level and weighted by 1/team size. Standard errors are clustered at the team level. $p < .10$, ** $p < .05$, *** $p < .01$.

Table 7: Specialization and Acceptance Rate

	(1)	(2)	(3)	(4)
SPE	-0.011***	0.020***	0.015***	0.015***
	(0.002)	(0.001)	(0.001)	(0.001)
Team age	-0.006***	-0.005***	-0.007***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.000)
Task type share control	Y	Y	Y	Y
Team size fixed effect	Y	Y	Y	Y
Firm fixed effect	Y	Y		
Project type fixed effect		Y		
# code submission		Y	Y	Y
Team fixed effect			Y	Y
Member fixed effect				Y
R-squared	0.217	0.214	0.417	0.442
Dependent mean	0.824	0.825	0.824	0.824
Observations	3,213,677	2,982,846	3,213,677	3,125,570

Notes: This table presents OLS regression results examining the relationship between team specialization and code submission acceptance rate. The dependent variable is the average code submission rate by team members. Team age is calculated as the difference between the calendar year and the year the team was created. Observations are at the individual level and weighted by 1/team size. Standard errors are clustered at the team level. $p < .10$, ** $p < .05$, *** $p < .01$.

Table 8: Specialization and Problem-Solving Time

	(1)	(2)	(3)	(4)
SPE	18.422***	11.894***	4.704***	4.403***
	(1.510)	(1.547)	(1.386)	(1.330)
Team age	-5.322***	-6.255***	-17.360***	-17.235***
	(0.313)	(0.342)	(0.691)	(0.739)
Task type share control	Y	Y	Y	Y
Team size fixed effect	Y	Y	Y	Y
Firm fixed effect	Y	Y		
Project type fixed effect		Y		
# question fixed effect		Y	Y	Y
Team fixed effect			Y	Y
Member fixed effect				Y
R-squared	0.166	0.166	0.355	0.398
Dependent Mean	51.182	51.372	51.182	51.579
Observations	2,364,271	2,267,892	2,364,271	2,294,737

Notes: This table presents OLS regression results examining the relationship between team specialization and issue solving time. The dependent variable is the average time a team takes to resolve an issue raised by users. Team age is calculated as the difference between the calendar year and the year the team was created. Observations are at the individual level and weighted by 1/team size. Standard errors are clustered at the team level. $p < .10$, ** $p < .05$, *** $p < .01$.

Table 9: Balance Table for Even Task Assignment Adoption

	Control		Treatment		Difference
	Mean	# Obs.	Mean	# Obs.	
Team size	6.93	3864	7.86	3864	-0.92
					(0.21)
Task types	4.79	3864	5.84	3864	-1.05
					(0.05)
Lines of code	49824.41	3864	55258.12	3864	-5433.72
					(11440.01)
Monthly stars	24.11	3864	59.73	3864	-35.61
					(3.71)
Specialization index	0.56	3864	0.53	3864	0.02
					(0.01)
Code acceptance	0.83	3818	0.83	3834	0.01
					(0.00)
Problem-solving time	49.21	2281	43.02	2969	6.19
					(2.69)
Create year	2018.38	3864	2018.16	3864	0.22
					(0.06)

Notes: The table presents summary statistics for both the control and treatment groups at event month $k = 0$, each consisting of 3,864 teams. The treatment group includes teams that adopted even task assignments in event month t. Notably, I only matched on decile of average team size, task type, and lines of code but did not specifically match the characteristics of teams.

References

Bassi, V., L. Jung Hyuk, A. Peter, T. Porzio, R. Sen, and E. Tugume (2023). Self-employment within the firm. *Working paper 123*, 1–18.

Baumgardner, J. R. (1988). The division of labor, local markets, and worker organization. *Journal of Political Economy 96*(3), 509–527.

Becker, G. S. and K. M. Murphy (1992). The division of labor, coordination costs, and knowledge. *The Quarterly journal of economics 107*(4), 1137–1160.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of machine Learning research 3*(Jan), 993–1022.

Borges, H. and M. T. Valente (2018). What's in a github star? understanding repository starring practices in a social coding platform. *Journal of Systems and Software 146*, 112–129.

Casalnuovo, C., B. Vasilescu, P. Devanbu, and V. Filkov (2015). Developer onboarding in github: the role of prior social links and language experience. In *Proceedings of the 2015 10th joint meeting on foundations of software engineering*, pp. 817–828.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

El-Komboz, L. A. and M. Goldbeck (2024). Career concerns as public good the role of signaling for open source software development. Technical report, ifo Working Paper.

Eldeeb, Y. and A. Sikora (2023). How much are github stars worth to you? (the guild). Accessed: 2024-10-11.

Forderer, J. and G. Burtch (2024). Estimating career benefits from online community leadership: Evidence from stack exchange moderators. *Management Science*.

Garicano, L. and T. N. Hubbard (2009). Specialization, firms, and markets: The division of labor within and between law firms. *The Journal of Law, Economics, & Organization 25*(2), 339–371.

Garicano, L. and E. Rossi-Hansberg (2006). Organization and inequality in a knowledge economy. *The Quarterly journal of economics 121*(4), 1383–1435.

Gong, J. and I. P. Png (2024). Automation enables specialization: Field evidence. *Management Science 70*(3), 1580–1595.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review 35*(4), 519–530.

Hong, L. and S. E. Page (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences 101*(46), 16385–16389.

Jaravel, X., N. Petkova, and A. Bell (2018). Team-specific capital and innovation. *American Economic Review 108*(4-5), 1034–1073.

Kahane, L., N. Longley, and R. Simmons (2013). The effects of coworker heterogeneity on firm-level output: assessing the impacts of cultural and language diversity in the national hockey league. *Review of Economics and Statistics 95*(1), 302–314.

Krebs, C. J. (1989). Ecological methodology. *(No Title)*.

Laohaprapanon, S., G. Sood, and B. Naji (2017). ethnicolr algorithm.

Lazear, E. P. (1999). Globalisation and the market for team-mates. *The Economic Journal 109*(454), 15–40.

Levinthal, D. A. and J. G. March (1993). The myopia of learning. *Strategic management journal 14*(S2), 95–112.

Prat, A. (2002). Should a team be homogeneous? *European Economic Review 46*(7), 1187–1207.

Rosen, S. (1983). Specialization and human capital. *Journal of Labor Economics 1*(1), 43–49.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights 4*(3), 305–322.

Sharoni, B. (2024). The effect of inventor mobility on network productivity. Technical report, Working Paper. Harvard University.

Smith, A. (1776). *The wealth of nations [1776]*, Volume 11937. na.

Vasilescu, B., Y. Yu, H. Wang, P. Devanbu, and V. Filkov (2015). Quality and productivity outcomes relating to continuous integration in github. In *Proceedings of the 2015 10th joint meeting on foundations of software engineering*, pp. 805–816.

Wagner, S. and M. Ruhe (2018). A systematic review of productivity factors in software development. *arXiv preprint arXiv:1801.06475*.

Online Appendix of:

How does Division of Labor affect Team Productivity?

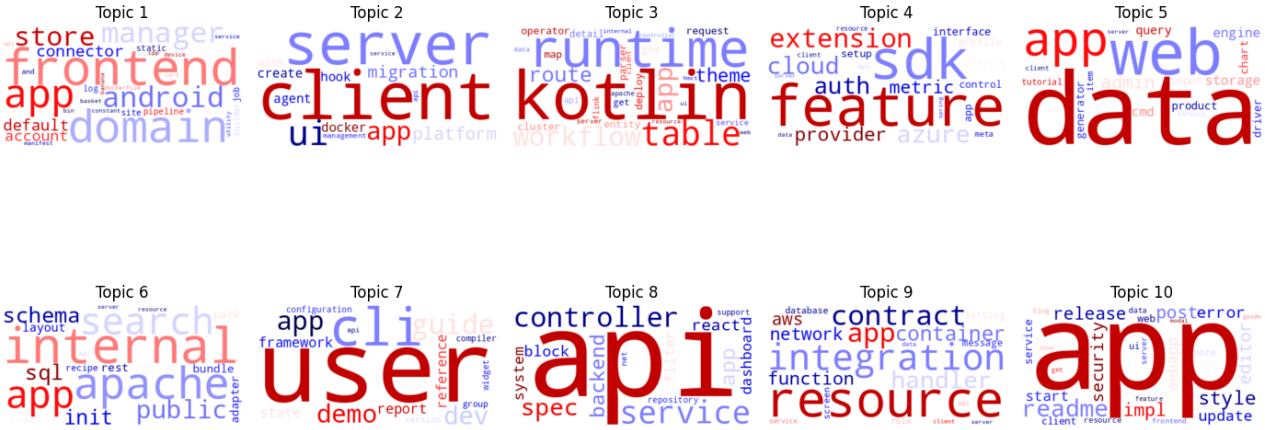
Jinci Liu

February 27, 2025

A Task Classification Process

[Figure A1 about here.]

Figure A1: Word Clouds for LDA 10 Task Types

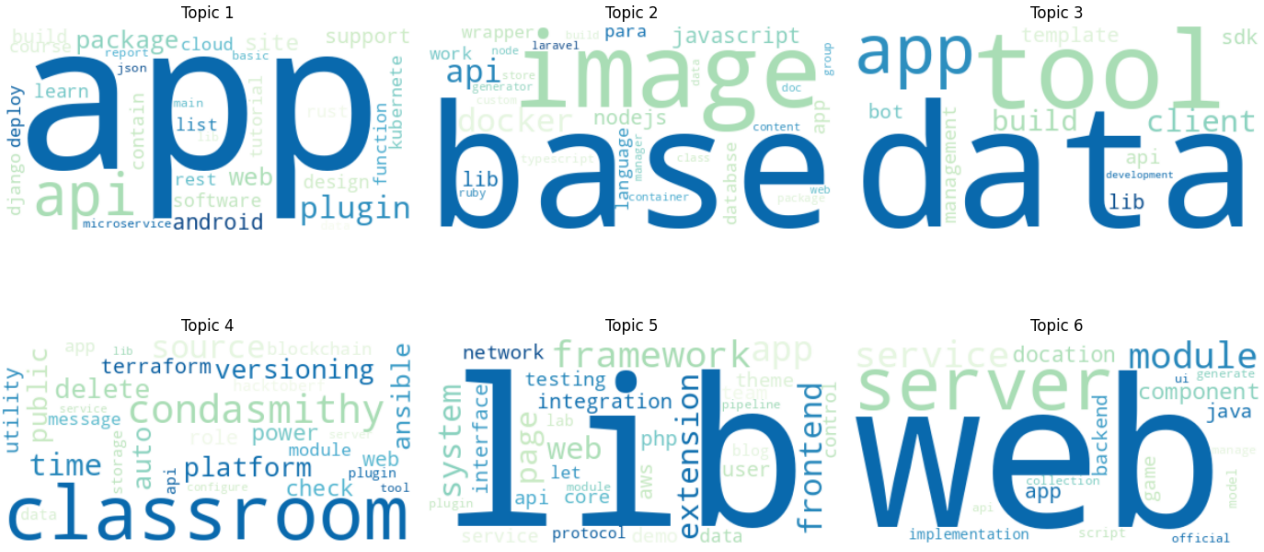


Notes: This figure presents word clouds visualizing the 10 task types identified through Latent Dirichlet Allocation (LDA) topic modeling. Each panel shows frequently occurring terms within different development domains: data management (Topic 5), server-side development (Topic 2), runtime environments (Topic 3), client applications (Topic 7), API development (Topic 8), and web applications (Topic 10). The size of each word represents its frequency within the topic, with terms like "data," "api," "user," and "app" being particularly prominent. This classification helps measure team specialization by categorizing tasks into distinct technical domains. Topics reflect common software development tasks and technological components.

B Project Type Classification Process

[Figure B1 about here.]

Figure B1: Word Clouds for LDA 6 Project Types



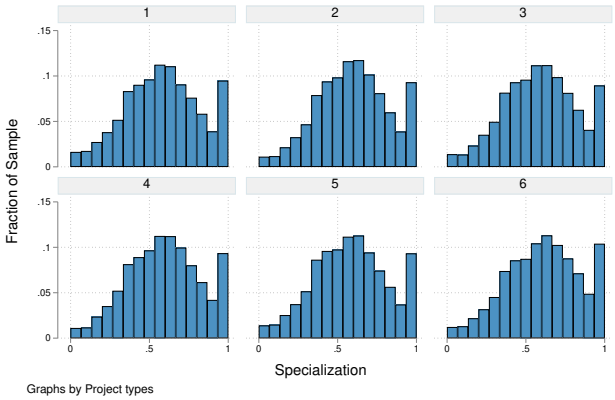
Notes: This figure presents the six project types identified through Latent Dirichlet Allocation (LDA) topic modeling. Each panel represents a distinct project category: mobile applications (Topic 1, dominated by "app"), database management systems (Topic 2, centered on "base"), development tools (Topic 3, highlighted by "tool"), educational projects (Topic 4, featuring "classroom"), library development (Topic 5, showing "lib"), and web applications (Topic 6, emphasized by "web"). The word clouds display terms scaled by their frequency within each project type, helping classify repositories into broad development domains. Project type classification helps control for heterogeneity across different software development contexts when analyzing the relationship between specialization and productivity. LDA analysis performed on repository metadata, README files, and project descriptions.

C Additional Results

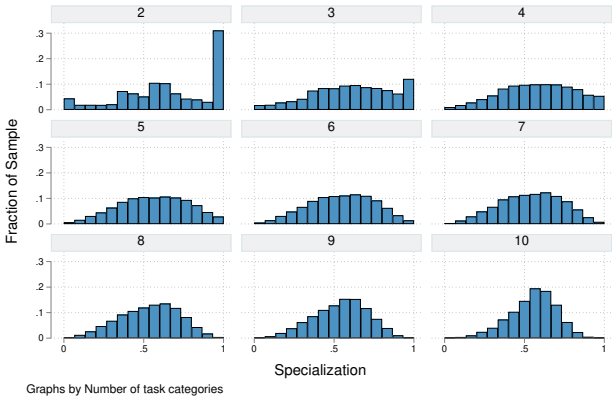
[Figure B1 about here.]

Figure C1: Distribution of Team Specialization Index

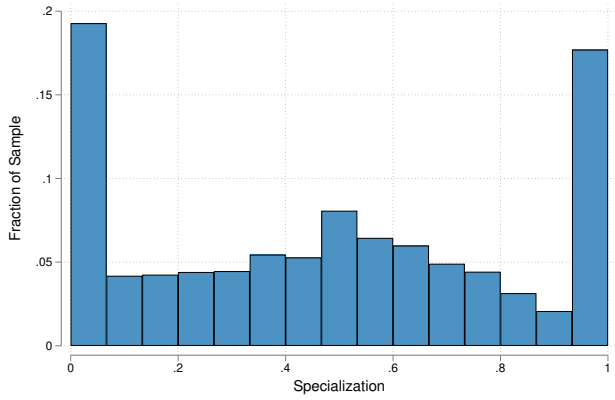
(a) By Project Type



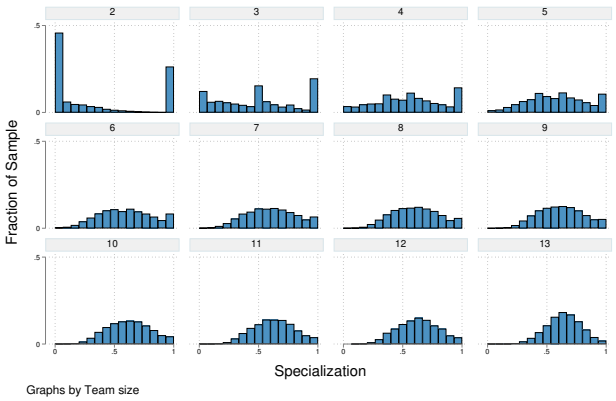
(b) By Number of Task Type



(c) Include Small Teams



(d) By Team Size



Notes: This figure illustrates the distribution of the team specialization index across various dimensions. Figures C1a and C1b display the distribution by project type and number of tasks for all team-month observations, excluding teams with fewer than four workers. Additional details on project types and task types are provided in Sections 2.5 and 2.4. Figures C1c and C1d also present the distribution for small teams consisting of 2 or 3 workers. These smaller teams are excluded from the main analysis because their limited size restricts their ability to vary their degree of specialization, resulting in most being either highly specialized or generalized.

D Examples

[Figure D1 about here.]

[Figure D2 about here.]

[Figure D3 about here.]

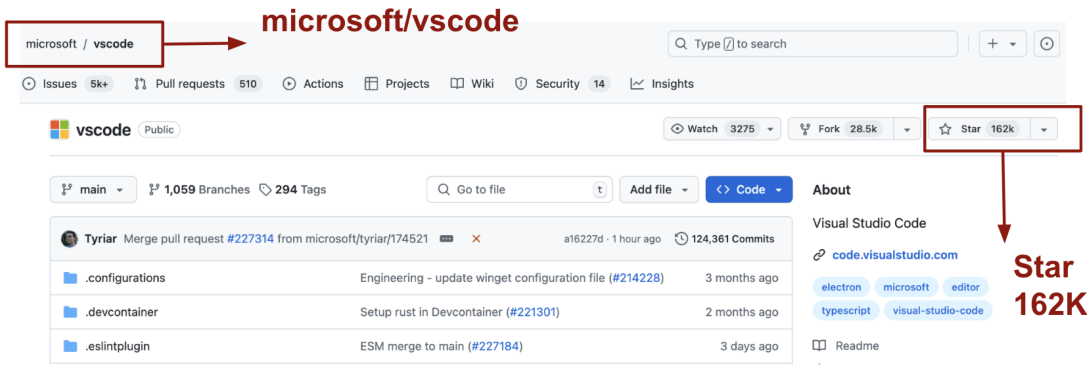
[Figure D4 about here.]

[Figure D5 about here.]

[Figure D6 about here.]

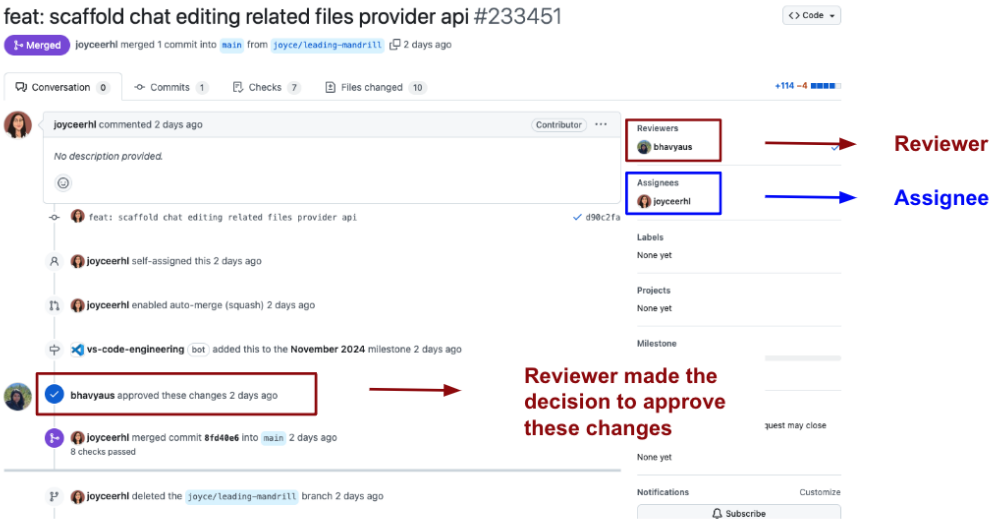
[Figure D7 about here.]

Figure D1: Star Example



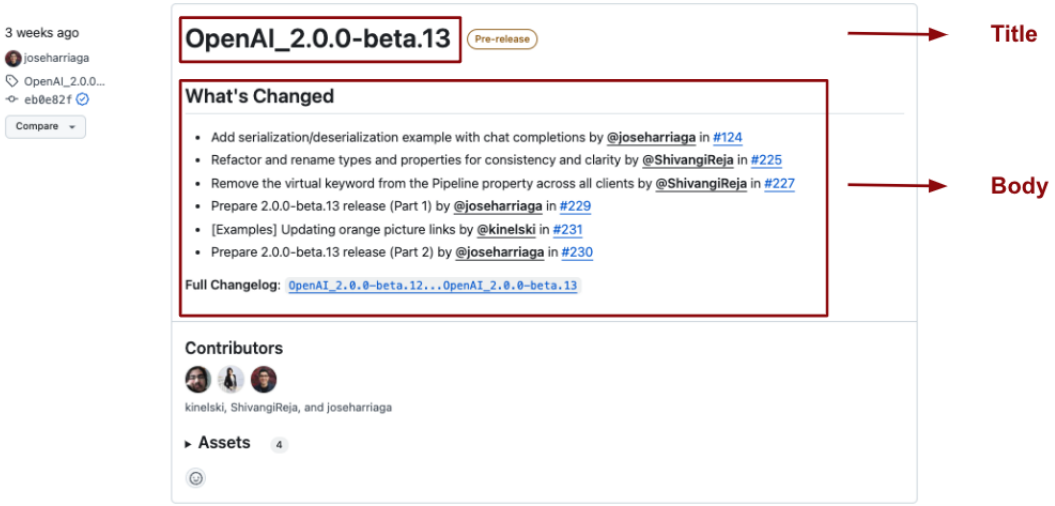
Notes: This figure illustrates how project popularity is measured using GitHub's star system, using Microsoft's Visual Studio Code repository as an example. Stars represent a form of peer evaluation where users can bookmark and show appreciation for repositories they find valuable. In this example, Visual Studio Code has received 162,000 stars, indicating its high popularity within the developer community. This metric serves as one of our key measures of team productivity, as it captures the project's impact and perceived value among potential users and other developers. While not a perfect measure of quality, stars provide a standardized way to compare project impact across different domains and team sizes.

Figure D2: Reviewer and Assignee



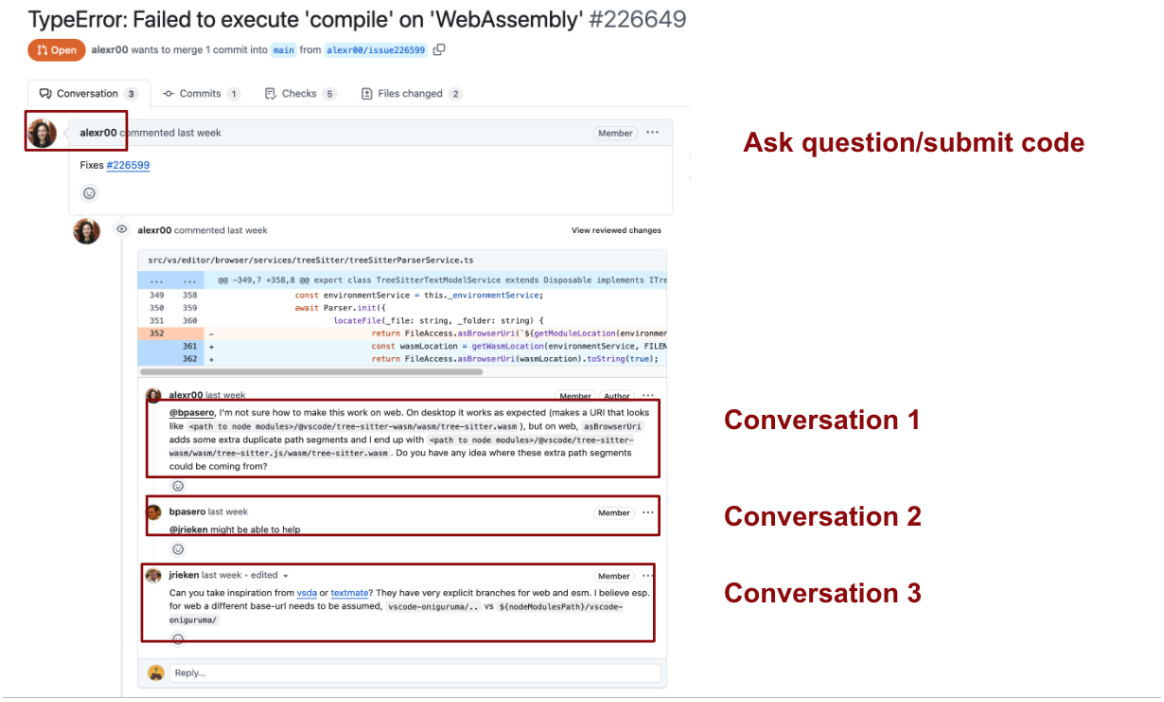
Notes: The example shows a typical Pull Request (#233451) with its associated review thread. Two distinct roles are highlighted: Reviewers (red box) who evaluate code changes and Assignees (blue box) who are responsible for implementing the changes. The bottom of the image demonstrates a completed review process where a reviewer has approved the proposed changes.

Figure D3: Project Release Example



Notes: This figure illustrates how software releases are documented on GitHub, using OpenAI_2.0.0-beta.13 as an example. A release consists of two main components: a title indicating the version number (highlighted in red) and a detailed body describing changes (highlighted in red box). The body contains specific improvements, bug fixes, and feature additions.

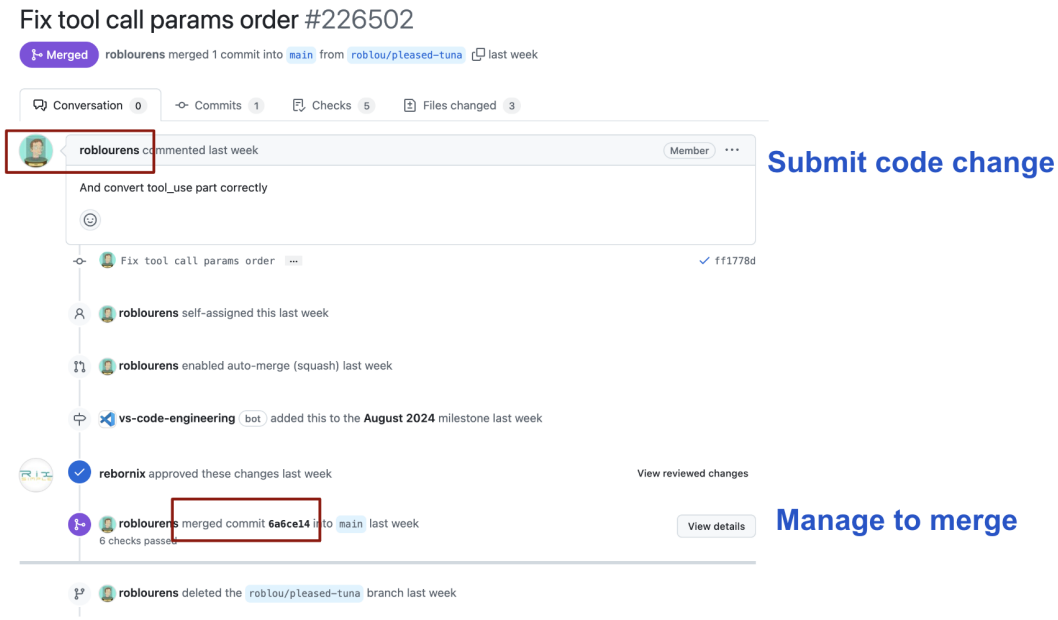
Figure D4: Conversation Example



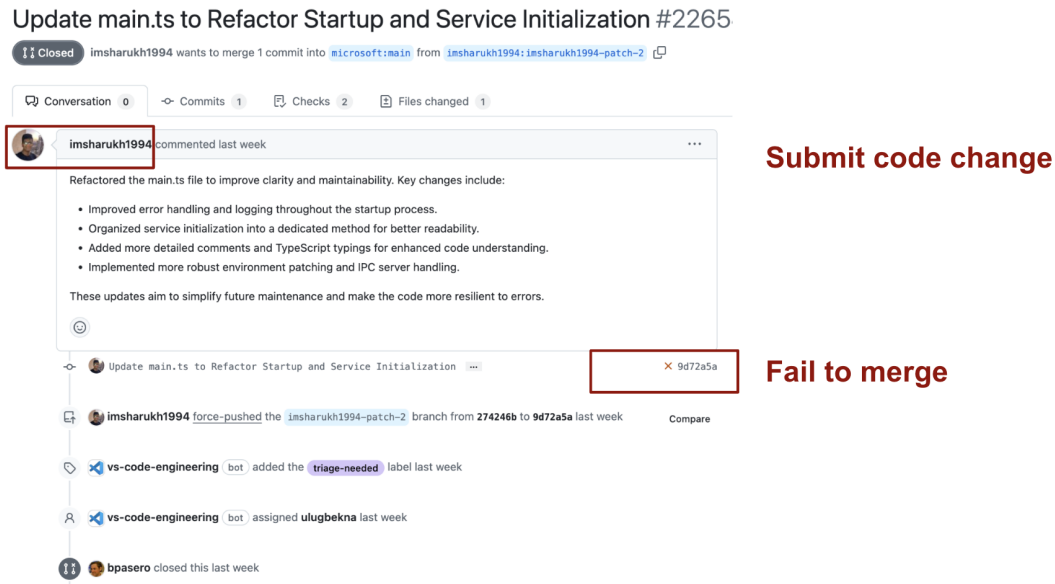
Notes: This figure demonstrates how technical discussions unfold in GitHub's issue tracking system. The example shows a TypeError issue (#226649) and the subsequent problem-solving conversation. The interaction is structured in three parts: (1) initial question/code submission, (2) technical discussions and proposed solutions, and (3) follow-up clarifications. This conversation structure allows us to measure the quantity of team communication (number of comments).

Figure D5: Pull Request Merge Example

(a) Manage to merge code

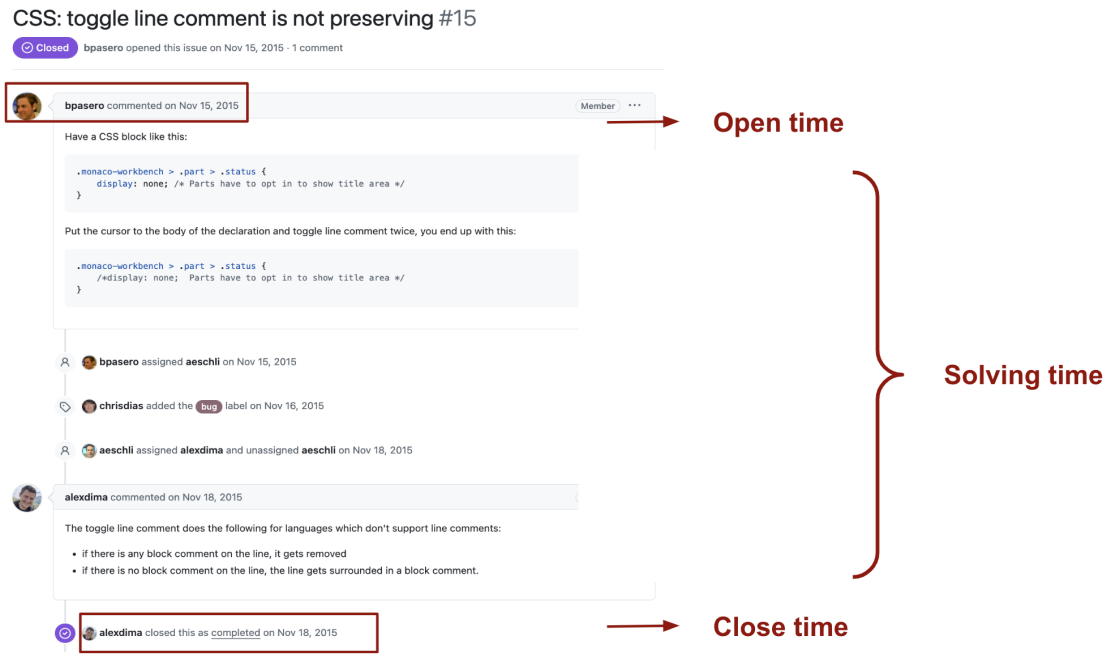


(b) Fail to merge code



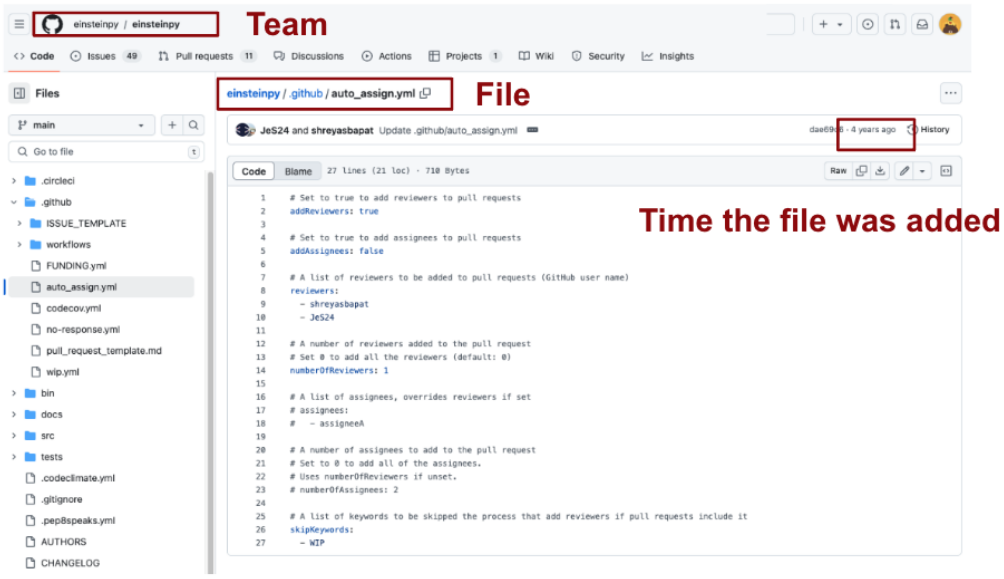
Notes: This figure illustrates two outcomes in the code review process: successful and failed merges. Panel (a) shows a successful merge (#226502) where submitted code changes are approved and integrated into the main codebase. Panel (b) demonstrates a failed merge (#2265) where proposed changes are rejected. Code acceptance rates serve as one of our key productivity metrics.

Figure D6: Problem solving example



Notes: This figure demonstrates how we measure problem-solving efficiency in teams. Using issue #15 as an example, we track three key timestamps: (1) Open time - when the issue is first reported, (2) Solving time - duration of technical discussion and solution development, and (3) Close time - when the issue is resolved. The time between open and close serves as our measure of problem-solving speed.

Figure D7: Auto Assignment File Example



Notes: This figure shows the implementation of GitHub's random assignment feature through a configuration file. The left panel shows the team's repository structure, while the right panel displays the auto-assignment configuration. The timestamp of file addition marks when teams transition to more generalist task allocation, providing our second source of identification.