Resolving governance disputes in communities: A study of software license decisions

Originally published in:
Strategic Management Journal 41(10), https://doi.org/10.1002/smj.3181

Funding acknowledgement:
169441 - Leading and Coordinating Knowledge Creation in Online Communities (SNF)

Received: 31 January 2018 | Revised: 2 April 2020 | Accepted: 2 April 2020 | Published on: 24 June 2020

DOI: 10.1002/smj.3181

RESEARCH ARTICLE

STRATEGIC MANAGEMENT JOURNAL WILEY

Resolving governance disputes in communities: A study of software license decisions

Vivianna Fang He[1] | **Phanish Puranam**[2] | **Yash Raj Shrestha**[1] | **Georg von Krogh**[1]

[1]ETH Zurich, Zurich, Switzerland

[2]INSEAD, Singapore, Singapore

Correspondence
Vivianna Fang He, ETH Zurich, Weinbergstrasse 56-58, Zurich CH 8092, Switzerland.
Email: fhe@ethz.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 169441

Abstract

Research summary: Resolving governance disputes is of vital importance for communities. Gathering data from GitHub communities, we employ hybrid inductive methods to study discussions around initiation and change of software licenses—a fundamental and potentially contentious governance issue. First, we apply machine learning algorithms to *identify* robust patterns in data: resolution is more likely in larger discussion groups and in projects without a license compared to those with a license. Second, we analyze textual data to *explain* the causal mechanisms underpinning these patterns. The resulting theory highlights the group process (reflective agency switches disputes from bargaining to problem solving) and group property (preference alignment over attributes) that are both necessary for the resolution of governance disputes, contributing to the literature on community governance.

Managerial summary: Online communities play an increasingly important role in how companies innovate across organizational boundaries and attract talent across geographic locations. However, online communities are no Utopia; disputes abound even (more) when we collaborate virtually. In particular, governance disputes can threaten the functioning and existence of online communities. Our study suggests that

governance disputes in online communities either unfold as bargaining over which solution is better or searching for a satisfactory solution. The latter is more likely to reach a resolution, when there is common ground. Companies interested in leveraging the power of online communities should (a) identify or train certain participants to transform endless bargaining into collective problem solving and (b) foster shared knowledge and value basis among participants through recruitment and strong organizational culture.

KEYWORDS

community governance, dispute resolution, inductive theorizing, machine learning, problem solving

1 | INTRODUCTION

Communities are an increasingly important form of organizing (Adler, 2001; Adler & Heckscher, 2017; Lee & Edmondson, 2017). Prominent examples come from domains as diverse as open source software (OSS) (Dahlander & Frederiksen, 2012; Foss, Frederiksen, & Rullani, 2016; O'Mahony & Ferraro, 2007; von Hippel & von Krogh, 2003), common pool resources (CPR[1]) (Dietz, Ostrom, & Stern, 2003; Ostrom, 1990), scientific consortia (e.g., Human Genome Project, Collins, Morgan, & Patrinos, 2003), and communities of practice within and across firms (e.g., 3D printing design, Stanko, 2016).

Like any organization that draws on diverse resources and efforts to pursue a set of goals, communities encounter disputes among their constituent individuals and coalitions (Deutsch & Krauss, 1962; Jehn & Bendersky, 2003; March, 1962; see also Amit & Schoemaker, 1993; Danneels, 2008). Dispute resolution is therefore a hallmark of effective and efficient community governance (Ostrom, 1990). An organization's core governance principles—about the direction, control, and coordination of its members—are meant (at least in part) to prevent the emergence of destructive disputes, and when disputes do occur, to aid in their resolution (Gulati, Lawrence, & Puranam, 2005; Gulati & Singh, 1998; March & Simon, 1958; Williamson, 1975).

However, disputes can also erupt around the very governance principles themselves. If not resolved, disputes over fundamental principles such as the distribution of resources, rights, and responsibilities, can derail the functioning or even jeopardize the existence of a community. Disputes pertaining to core governance principles that affect the entire community (*governance disputes* hereafter) differ substantially from disputes concerning operational matters (e.g., the division and allocation of tasks, or management of resources) that typically affect only a subset of community members at a time (henceforth *operational disputes*). Whereas theory and research on mechanisms for resolving operational disputes are rapidly growing, studies on

[1]A type of good consisting of natural or human-made resource systems (e.g., irrigation systems, forests, or fisheries) that is characterized by nonexcludability (difficult to exclude actors from gaining utility from the resource) and subtractability (use by one actor precludes use by another).

governance disputes remain rare (Saeed, McDermott, & Boyd, 2017). The nascent state of knowledge on how communities resolve governance disputes (Dietz et al., 2003) warrants empirically driven, inductive investigation.

One plausible reason why research is still scant on governance disputes in communities is that records of such disputes are often inaccessible. Those momentous occasions, when most if not all members were involved in resolving disputes over governance principles remain largely buried in the communities' history, not documented or available for research. It is also possible that unresolved governance disputes removed those protocommunities from possible scholarly scrutiny today. As a result, analysis of governance disputes may be difficult for researchers working with data limited to functioning communities (Ostrom, 1990), since by virtue of their existence, these communities have escaped or resolved significant governance disputes and their counterparts that did not may no longer exist.

We leverage an unusual research opportunity found in the public discussions on GitHub, an online platform hosting numerous open source software development (OSS) communities, to explore the question of *how communities resolve governance disputes*. We aim to understand the conditions that increase or decrease the likelihood of resolution, and the processes that underpin differences in outcomes. Two factors motivate our choice of empirical context: (a) OSS communities are an important contemporary type of community for organizing economic activity (Dahlander & Frederiksen, 2012; O'Mahony & Ferraro, 2007), and (b) they provide a unique window into the phenomenon of governance dispute resolution with a nondecaying electronic record.

Our exploration focuses on a particular type of governance dispute: disputes vis-à-vis the software license that governs a given community's core software development project (Markus, 2007; O'Mahony, 2003). License disputes are of strategic importance for an OSS community, because they determine how the software is used, modified, and combined with other software(s) (Lerner & Tirole, 2005), thereby affecting all downstream users' innovation capabilities (Murray & O'Mahony, 2007). Furthermore, they have significant impact on the community's ability to attract not only developers (Lerner & Tirole, 2005; Subramaniam, Sen, & Nelson, 2009) but also corporate partners, who then combine OSS products or services with proprietary ones (Deodhar, Saxena, Gupta, & Ruohonen, 2012; Economides & Katsamakas, 2006).

We sample disputes surrounding the initiation or revision of a software license decision, collecting quantitative and qualitative data on these disputes. The dispute instances we discovered involve a large volume of data, making it difficult to rely solely on traditional inductive methods that involve manually analyzing each case separately. Therefore, we combine machine learning analysis with in-depth content analysis to develop a novel theory (Puranam, Shrestha, He, & von Krogh, 2020; Tidhar & Eisenhardt, 2020). In effect, we pioneer a two-pronged approach to inductive research: First, using machine-learning algorithms we document robust patterns across the entire sample (183 cases). These patterns are unlikely to be the result of sampling error or functional form idiosyncrasies, because the algorithms we apply vary in the functional forms and we repeatedly split the data to build and test models using out-of-sample prediction accuracy as a criterion. Second, we focus on analyzing in depth a randomly selected sub-sample (61 cases) to enhance the theoretical understanding of mechanisms underlying the patterns we detected.

Our key finding is that governance disputes in OSS communities can take the form of either bargaining or problem solving processes. The latter group process is more likely when "reflective agency" is manifested. This involves individual interventions that can help steer the discussion toward the attributes of the software licenses (e.g., openness and permissiveness) rather than the alternatives themselves (e.g., GPL and MIT). To the extent that the group has the property that there is pre-alignment of preferences on attributes among members, attribute-based

discussions (i.e., discussions focusing predominantly on license attributes) can aid resolution. If either condition (i.e., reflective agency and pre-alignment of attribute preference) is absent, then the dispute remains a hard to resolve bargaining problem.

This study is the first to document the resolution mechanisms in governance disputes on license decisions, and thus provides a complement to the growing literature on operational disputes in online communities (Forte, Larco, & Bruckman, 2009; Klapper & Reitzig, 2018). Its insights help to understand resolution of not only disputes of strategic and economic importance (i.e., how to manage intellectual property and appropriate value) to online communities in particular (Lerner & Schankerman, 2010) but also disputes of existential significance (i.e., how to establish fundamental governance principles) to communities in general (Dietz et al., 2003; Ostrom, 1990).

2 | LITERATURE REVIEW

2.1 | Governance disputes in communities

The central problem in community governance concerns the alignment of members' interests and actions to collectively create and capture value (Gulati, Puranam, & Tushman, 2012; Olson, 1965; O'Mahony & Bechky, 2008; von Krogh, Haefliger, Spaeth, & Wallin, 2012). As members seek to realize both collective and individual interests through community involvement, they inevitably experience disagreements or encounter disputes at some point. As a result, mechanisms for effective dispute resolution are a necessary condition to sustain collective action over time (Hardin, 2005; Morrill & Rudes, 2010).

Studying communities such as those that organize around common pool resources (CPR), scholars reasoned that explicit collective goals to sustain and build resources would be sufficient to align members' interests and agreement on rules for use of those resources (Adler & Heckscher, 2017; Heckscher & Adler, 2006; Sherif, 1958). However, Ostrom and colleagues (Dietz et al., 2003; Ostrom, 1990) found that disputes do arise across communities with such goals, and concluded that community viability depended on one or more (explicit) dispute resolution mechanisms (Dietz et al., 2003; Wilson, Ostrom, & Cox, 2013).

Ostrom (1990) laid out a set of eight governance principles that she argued would allow communities to effectively manage their resources (see Cox, Arnold, & Tomás, 2010 for a recent discussion): (a) clear definition of community boundaries (who is in and out), (b) fair system of rewards, (c) collective choice rules that have community approval, (d) monitoring mechanisms, (e) applying gradated sanctions on free riders, (f) conflict resolution mechanisms, (g) autonomy for community self-governance, and (h) coordination with other relevant communities and organizations. Yet, how these governance principles are agreed upon within communities and what happens when disputes about them arise are unknown. Put differently, it is clear that governance principles are necessary to resolve operational disputes, but how disputes about governance principles themselves are resolved remains unspecified (Dietz et al., 2003: 1911; Saeed et al., 2017: 590). We turn next to relevant work on dispute resolution in organizations with a bearing on this issue.

2.2 | Dispute resolution in organizations

Early studies in a structural-rationalist tradition treated disputes as a threat to efficient governance in organizations (Kolb & Putnam, 1992; Weber, 1946), and placed considerable emphasis

on the institution of formal authority as a primary means of dispute resolution (March & Simon, 1958; Simons, 1947). Later, the structural-rationalist tradition was complemented by a cultural tradition rooted in Selznick (1948) portrayal of organizations as simultaneously bureaucratic and relational. Culture, values, norms, and ultimately preferences, were recognized to play an essential role in how disputes are understood among members (Menkel-Meadow, 2001), and dispute resolution therefore "include(s) subtle actions couched in everyday activities (e.g., remedial exchanges, avoidance), as well as formalized negotiation and law, quasi-legal structures (e.g., arbitration, mediation, and employee grievance mechanisms), and collective action" (Morrill & Rudes, 2010: 633). An important contribution of the cultural tradition (which guides our approach) has been the recognition that theorizing about dispute resolution demands careful empirical examination of the contexts in which disputes unfold (Barley & Tolbert, 1991; Budd & Colvin, 2014; Roche, Teague, & Colvin, 2014).

The cultural tradition gave rise to two streams of studies on work-related disputes in organizations, focusing on *roles* and *systems*, respectively (Currie, Gormley, Roche, & Teague, 2017). The first stream reveals the role of managers for exercising their formal organizational authority, expertise, and interpersonal skills to enforce a solution (Karambayya & Brett, 1989), when disputes cannot be resolved laterally. The second centers around alternative dispute resolution (ADR) *system*s, which cover various methods (negotiation, mediation, and arbitration) for settling disputes outside of court, and show that these systems offer flexibility in the choice of mechanisms and their combination to fit the specific context of a dispute (Löhr, Hochmuth, Graef, Wambura, & Sieber, 2017; Ross & Conlon, 2000).

Both structural and cultural approaches have proceeded on the assumption that dispute-related roles and systems exist in organizations. For example, managers are responsible for resolving disputes, and an optimal mix of ADR systems, judicial process, or other resolution mechanisms are readily deployable (Edelman, Krieger, Eliason, Albiston, & Mellema, 2011; Roche & Teague, 2012; Ross & Conlon, 2000). Yet, communities operating in non-hierarchical settings often lack dispute-related roles and pre-designed systems (Currie et al., 2017). In fact, roles and systems are themselves governance principles (Ostrom, 1990), leaving open the question of how disputes are resolved when they arise around the governance principles themselves (see Helms, Oliver, & Web, 2012).

2.3 | Dispute resolution in online communities

Digital technology enables individuals, firms, and other organizations separated by time and space to convene, interact, share, and create resources in online communities, largely on a voluntary basis (Boudreau & Jeppesen, 2015; Foss et al., 2016; Gulati et al., 2012; Lee & Cole, 2003; Miller, Fabian, & Lin, 2009; Moon & Sproull, 2002; Preece, 2000; Sproull & Arriaga, 2007). While members of online communities tend to voluntarily assemble and contribute to a common goal (e.g., developing software), such communities—like their offline counterparts—are by no means exempt from disputes (Arazy, Ortega, Nov, Yeo, & Balila, 2015; Kane, Johnson, & Majchrzak, 2014; Kittur & Kraut, 2010; Lee & Cole, 2003; Raymond, 1999; Wang, Shih, Wu, & Carroll, 2015; Weber, 2004).

Yet, the absence of traditional organizational features such as hierarchy and formal authority distinguishes disputes in online communities from other organizational disputes (Faraj, Jarvenpaa, & Majchrzak, 2011; Faraj, von Krogh, Monteiro, & Lakhani, 2016; Gulati et al., 2012; Lakhani, Lifshitz-Assaf, & Tushman, 2013). Moreover, interactions among community members are (a) anonymous by choice (i.e., individuals may choose not to reveal their identities); (b)

simultaneous (i.e., many participants can interact at the same time creating a "cacophony" of opinions) or asynchronous (there are long time gaps between contributions to a discussion); (c) unrestricted (e.g., participants are free to enter or exit and their participation is not easily controlled); (d) recorded (i.e., past discussion and actions are kept for all to inspect); and (e) low in media richness (i.e., without facial expression or body language). It is reasonable to expect these attributes to give dispute resolution in online communities a distinctive flavor.

An emerging stream of literature has taken on this challenge and has examined operational disputes on (editing) tasks in the Wikipedia online community (Forte et al., 2009; Kittur, Suh, Pendleton, & Chi, 2007; Klapper & Reitzig, 2018; Piskorski & Gorbatâi, 2017). Collectively, these studies point to two surprising aspects. First, the mechanisms (roles and systems) set up with a *global* mandate (i.e., to resolve issues wherever they occur in the community) rarely seem necessary to resolve *local* disputes (i.e., those involving a few participants or a particular task). For instance, since its formation in 2003, the Wiki community's arbitration committee had ruled on less than 500 disputes by 2014 (Piskorski & Gorbatâi, 2017; see parallels to Kellogg, 2009; or Jensen & Scacchi, 2005 study on NetBeans). Conversely, in a similar period Wikipedians had resolved around 640,000 cases of editing conflicts on the "talk" page—an informal discussion forum—with voluntary intervention by a fellow contributor occupying an "administrator" role (Klapper & Reitzig, 2018). It thus appears that Wikipedia contributors are able to resolve operational disputes without invoking the globally accessible resolution mechanism (Forte et al., 2009; see parallels to Löhr et al., 2017). Instead, forms of distributed and informal authority seem more prominent in reaching resolution (Dahlander & O'Mahony, 2011; Klapper & Reitzig, 2018). Second, governance disputes involving basic principles that involve the entire community have either been rare or largely escaped scholarly scrutiny. Thus, questions remain about why governance disputes are rare in online communities and how they are resolved when they do occur.

3 | RESEARCH DESIGN

3.1 | Context

We chose GitHub as the context to examine governance dispute resolution. GitHub is currently the world's largest OSS development platform (Gousios, Vasilescu, Serebrenik, & Zaidman, 2014), hosting 35 million projects and 14 million users. Many GitHub projects comprise communities of individuals (i.e., owners, collaborators, and other contributors) who jointly develop a software product.[2] Each project has at least one owner who initiates the project and holds the copyright of the project's software repository as long as an open source license is not in place. Collaborators are those who have contributed a significant amount of code to the project. The owner has the power to change the visibility of the repository from private to public or vice versa, and to delete parts or the entire repository when she holds the sole copyright. However, once a license is adopted, ownership is distributed according to license terms. The adoption or change of a license is thus a significant event in the community, as it establishes a basic governance principle.

[2]Each project has a repository—the most basic element in GitHub—which contains all project files and stores each file's revision history. Repositories can have multiple collaborators and can be public or private.

In the absence of an employment contract, the owner has no formal authority over collaborators or other contributors. Further, the owner has limited power to restrict access to the software code. Other participants (neither owners nor collaborators) hold various rights ranging from pushing, merging, and closing pull requests for code to creating and editing releases of the software. Participants may exit the project or even "fork" it (i.e., take a copy of source code and start a new project to produce a separate software package), thereby posing a restraint to autocratic behavior by project owners (Kane et al., 2014; Raymond, 1999).

The GitHub online platform has a discussion page where community members can express their opinions or dissatisfaction (e.g., with software performance or project direction) and seek solutions (Bissyande et al., 2013). These discussions provide an opportunity to examine various topics important to the functioning of online communities, ranging from operational tasks such as bug fixes to major governance decisions such as license choices.

Among various discussions, license decisions tend to generate broad participation, conflicting views, and heated debates (Alspaugh, Scacchi, & Asuncion, 2010; DiBona, Ockman, & Stone, 1999). Because it is almost impossible for a single developer to be familiar with the plethora of licenses available, license decisions are among the most difficult governance decisions to make (Almeida, Murphy, Wilson, & Hoye, 2017). Even when developers know the technical aspects of a license, they may not fully comprehend its intellectual property implications. Alspaugh et al. (2010) report that developers often unintentionally write software components under different licenses that legally conflict with each other, or are technically incompatible with the overall software architecture. With such incompatibility, components cannot be merged into a software product unless there is a fundamental change in the license regime.

Given the importance of license decisions (Singh & Phelps, 2013) and the high potential for discussion around them to produce disputes in online communities, we use license discussions in OSS projects as a useful window into dispute resolution on core governance principles in communities.

3.2 | Sample and data

GitHub has a built-in issue tracking system. An issue opens with a comment, proceeds with a discussion, and ends with an explicit or implicit closure. Issue tracking offers an unprecedented opportunity to map a dispute resolution process from start to finish. We used the GitHub issue search function within the Application Programming Interface (API) to identify all projects that contain issues with "license" in its title.

More than two thirds (71.6%) of GitHub's 35 million projects have only a single participant (its owner) (Kalliamvakou et al., 2016) and thus are unable to generate any discussion. Most of the remaining projects either do not have a complete discussion history available or do not require a license discussion. We therefore focus on multi-member projects that are recent (i.e., less than 5 years old) and independent (i.e., without a parent project), for two reasons. First, more recent projects are more likely to have a complete archival record of license discussions (see St. Laurent, 2004), which enables us to observe dispute (non-)resolution. Second, independent projects do not inherit license decisions from a parent project or those established by a legacy owner, potentially requiring a license discussion. The search for issues with the term "license" in its title provided us with about 80,000 issues, but less than 1,000 satisfied the selection criteria (multi-member, recent, independent). Among the remaining issues, a manual check for dispute episodes resulted in the identification of 183 issues from 183 projects, totaling 1998 discussion participants.

Projects in our sample were intended for a variety of applications, and as a result, the projects' code bases vary in complexity and volume, ranging from only 4 software commits to 28,887 commits (average = 2,577.09; s.d. = 4,863.49). At the time when we collected our data, projects had been in existence for an average of 1,162.05 (s.d. = 590.56) days, with the youngest project being 32 days old, and the oldest 3,806 days old. Including projects of different types, commit levels, and ages improved the robustness of our inductive theorizing.

The 183 discussions in our sample comprised a total of 983 web-format pages with a mean of 5.4 pages per discussion. An average discussion contained 26.54 comments (s.d. = 28.40), with 1,566.28 total words (s.d. = 2,379.77), from 11 participants (s.d. = 12.19).

3.3 | Analytical strategy

We build theory using algorithm-supported induction, where machine learning (ML) algorithms in conjunction with human interpretations yield robust and interpretable conclusions about patterns in data (Puranam et al., 2020). Specifically, we follow a multistep analytical strategy as presented in Figure 1. In Step 1, we develop a basic understanding of the phenomenon from manually analyzing eight cases, specifying variables that are potentially relevant to governance dispute resolution in online communities. In Step 2, we establish robust associations among variables, or the "stylized facts" pertaining to the entire sample, using ML algorithms. In Step 3, to explain the patterns detected in Step 2, we conduct detailed textual analysis of disputes in a random sub-sample (one third) of our data.

4 | RESEARCH PROCEDURE

4.1 | Step 1: Understanding the phenomenon

We first carefully examine an initial (randomly selected) set of eight discussions. Each of the four authors independently read two discussions in detail. Next, we swapped the discussions and repeated the first step until each author had read all eight discussions. We then shared our interpretations and reviewed the commonalities and differences among discussions. In addition, we consulted documents on OSS development and software licenses to understand the technological dimensions involved in the license decisions.

This process of reading, discussion, and consultation uncovered 11 variables (i.e., issue type, project age, project size, admin group size, discussion group size, initiator status, initiator contribution, polarization of opinions, imbalance in participation, information intervention, and procedural intervention), capturing important structural and processual features of disputes over license decisions. Some of these variables (e.g., project age and size) were measured by information directly available on GitHub, whereas others (e.g., participants' interventions) were coded in the spirit of grounded theory (Strauss & Corbin, 1990). The definitions and descriptive statistics of these variables are presented in Appendices I and II. While coding these variables was a laborious yet feasible manual task, detecting patterns across 183 cases is infeasible to do manually. This is perhaps one reason why inductive researchers have typically focused on only a few cases. Our approach allows us to escape this constraint, by using ML algorithms instead of manual analyses to induce patterns of inter-relationships in a much larger sample.

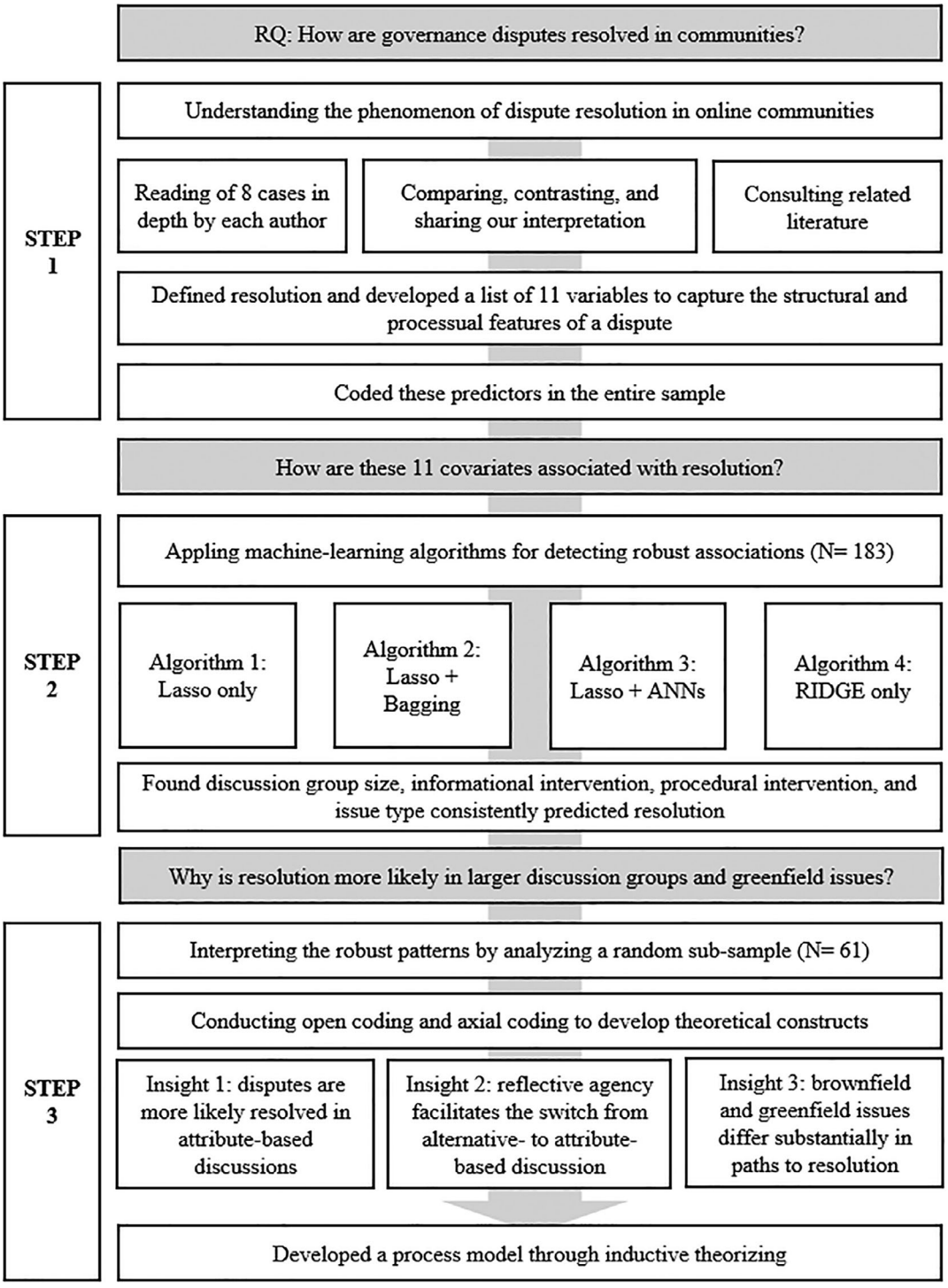

FIGURE 1 Research questions and analytical steps

4.1.1 | Dispute resolution

In our context, *disputes* are conflict or controversy situations where participants advocate different opinions or make distinctive claims about a software license and/or its attributes. The resolution of a dispute over a license requires an explicit decision (typically recorded by the owner) to (a) adopt or not adopt a license or (b) change or not change the license. The variable *Resolution* is coded as "1" when an explicit decision is taken, either to retain the current IP status (current license or owner's copyright status) or change the current IP status (switch to or adopt a new license). In all such cases in our sample, we found no explicit dissent on the recorded decision among the majority of participants in the discussion. *No resolution* therefore implies that no decision is taken due to failure to agree on a decision either to adopt a software license or to change the existing license.

4.1.2 | Context of the dispute

Issue type

We found that disputes targeted two types of license issues: the adoption of a first license, which we term a "greenfield issue," and the revision of an existing license, which we term a "brownfield issue." The variable *issue type* captures this distinction and was coded based on reading all 183 cases manually. In our analysis, we examine whether dispute resolution processes are qualitatively different across these types of issues.

Participation

We identified three ways to conceptualize the scale of participation: *project group size*—the number of contributors in the project; *admin group size*—the number of participants who hold some decision rights (i.e., owners, collaborators, and administrators); and *discussion group size*—the number of commenting participants involved in the dispute discussion. Larger groups involved in a dispute may indicate the more contentious nature of the problem, and also make it harder to reach an agreement that is acceptable to all (March & Olsen, 1976; Olson, 1965; Ostrom, 1990).

4.1.3 | Characteristics of the dispute

Initiator characteristics

Whether the initiator of the discussion was a project owner or not, and the amount of source code contributed by the initiator (contribution), could shape the amount of attention and support she receives from other participants. From prior research on operational disputes in online communities, we believe these variables may signify the competence-based status of the initiator (Dahlander & O'Mahony, 2011; Klapper & Reitzig, 2018). We obtained relevant information on the initiator's characteristics from the GitHub interface.

Initial polarization of opinions

Discussion groups could vary in terms of how different their participants' opinions are on licenses initially (support for alternatives). The *initial polarization of opinions* could, for

example, make convergence more difficult (Maciejovsky & Budescu, 2013). We measured this by reading all 183 cases and coding the diversity of opinions in the initial stages of the discussion.

Project age

We measured project age by the days elapsed since initiation of the project and the time the license discussion arose. Project age is a common correlate to many aspects of online communities (Grewal, Lilien, & Mallapragada, 2006). In particular, project age may be related to the extent participants in the community feel a need to settle on a license in order to safeguard project success (Subramaniam et al., 2009).

4.1.4 | Resolution process

Interventions

Participants could intervene in the dispute resolution process either by providing relevant information (e.g., answering a specific question about licensing) or by influencing the procedures of the resolution process (e.g., initiating a vote, adding a member). These interventions were hand-coded by reading all 183 cases.

Imbalance in participation

While certain participants are very active and could sometimes dominate a discussion, others are relatively silent and only comment occasionally during the entire course of the discussion. Imbalance in participation has been shown to be an important parameter in the life of online communities (Arazy, Nov, Patterson, & Yeo, 2011; Shah, 2006). We computed this using the Gini coefficient of participation across all discussion participants.

4.2 | Step 2: Establishing robust patterns

We utilize a core function of machine learning (ML)—the selection of important predictor variables (also known as *feature selection*) for removing variables that are redundant (i.e., adding little unique information) or overfitting (i.e., failing to predict across subsets of the data). Because different algorithms implicitly feature different families of functional forms, they may not generate the same predictive model (Abu-Mostafa, Magdon-Ismail, & Lin, 2012). We therefore searched for variables that are consistently identified across different algorithms and subsamples, and appear in models that make good out-of-sample predictions. Specifically, we chose LASSO (least absolute shrinkage and selection operator), its variants with bootstrap aggregation, LASSO in combination with artificial neural networks (similar to the approach of Zhou & Jiang, 2003), and ridge regression (Tibshirani, 1996). These algorithms are very useful for selecting the most important variables that have robust associations with the outcome of interest. More details on these algorithms are available in Appendix III.

Our approach holds three key advantages over using standard statistical tools such as regressions to identify associations that seem statistically significant (e.g., drawing on linear regression or comparison of means and cross-tabs; see Glaser, 2008). First, when searching for robust patterns rather than testing hypotheses, tests of significance can be misleading (e.g., significant relationships need not be important predictors). Second, the associations established through

regression models that fit the data may be sample specific (i.e., relationships need not be robust beyond the sample, the primary challenge created by "p-hacking"). Techniques for assessing the robustness of findings to sampling error are necessary for an exercise such as ours. Third, regression models or correlations assume particular functional forms (outcomes or likelihoods of outcomes are modeled via linear functions of covariates), which may become too restrictive if the study aims to inductively build theory.

4.3 | Step 3: Interpreting and explaining the patterns

While ML algorithms generate a predictive model based on the robust associations detected in the data, such a model by itself is not an explanation (Mullainathan & Spiess, 2017). In the final analytical step, we explore more detailed qualitative data and reason abductively (Bamberger, 2018; Peirce, 1878) to provide internally consistent explanations for the patterns detected in Step 2. To this end, we created a random sub-sample of 61 cases for textual analysis and explanation, aimed at developing an in-depth understanding of the processes. Taking a grounded theory approach (Strauss & Corbin, 1990), two researchers independently coded these cases using Nvivo. This round of qualitative analysis generated 27 open codes (Strauss, 1987) of which 7 codes (Table 1) capture the dispute resolution process, 10 codes (Table 2) capture the interventions exercised by participants, and 10 codes (Appendix IV) capture license attributes.

Next, we conducted axial coding (Strauss & Corbin, 1990) and grouped open codes into broader themes. These themes reflect the temporal order in which a dispute unfolds (i.e., genesis, dispute process, and resolution outcome). Further, we analyze the theoretical relationships among these broader themes and position them in the causal link accordingly. Our interpretations, generated abductively through iterations with data and literature where necessary, gave rise to a process model of governance dispute resolution in online communities.

5 | FINDINGS

In what follows, we describe what we found at each stage and the questions our findings further engendered in the sequence they occurred. Our objective is to enable readers to understand (and replicate) the process of our investigation.

5.1 | Identifying robust patterns in the data

Using a combination of ML algorithms (detailed results are presented in Appendix V), we were able to identify models that (i.e., with ANN plus LASSO) had between 72.24% (F1 score = 0.072 and MCC = 0.43) and 75.40% (F1 score = 0.74 and MCC = 0.52) accuracy.

The pattern of robust association we found is striking, both in what it includes and excludes. Consistently across all algorithms, four variables—discussion group size, informational intervention, procedural intervention, and issue type (i.e., greenfield vs. brownfield)—emerged as robust predictors of resolution. Project age, project group size, admin group size, initiator status, initiator contribution, initial polarization of opinions, and imbalance in participation turned out *not* to be important in predicting resolution or nonresolution. Recall that our initial

TABLE 1 Dispute resolution process

Aggregate themes	Open codes with examples of representative quotes
Dispute genesis	
• Disputes on alternatives	**Expressing clear preferences over alternatives** "I've always been a fan of MIT. I don't know what we should be thinking about legally here, but it's my go-to license and I love seeing GitHub open source software licensed under it." (A participant in project *Atom*)
	Expressing indifference between alternatives "I'm fine with all 4 of those licenses (LGPL, Apache2, BSD, MIT) if they're necessary to get wire onto workshop." (A participant in project *Wire*)
• Disputes on attributes	**Disputing attributes via dialogue** "The Ms-Pl is incompatible with many free software licenses, as it imposes a unique restriction: (D). If you distribute any portion of the software in source code form, you may do so only under this license by including a complete copy of this license with your distribution..." "There are many GPL-, LGPL-, and AGPL-licensed libraries, yet we cannot legally combine these with DotNetZip due to the terms of the Ms-Pl." "Dual-licensing under Apache 2 or MIT would resolve the problem. As GitHub notifies people who are mentioned, re-licensing would involve pasting the list of contributors into this thread and asking for a release of their copyright. Getting approval can take anywhere from a few days to several months, depending upon contributor awareness. We would also need to get approval from @DinoChiesa for the original code on SVN." (Three participants in project *DotNetZip.Semverd*)
	Disputing attributes by increasing the search space "I personally recommend strongly staying away from AGPL though, and that's not without a lot of fair reasons. The best license for an app that has 'Open Source' at the heart of its feature set, should definitely consider truly open licenses like the BSD or MIT or even Apache 2.0." (A participant in project *Laverna*)
Resolution process	
• Bargaining in alternative space	**Remaining in alternative space by not defending options** "CC-BY SGTM." "I'm happy with CC-BY." (Two participants in project *Presentations*)
	Remaining in alternative space through failure to switch to attributes "@Riamse, care to expand on your 'no,' please?" "No. Just close the issue, you're never going to change my mind." (Two participants in project *Afch*)
• Problem solving in attribute space	**Collectively solving issues of license fit for project** "I think that you guys had misunderstood AGPL. AGPL protect[s] the source code server over the network, like webpages (that's why diaspora, wordpress, is AGPL), but telegram-bot doesn't serve the source code over the network, [it] is just running like if it were a binary, so GPL protect[s] it exactly the same ... [as] AGPL. Even more, you can't communicate telegram-bot over the network. It's tg who communicates with telegram-bot, and the telegram servers with tg. Check this and this." (A participant in project *Telegram-bot*)
	Building a common understanding of license specifications "@Scotchester, that's a great question. Luckily, I don't think we have to worry about it. All external dependencies are pulled in after checkout: JavaScript

TABLE 1 (Continued)

Aggregate themes	Open codes with examples of representative quotes
	and CSS via Bower, and Java and Clojure via Leiningen. The repo should have no external code in it currently. Given @NoahKunin's comments and @tauberer's document, I'm convinced. Let's take @marcesher's advice and wait until @virtix can comment, but this pull request has my +1." (A participant in project *Qu*)
• Switch	**The point where the focus of a discussion shifts from alternatives to attributes** "Yes, a dual license does not need to be compatible, but I was pointing out that if the default license was CC0, CC-BY, or CC-BY-SA, then it would not matter that a user added the CC-BY-ND as alternate license. The default, freeer license would already grant more rights and there would be no way for users to force the non-default license. I'm against CC-BY-ND as the default because most people won't change it and my opinion is that we should promote free culture here. [Segue to next debate point.]... "In summary, Freeeeeeeeeeeeeeedooooooooooooooommmmmm." (*A participant* in project *terms of serive_206*) "Yes, but basically you don't expect any reward from dishonest people and honest people won't hesitate rewarding you if they succeed. This is proven to work by many open source projects, like Node.js itself for instance. I agree with Alfredo and would recommend following the best practices of the open source world by choosing either MIT or Apache. The more limitations are imposed on a project, the less likely it is that people (and companies, which is important) keep contributing to it over time." (A participant in project *jsfbp_3*)
Resolution outcome	
• No resolution	**Failing to reach a resolution in brownfield** "If we can't get a sign-off, I don't think relicensing is possible." (A participant in project *DotNetZip.Semverd*) **Failing to reach a resolution in greenfield** "As I read it currently, even using the modern.IE images as they are clearly intended by MS is not compatible with their own licensing. Until all of that gets sorted out, I can't license ievms in a traditional sense. I hope you understand." (A participant in project *Ievms*)
• Resolution	**Reaching a resolution in brownfield** "I have made a pass at replacing the LGPL license with BSD throughout the codebase, and I'm going to mark this fixed. If there's any additional changes needed to migrate, file new issues for those please." (A participant in project *Ffi*) **Reaching a resolution in greenfield** "Thanks for bringing this to my attention. I wasn't aware that certain countries had such restrictive default licenses. I've added a license file (WTFPL) so people know where they stand." (A participant in project *Awesome-php*)

reading of eight cases led us to identify a list of 11 variables. If we had stopped at Step 1, we might have concluded that all those variables were relevant for explaining dispute resolution in online communities. However, in the full sample not all variables identified in Step 1 turned

TABLE 2 Interventions during a dispute process

Actions	Specific activities	Illustrative quotes
Informational intervention	Advocate the merit of an attribute or an alternative	"[O]ne reason it's important for code like this to be truly open source: Someone—The good guys—Could use it to simulate various voter registration fraud scenarios in a controlled environment, in order to come up with ways to detect and defeat such fraud. One never knows the unexpected but legitimate uses to which code might be put; therefore one never knows in advance the full effect of seemingly innocuous restrictions." (*Kfogel* in project *Voter-registration*) "Apache license includes permission to use patented material freely, and really spells out the BSD license. That's why I like it. Plus, we can change the license as we want without the silliness with the GPL." (A participant in project *Wire*)
	Reveal personal preference of an attribute or an alternative	"I'm in favor of free (as in speech), so the most permissive possible:)" (*Patcon* in project *World Citizenship*) "My requirements for a license are just two things: 1. I want as many people to use my code as possible. I want as few barriers as possible, both legal and mental. 2. I want it to be as clear as possible to developers who are not lawyers that the above is true." (*Natefinch* in project *Hugo*)
	Volunteer information about an attribute or an alternative	"Nothing about a project being GPL has any bearing on what license I as a contributor use for my contributions. The GPL 2 states: You may not copy, modify, sublicense, or distribute the Program except as expressly provided under the License." (M*ahkoh* in project *Neovim*) "Here's a great example of how project-open-data handled their license migration: project-opendata/project-open-data.github.io#135." (*Willnorris* in project *Hugo*) "Just for tracking purpose: question for Timotheus' code relicense is at http://en.wikipedia.org/wiki/User_talk:Timotheus_Canens#AFCH_relicensing" (*Wikipedia-mabdul* in project *Afch*)
	Ask questions and provide answers about an attribute or an alternative	"Is that compatible with the license for our resources?" (*LB* in project *ChessPlusPlus*) "@mlubin: Yes, we should add a clause to the site license saying that all content is MIT unless otherwise states, in which case it is licensed under whatever license is indicated." (*StefanKarpinski* in project *Julialang*)
Procedural interventions	Include a participant	"@wikipedia-mabdul—Still awaiting your! vote." (*Nathan2055* in project *Afch*)

TABLE 2 (Continued)

Actions	Specific activities	Illustrative quotes
	Exclude a participant	"@albanpeignier's 1 contribution was merged post-LGPL, but technically written against the BSD license. Can probably scratch that off the list." (*stouset* in project *Ffi*)
	Conduct a vote	"Unless someone disagrees, I'll create an issue in the next hours pinging everyone that has contributed to this project to ask for their permission to add an MIT license to this project." (*Bpinto* in project *Oh-my-fish*) "At this point, each member of the education committee should make an informed decision and submit their opinion on the following Doodle poll. Please write your full name to the left of your choice. The poll will be open until Saturday, July 19, 11 am Eastern (5 pm Central European time)." (*Nikosbentenitis* in project *Btcfoundationed*)
	Make a unilateral decision	"Apache 2 license it is. Sorry about not waiting too long, everyone, but we would like to get the ball rolling." (*Anticept* in project *Wire*) "The poll for the license of our materials is now closed. See the votes here. All of those who voted chose the MIT license for the materials. I modified the MIT license based on the license used in the Foundation's bylaws and posted it here." (*Nikosbentenitis* in project *Btcfoundationed*) "Update: Let's hold off for a little while. I have a call in to somebody more knowledgeable." (*Kleinmatic* in project *Tabula*)
	Close a discussion	"Just close the issue, you're never going to change my mind." (*Riamse* in project *Afch*)
	Re-open a discussion	"johanneswilm reopened this on Feb 21, 2014." (Project *diffDOM*)

out to be robust predictors of resolution. Put differently, Step 2 helped us remove the redundant variables, avoid overfitting our explanations to the selected sample of eight cases, and allow the subsequent analysis to focus on the robust and important set of variables in our context.

Further, among the variables identified as having a robust association with resolution, not all results are intuitive. In particular, discussion group size has a positive relationship with dispute resolution (i.e., discussions with more participants are systematically *more* likely to reach resolution). In most bargaining and conflictual situations, large size makes it *harder* for a group to reach a resolution (Olson, 1965; Ostrom, 1995; Roth, Murnighan, & Schoumaker, 1988). A possible explanation for our counterintuitive finding about group size is

that a large number of discussants reflects the importance of a project, thereby creating a stronger intent among participants to resolve intellectual property disputes or willingness to defer to others' opinions. However, if this reasoning were accurate, the variables project group size or administrator group size should have appeared as robust predictors. Because project and administrator group sizes were not robust predictors, we ruled out this alternative explanation.

5.2 | Understanding the antecedents of resolution

To explain the causal structure in the pattern of relationships among the variables we uncovered through machine learning, we next conducted in-depth qualitative analysis of the governance disputes.[3] We took a random sample comprising one third of the total cases ($N = 61$) for detailed examination, using open and axial coding. In terms of pattern detection, this sub-sample is representative of the full sample, since it is a random sub-sample. We also verified that the ML analysis revealed the same key predictors of resolution in this sub-sample.

5.2.1 | Attribute-based discussions are resolved more often than alternative-based discussions

After coding the textual data of the sub-sample 61 cases, we found that a governance dispute could be initiated around one of the two themes—license alternatives or attributes of these alternatives. Commonly discussed license alternatives included MIT, BSD, LGPL, and GPL. Commonly discussed license attributes included permissiveness, protectiveness, compatibility, and ease of contribution by developers. Various license alternatives differed from each other on at least one of these attributes. Appendix IV presents the license attributes and alternatives discussed in our sample.

In the following excerpt from project *Elegant-mind*, one owner of the project started a discussion and suggested changing the current license from GPL to LGPL. Another owner joined and suggested that MIT might be a better alternative and the debate went on. This case is an example of an alternative-based discussion and it did not reach a resolution.

> *Sjstoelting*: I would like to start a license discussion. Evolution itself is licensed with GPL2, which is OK so far. But I think we do better, if we take LGPL for the core, because the core is a library and is used as a library (framework). Is it OK for you that we publish our enhancements under LGPL instead of GPL?
>
> *Yama*: I think MIT is best, but is it difficult? If it is so, I think that LGPL is better.

Conversely, project Vis.js illustrates an attribute-based discussion, in which the dispute was resolved by adopting a dual license of Apache and MIT. One owner of this project states:

[3]We also conducted fuzzy set QCA with the key predictors (reported in Appendix VI). The results are consistent with the insights from the content analysis reported here.

> *Josedejong*: Currently vis.js is licensed under the Apache 2.0 License. This license is incompatible with, for example, the GPLv2 licenses. I'm not sure how big this issue really is. But in order to better serve our users, we are considering whether it is useful to change the license. A few things are important for us: 1. as much as possible freedom for the user; 2. protection for the user (like against patent threats or changing the license conditions later on); 3. keep attribution (honor the people and our company investing so much in this library). While MIT is a nice fit for (1), Apache 2.0 better serves (2) and (3). We could choose one of the two, or even choose a dual license allowing both. Do you have any opinion on the licensing of vis.js?

This distinction between attribute- and alternative-focused discussions did not arise from the application of machine learning; rather it resulted from qualitative coding the textual data. The distinction turned out to be critical, because disputes that did not reach resolution (14 of 61 cases) consistently focused on license alternatives. Among these 14 unresolved governance disputes, 12 started with discussing alternatives. Although these discussions touched on attributes at some point, they failed to completely turn the focus to attributes. In contrast, in the majority of disputes that reached a resolution (34 out of 47), participants either began with discussing attributes (16) and focused on them throughout or switched the focus of discussion from alternatives to attributes (18). When alternative-based discussions did achieve resolution (in 13 cases), it was through procedural interventions, either by a unilateral decision made by the owner (10) or by a voting procedure suggested by the participants (3). *Our first interpretive insight is that disputes are more likely to be resolved when discussions focus on the license attributes, instead of on license alternatives.*

To triangulate this insight against the patterns we had discovered through ML, we also assessed whether the predictors of resolution were positively associated with an attribute-based discussion. Since we had data on attribute-versus alternative-based discussions only in 61 cases, we created 1,000 bootstrap samples of these cases and ran the same logistic regression model in each sample, with discussion group size, issue type, procedural intervention, and information intervention as predictors of an attribute-focused discussion. This procedure is known as bootstrap aggregation or bagging (Abu-Mostafa et al., 2012).

The results show that in 815 of the 1,000 samples, discussion group size had a positive effect on a discussion being attribute focused (with an average coefficient of 0.043). Similarly, in 1,000 out of the 1,000 samples, brownfield issue type had a negative effect on a discussion being attribute focused (with an average coefficient of −1.175). These results strongly suggest that attribute-focused discussion is the mechanism that connects the two predictors of resolution we found through ML (i.e., discussion group size and issue type) and resolution outcome. Additionally, informational intervention had a consistently negative (822 of 1,000 samples) but small effect on attribute-focused discussions (−0.02), indicating that more such interventions inevitably arise as the discussion progresses without resolution. Procedural intervention had a consistently positive (716 of 1,000 samples) effect on attribute-focused discussions (0.08). It is also likely that informational interventions complicate the discussion and create further divergence whereas procedural interventions simplify the discussion and facilitate convergence.

Next, we qualitatively analyzed the discussion data with the aim to understand why larger discussion groups or greenfield projects might systematically be more likely to feature attribute-based discussions around licenses.

5.2.2 | Switching from alternative-based to attribute-based discussions: The role of reflective agency

We found it noteworthy that even those governance disputes that start with discussing alternatives, at some point may undergo an important change toward attributes. We define this critical point where the focus of a dispute pivots from license alternatives to attributes as a "switch." The following snippet illustrates such a switch:

> *Trustmaster*: "I agree with Alfredo and would recommend following the best practices of the open source world by choosing either MIT or Apache. The more limitations are imposed on a project, the less likely it is that people (and companies, which is important) keep contributing to it over time." (Project *Jsfbp*)

In the above example, the discussion pivoted toward reflecting on what attribute is really important for their communities (e.g., permissiveness or little constraint). Once the desired attribute became clearer to the participants, the process looked less like bargaining in a search space defined by alternatives, and more like collective problem solving in a search space defined by attributes. The dispute in project *Jsfbp* was eventually resolved by adopting the MIT license.

A focus on the attributes allows the participants to learn about their individual preferences regarding attributes, discover the convergence (or the lack of it) of preferences, and search for one or more alternatives that satisfy those preferences. Therefore, we conceptualize this mode of dispute resolution as "problem solving." In contrast, when focusing on alternatives, disputants engage in bargaining over the merits or drawbacks of different alternatives. Accordingly, this mode of governance dispute resolution can be conceptualized as "bargaining." In the following paragraphs, we show in detail how these two modes of dispute resolution distinctly unfold.

We found that participants in some disputes anchored their advocacy of a preferred alternative on the group's preferred attribute in order to gain support. For instance, a contributor *kripken* in project *Servo* advocated the MIT license as a feasible alternative because of its permissiveness—an attribute that later turned out to be preferred by most participants in the group. After an extensive discussion a resolution was reached—the main repository of *Servo* was licensed under MPL2, and the rest under a dual license of Apache and MIT.

> *Kripken*: There could be a big benefit to using the MIT license as this would be the only browser available under a permissive license (WebKit is LGPL, Gecko is MPL, others proprietary). Even without debating the benefits of permissive licensing, there is a benefit to having different licensing than existing browsers to fill a different need and attract new users.

Further, we found that focusing on attributes enabled resolution by making participants more open to being persuaded. For example, in project *Telegram-bot*, the discussion started as a debate over alternatives between the owner *Yagop* and a contributor *Asdofindia*.

> *Asdofindia*: AGPL fixes a loophole in GPL license that renders GPL useless when it comes to telegram-bot... That's why Diaspora, Wordpress, etc. use AGPL. I think AGPL is the best for our telegram-bot too. Since Lua isn't compiled to binary, GPL offers no protection at all, at the moment.

> *Yagop*: I want to keep GPL, at least at the moment.
>
> *Asdofindia*: I'm just telling you that you may as well release this in public domain because it gives the same protection that GPL does to your hard work—no protection...

Fortunately, a collaborator *Rockneurotiko* pointed out that "I think that you guys misunderstood AGPL. AGPL protects the source code server over the network, like webpages, but telegram-bot doesn't serve the source code over the network, it is just running as if it were a binary, so GPL protects it exactly the same way as AGPL" and established code protectiveness as a preferred attribute among disputants. From that point on, the discussion turned into a collective problem-solving process and participants converged toward AGPL license.

The qualitative analysis of the textual data revealed a second insight: *the switch from alterative-based discussion to attribute-based discussion results from a set of actions that we conceptualize as "reflective agency."* Specifically, we define reflective agency as social actions that are rooted in critical interpretation and mindful reflection and that can be exercised by one or more individuals. In our sample, we found two different yet interrelated forms of reflection: challenging or re-examination of an underlying attribute of a particular license and analyzing and comparing underlying attributes across different license alternatives. The following snippet illustrates how a participant challenged the current understanding of GPL:

> *Hopeseekr*: Man. Whomever thinks the GPL is "open source" needs to get their heads checked. It's like saying Communism's forced appropriation of everyone's creations is the ultimate form of "Charity": D (Project *Utilphp*, resolved, decided against GPL)

On the other hand, the quote below shows how a participant compared the underlying attributes across different license alternatives and how such a comparison revealed commonality and differences across licenses:

> *Josdejong*: In order to better serve our users, we are considering whether it is useful to change the license. A few things are important for us: 1. as much as possible freedom for the user, 2. protection for the user (like against patent threats or changing the license conditions later on), 3. keep attribution (honor the people and our company investing so much in this library). While MIT is a nice fit for (1), Apache 2.0 better serves (2) and (3). (Project *Vis*, resolved, adopted a dual license of Apache and MIT)

These reflections help to reframe the issue at hand by finding the underlying attributes that are at the heart of the community's governance principles. For example, in project *Agar.io-clone*, after the owner *Huytd* expressed his confusion, developer *Ariamiro* summarized compatibility and simplicity as the attributes that differentiated the large subset of alternatives discussed, substantially reducing the complexity of comparing and contrasting license alternatives. This dispute was soon resolved by adopting MIT as the final license.

> *Huytd*: Actually, for me, this product is 100% free and open for everyone, you can use it in any purpose. I don't know if there is any license to match this. And how about what other contributors think?
>
> *Ariamiro*: If you want people using the game for any purpose choose Public Domain; if you want a good and simple license choose MIT. I don't recommend that you continue without a license.

We did not find any evidence that status (e.g., project owner or contributor) mattered for the expression of reflective agency. Instead, we observed that any participant engaged in such reflections could transform the discussion from bargaining to problem solving, by challenging and analyzing the current state of discussion. Our concept of reflective agency is consistent with Dewey's work that emphasizes the role of individual actions in transforming public communication (Dewey, 1922).

Online communities attract people with a vast array of expertise and backgrounds (Wang et al., 2015). This diversity creates two collaboration challenges in license decisions: First, members may not agree on which license alternative is more desirable. Second, members have different understandings of license alternatives, and thus may find it difficult to find the alternative that scores the best on the commonly desired attributes. When disputes arise, reflective agency solves the first problem by reframing the discussion from bargaining over which alternative is better into comparing the alternatives in terms of various attributes. It then tackles the second challenge by reflecting on the discussions and generating a shared understanding of the different alternatives. When this understanding is collectively accepted by group members, it creates a common ground for agreement (Bechky, 2003; Leonardi, 2015). Coupled with alignment of preferences over attributes, the result is a successful resolution.

Note that although switches make the functioning of reflective agency more visible, it is quite likely the discussions that already began with a focus on attributes are just as much an instantiation of reflective agency. Some participants may be able to conceive of a discussion as being more likely to be productive if it can be initially focused on attributes.

5.2.3 | Reflective agency in different issue configurations

The value of procedural interventions is intuitive, because these interventions resemble mediation in an ADR system, although interventions in our context do not need to be carried out by individuals who hold specific roles. The negative association between informational interventions and resolution is understandable once we realize that the longer the discussion remains unresolved, the more information is likely be introduced by participants. While procedural interventions simplify the discussion and facilitate convergence, informational interventions may also complicate the discussion and create further divergence. However, why resolution should be easier in greenfield issues and larger discussion groups is less clear, particularly since reflective agency is not specific to project owners (i.e., contributors and regular participants also exercise this function). To answer this question, we conducted further qualitative analyses (Appendix VII), comparing greenfield versus brownfield issues and large ($N > 11$, which is the mean size) versus small discussion groups ($N < 11$).

This comparison leads to a third insight: *brownfield and greenfield issues differ substantially in their possible paths to resolution*. We found that for a project to preserve the integrity of

software written under an existing license, changes in license (i.e., brownfield issues) can only be made to a new license that is backward compatible. The requirement of license compatibility restricts the space of available alternatives from which the next license can be selected. In brownfield issues, the existence of a license may itself center the discussion on alternatives to that license, preventing the discussion from switching toward attributes, and eventually from reaching resolution. This is evident in project *Toxcore*, in which participants repeatedly went back to the original license LGPLv3, and were not willing to deviate from it.

In contrast, in greenfield issues the search for a license can proceed more freely along both license attributes and alternatives. For example, in project *Oh-my-fish*, although there were only four participants, they were able to cover a broad range of possible licenses. Were the dispute resolution process a pure bargaining one, the existence of a status quo and the subsequent fixation on it would act as a focal point that promotes convergence to a negotiated outcome (Schelling, 1960). As bargaining is a mixed-motives game involving both elements of conflict and coordination (Schelling, 1960), anchoring by all participants on a common reference point can promote rapid agreement, though not necessarily an integrative one. The absence of such an effect reiterates that a pure bargaining perspective does not offer a good explanation for our data.

The qualitative comparison also allows a more comprehensive explanation of *why a larger discussion group size is advantageous for dispute resolution*. When we examine and compare the dispute processes in depth, it became clear to us that the likelihood that someone displays reflective agency by reframing the discussion around attributes is higher in larger discussion groups. Larger groups such as *WordPress* and *Toxcore* typically featured reflective agency, whereas smaller discussion groups such as *Elegant-mind* and *Drawille* did not.

An explanation for this size effect can be given as follows: consider x as the probability of any individual in a discussion group being able to convincingly frame the discussion in terms of attributes (i.e., a particular individual would exercise reflective agency). Even if x is low, the probability (denoted as p) that at least one individual will do so successfully (i.e., the emergence of reflective agency in the group), which is $1–(1–x)^n$, grows rapidly with group size n. For instance, p is close to 80% even if x is only 30% for $n = 5$. This is a corollary of the Condorcet jury theorem well established in experimental and field research on group problem solving and decision making (Laughlin, 2011; Lorge & Solomon, 1955; Nemeth, 1986; Wittenbaum & Stasser, 1996). Therefore, a problem-solving perspective (Laughlin, 2011) provides a fitting explanation for what we have observed.

In sum, reflective agency makes a discussion more likely to be attribute-based (rather than alternative-based). It more often emerges in greenfield discussions because of unconstrained search space and in large groups because of a higher probability that at least one individual will display reflective agency.

5.2.4 | Alignment of preferences on attributes as a necessary condition

An important necessary condition also emerged from our analysis: the high rates of resolution in governance disputes that are attribute focused are possible only when there is alignment in the community on what the important attributes are. In theory, participants could disagree just as violently over attributes as they would over alternatives. However, our analysis revealed a strong alignment in participants' attribute preferences in most of the projects in our sample.

Our open coding of the textual data revealed that even though the disputes touched upon numerous attributes, only three—permissiveness, protectiveness, and compatibility—remained prominent throughout the sample. Further, among the discussions in which permissiveness was raised, 31 discussion groups subsequently adopted a permissive license (e.g., MIT and Apache).

For those familiar with OSS development, this strong alignment of preferences over attributes is hardly surprising. Participants join OSS communities in part because they value sharing and collaboration (Belenzon & Schankerman, 2015; Ren, Chen, & Riedl, 2016; Shah, 2006). Individuals embracing an open source ideology self-select into such online communities (Stewart & Gosain, 2006). Therefore, not only do participants in this context have alignment of preferences on attributes, but they are also likely to be aware of the social norms in the communities that conduct such projects (Stewart & Gosain, 2006). Self-selection by participants thus ensures that within a specific project, they are likely to find at least some common ground around certain attributes. The following snippets provide evidence that within project *Btcfoundation*, participants converge on a fundamental value—openness. The participants eventually opted for the MIT license and resolved the dispute.

> *Pmlaw*: It's not about being religious, just read the licenses: MIT is more open than GNU/GPL. GNU/GPL imposes your view of the world on every downstream user. Of course, the committee is free to make whatever decision it likes, I was asked for my opinion and I offered it.

> *Johnmees*: I'm not interested in a license that forces a similar openness on future users, or demands attribution. I think the most open license possible would be best. From what has been presented so far, it sounds like the MIT license is our best bet. I recommend we go with that.

With strong preference alignment, attribute-based discussions could then proceed to resolve the governance disputes. However, when such alignment is absent, attribute-based discussions did not lead to resolution. For example, in project *Toxcore*, although the license dispute started off with a focus on the attributes (i.e., compatibility and freedom), the participants were rapidly divided into two camps, with one strongly favoring compatibility for maximizing market share, and the other favoring permissiveness (freedom) over commercial success. As a result, no resolution could be reached:

> *TheAustinHowell*: Free software is free software; I don't care what Stallman says. Ensuring the privacy of possibly hundreds of millions of people is more important than the license. To do this we should have a license that ensures maximum market share.

> *TrevorDorl*: All I can see LGPL doing for this software is creating non-free user interfaces for it, which just doesn't make sense. And frankly I think non-free interfaces to Tox is kind of a frightening idea, just a bad idea in general.

However, the alignment of preference on attributes is not a sufficient condition for resolution. Unless an attribute focused discussion is triggered through reflective agency, the alignment may not even be discovered, let alone useful. In the next section, we condense the causal

logic underpinning our findings into an inductively derived theoretical model of the dispute resolution process in online communities.

6 | EMERGING THEORY

Given our findings, we develop a theoretical model integrating group process and group property to explain the resolution of disputes pertaining to core governance principles in online communities. Our central argument is that disputes that take the form of collective problem solving in attribute space enjoy a higher likelihood of resolution than those that resemble bargaining in the alternative space. This is conditional on the alignment of participants' preferences over attributes—an alignment grounded on participants' self-selection into the OSS community.

6.1 | Group process: The importance of reflective agency

Our emerging theory stipulates that governance disputes in communities are resolved through two distinct group processes: problem solving and bargaining (see Figure 2). When a dispute discussion focuses on the attributes from the onset or manages to switch the focus to attributes at some point of the process, it resembles problem solving and is more likely to attain resolution. This is facilitated by "reflective agency," whereby one or more participant(s) reflects on the problem underlying the (potential) dispute and recognizes a productive path forward. Reflective agency steers participants in disputes away from advocating one over another alternative, to searching for an alternative that satisfies commonly preferred attributes.

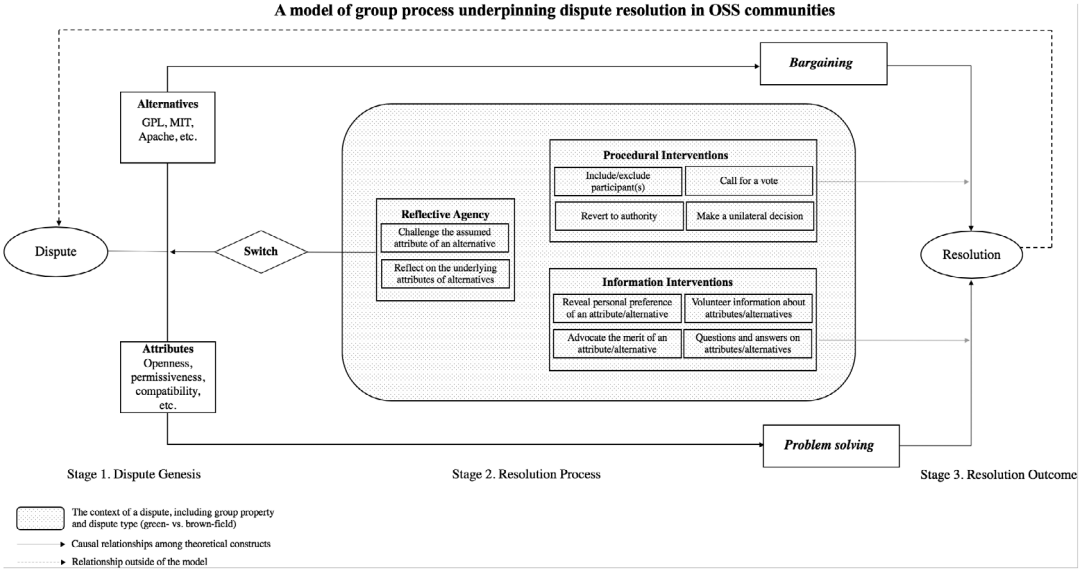

FIGURE 2 A model of group process underpinning dispute resolution in OSS communities

6.2 | Group property: The importance of preference alignment

Reflective agency by itself does not ensure resolution. A second necessary condition for resolution is that preferences over governance principle attributes are largely shared among discussion participants. In OSS communities, the salient role of self-selection (i.e., voluntary joining) makes this condition possible. In firms, the alignment of preferences on key attributes may also emerge through sorting—either in attention to such preferences for hiring new employees or allowing employees to self-select into projects.

7 | DISCUSSION

Governance disputes and their resolution are critical in the lives of communities, yet they have been under-studied and under-theorized. Exploiting the availability of complete archival records for discussions in GitHub communities, we studied how disputes vis-à-vis the software license—a core governance principle of these communities—are resolved. We found that when governance disputes in communities bear the signature of group problem-solving processes rather than a multi-party bargaining process, they are more likely to be resolved. However, such a resolution is far from inevitable, and our theory reveals two necessary conditions: the emergence of reflective agency and the alignment of preferences over attributes. Reflective agency is manifested when at least one individual in the group is able to reflect on the problem and transform or initiate the discussion along the lines that are attribute focused (rather than alternative focused). This conceptual "transposition" of the discussion into the problem-solving domain facilitates resolution under conditions where there is a pre-alignment of preferences among participants around key attributes of the governance principles. Although the counterfactual is unobservable, it is logical to expect that in the absence of reflective agency, even if participants have aligned preferences over attributes, they may not discover this important fact and therefore miss the opportunity for resolution.

Our theory has important boundary conditions regarding types of disputes and communities. First, this theory is derived from studying disputes arising from license decisions in governing intellectual property. While such decisions are core to any organization, they represent only one type of governance principle, albeit a central one. It is possible there are other governance principles at stake in online or offline communities, such as the development of procedures on how to resolve operational disputes, for which community members may share little alignment on preferences. Thus, we call for research that explores the process of dispute resolution over other types of governance disputes.

Second, our theory was inductively built from a context (OSS development) in which members, to a large degree, share common preferences because of self-selection (Butler, Bateman, Gray, & Diamant, 2014; Spaeth, von Krogh, & He, 2015; von Krogh et al., 2012) and contribute to the production of a public good (Lerner & Tirole, 2002). Shared preferences such as a commitment to voluntarism, learning, openness, and sharing of code are known to draw volunteers into such online communities (Dahlander & O'Mahony, 2011; Stewart & Gosain, 2006). The high degree of alignment in preferences over attributes is a result of effective sorting (self-selection) mechanisms in online communities. There are at least two reasons why sorting could be more effective in online communities.

First, online communities are not necessarily restricted by a common geography (Mahony & Lakhani, 2011) and thus may enjoy a broader base for sorting than other communities where

those restrictions apply. Second, the transparency resulting from permanent, electronic records (Marlow, Dabbish, & Herbsleb, 2013) creates clear guidelines for individuals to self-select in or out of an online community. In contrast, offline communities rarely have such transparent documentation of community members' preferences. New members often need to learn the encrypted value, norm, and culture through an extended process of socialization. Despite porous borders and freedom to exit, considerable time may elapse between joining and a member's decision to exit (due to incompatible preference) in offline communities. Thus, if one would take a cross-sectional look into an offline and online community at any point in time, the latter will possibly have a higher level of preference alignment.

However, we are cautiously optimistic about the relevance of our results beyond online communities. Preference alignment on attributes may be a broad-based feature of communities, more so than scholars have hitherto imagined. For example, in many of the communities surrounding the common pool resources (CPR) that Ostrom and colleagues (Ostrom, 1990; Wilson et al., 2013) studied, membership at birth together with the process of growing up within these communities must undoubtedly facilitate the internalization of group norms, thereby affecting and to some extent aligning individual preferences. This alignment may both prevent and help resolve governance disputes.

Furthermore, today even traditional (offline) communities—be they communities of practice within a firm, research collaboration, or social movements—rely extensively on digital tools. Many companies have already been using online discussions and virtual collaboration tools (e.g., Slack, Zoom, Monday.com) among employees and with partners or customers. The global pandemic created by COVID-19 has expedited this trend by propelling organizations around the world to adjust rapidly to remote collaboration. The insights generated by our study have implications for the leadership and coordination of all these forms of virtual collaboration.

Despite its limitations, our study contributes to the literature on community governance in general and dispute resolution in online communities in particular. It extends the understanding of dispute resolution in online communities from an operational one (Forte et al., 2009; Klapper & Reitzig, 2018) to a more strategic one. Our theory explains otherwise puzzling observations about OSS communities, such as the rare occurrence of governance disputes and the positive association between discussion group size and resolution. First, we contend that because participants' self-selection into these communities generates a degree of pre-alignment on preferences over attributes, governance disputes are rare in these communities. Second, our theory predicts that when disputes do break out, larger groups may be more likely to see resolution because of a higher probability for reflective agency to emerge. Our theory thus constitutes a starting point for future empirical work on communities operating in online and offline settings, possibly covering a broader range of governance principles that define, bestow, and distribute rights and rewards on community members (Lerner & Tirole, 2002; Ostrom, 1990) beyond intellectual property.

Beyond its theoretical implications, our two-pronged approach advances inductive methods by separating pattern detection and pattern explanation that are traditionally entangled. When detection and interpretation occur simultaneously, two risks arise. First, cognitive constraints limit the number of cases researchers may be able to analyze, thereby raising the risk of overfitting (i.e., the explanation may be valid in the small sample studied but does not extend to other samples). Second, researchers' confirmatory biases and motivated information processing may lead to false positives or false negatives (e.g., seeing patterns that are explainable in an intuitive or interesting manner versus not seeing others that are less intuitive). Our approach takes some steps toward mitigating both risks. Compared to traditional inductive research

methods, our approach can handle a larger number of cases—an improvement that allowed us to be confident that our detected pattern is not a result of overfitting. As a consequence, the resulting insights from our approach should have an advantage in making out-of-sample predictions—the ultimate objective of all (inductive) theorizing.

Further, our approach integrates insights from both quantitative and qualitative analyses, where neither alone could have fully explained the phenomenon of interest. The use of ML algorithms enabled us to effectively detect robust patterns in data. These robust patterns, which formed the basis for our abductive reasoning, were detected algorithmically, and are replicable by others who use the same data and procedures. Subsequently, we interpreted the textual data, narrowing down the cases to examine in depth based on results of the algorithmic analysis. We related our interpretations to existing literature—elements of inductive research that are not replaceable by algorithms—to offer our explanation to this pattern. Our work is among the first empirical studies in the field of strategy to combine ML with more traditional sense-making approaches for theory induction.

Finally, our emerging theory informs the practice of community governance in both online and offline settings. Given the importance of reflective agency we found in our study, an implication is that those who help govern communities, when they cannot rely on traditional hierarchical structures or ADRs to aid dispute resolution, may yet rely on a combination of reflective agency aided by alignment on preferences created through sorting. Rather than count on the chance that reflective agency will emerge, they might recruit individuals with the capacity for reflective agency, identify them (e.g., based on text analysis of prior online interactions), or even train moderators of communities in using such skills.

Similarly, insofar as the presence of a permanent, transparent record of the community's preferences and the absence of geographic constraints facilitate more effective sorting, offline communities or even traditional organizations could benefit from comparable practices. First, these communities and organizations could increase the transparency of the values and preferences shared in an organization or a community. An initial step involves simply better documentation of values and the next step requires more effective communication, which could be achieved by improving internal communications, especially during the on-boarding phase and transmitting clear, authentic signals when communicating externally. Second, these communities and organizations could also resort to online processes for dispute resolution. The full visibility and archiving available in online processes might help to resolve disputes in a manner superior to offline interactions (Marlow et al., 2013).

8 | CONCLUSION

This study explores the question of how communities resolve governance disputes. Gathering data from discussions surrounding software license decisions, we employed a combination of machine learning and qualitative methods to detect and explain patterns in the data. Our findings on how OSS communities settle principles for governing intellectual property led to an emerging theory, stipulating that governance disputes in these communities are resolved through two distinct process—problem solving and bargaining. When a dispute discussion focuses on the license attributes, it resembles problem solving and is more likely to attain resolution. The switch from an alternative-based to attribute-based discussion rests on reflective agency that reframes the discussion by decomposing conflicting opinions into shared preferences. Larger groups and discussions that are not constrained by a status quo, are more likely to

reach resolution because of a higher probability for reflective agency to emerge. Our work contributes to the rapidly growing literature on the effective functioning of online communities, and to the broader literature on community governance and dispute resolution.

ACKNOWLEDGEMENTS

This research received funding from the Swiss National Science Foundation (grant 169441). All authors contributed equally and are listed in alphabetical order. We are grateful for the excellent editorial guidance of Will Mitchell and the insightful comments of the three reviewers. We also thank Shiko Ben-Menahem, Roderick Swaab, Ilia Tsetlin, and Stefan Thau for their helpful feedback on previous drafts of this manuscript.

ORCID

Vivianna Fang He https://orcid.org/0000-0003-2591-7838

REFERENCES

Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H.-T. (2012). *Learning from data* (Vol. 4). New York, NY: AMLBook.

Adler, P. S. (2001). Market, hierarchy, and trust: The knowledge economy and the future of capitalism. *Organization Science*, *12*(2), 215–234.

Adler, P. S., & Heckscher, C. (2017). Collaboration as an organization design for shared purpose. In *Research in the sociology of organizations* (Vol. 57, pp. 81–111). Somerville, MA: Emerald Group Publishing Ltd.

Almeida, D. A., Murphy, G. C., Wilson, G., & Hoye, M. (2017). Do software developers understand open source licenses? In *IEEE International Conference on Program Comprehension* (pp. 1–11). IEEE Computer Society.

Alspaugh, T. A., Scacchi, W., & Asuncion, H. U. (2010). Software licenses in context: The challenge of heterogeneously-licensed systems. *Journal of the Association for Information Systems*, *11*(11), 730–755.

Amit, R., & Schoemaker, P. J. H. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, *14*(1), 33–46.

Arazy, O., Nov, O., Patterson, R., & Yeo, L. (2011). Information quality in Wikipedia: The effects of group composition and task conflict. *Journal of Management Information Systems*, *27*(4), 71–98.

Arazy, O., Ortega, F., Nov, O., Yeo, L., & Balila, A. (2015). Functional roles and career paths in Wikipedia. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1092–1105.

Bamberger, P. A. (2018). AMD—Clarifying what we are about and where we are going. *Academy of Management Discoveries*, *4*(1), 1–10.

Barley, S. R., & Tolbert, P. S. (1991). Introduction: At the intersection of organizations and occupations. *Research in the sociology of organizations*, 8(1–13). Greenwich, CT: JAI Press, Inc.

Bechky, B. A. (2003). Sharing meaning across occupational communities: The transformation of understanding on a production floor. *Organization Science*, *14*(3), 312–330.

Belenzon, S., & Schankerman, M. (2015). Motivation and sorting of human capital in open innovation. *Strategic Management Journal*, *36*(6), 795–820.

Bissyande, T. F., Lo, D., Jiang, L., Reveillere, L., Klein, J., & Traonm, Y. L. (2013). Got issues? Who cares about it? A large scale investigation of issue trackers from GitHub. In *2013 IEEE 24th International Symposium on Software Reliability Engineering, ISSRE 2013*, pp. 188–197.

Boudreau, K. J., & Jeppesen, L. B. (2015). Unpaid crowd complementors: The platform network effect mirage. *Strategic Management Journal*, *36*(12), 1761–1777.

Budd, J. W., & Colvin, A. J. S. (2014). The goals and assumptions of conflict management in organizations. In *The Oxford handbook of conflict management in organizations* (pp. 12–29). Oxford: Oxford University Press.

Butler, B. S., Bateman, P. J., Gray, P. H., & Diamant, E. I. (2014). An attraction-selection-attrition theory of online community size and resilience. *MIS Quarterly*, *38*(3), 699–728.

Collins, F. S., Morgan, M., & Patrinos, A. (2003). The human genome project: Lessons from large-scale biology. *Science*, *300*(5617), 286–290.

Cox, M., Arnold, G., & Tomás, S. V. (2010). A review of design principles for community-based natural resource management. *Ecology and Society*, *15*(4), 38–40.

Currie, D., Gormley, T., Roche, B., & Teague, P. (2017). The management of workplace conflict: Contrasting pathways in the HRM literature. *International Journal of Management Reviews*, *19*(4), 492–509.

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

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

Danneels, E. (2008). Organizational antecedents of second-order competences. *Strategic Management Journal*, *29*(5), 519–543.

Deodhar, S. J., Saxena, K. B. C., Gupta, R. K., & Ruohonen, M. (2012). Strategies for software-based hybrid business models. *The Journal of Strategic Information Systems*, *21*(4), 274–294.

Deutsch, M., & Krauss, R. M. (1962). Studies of interpersonal bargaining. *Journal of Conflict Resolution*, *6*(1), 52–76.

Dewey, J. (1922). *Human nature and conduct: An introduction to social psychology*. NY: Henry Holt and Company.

DiBona, C., Ockman, S., & Stone, M. (1999). *Open sources: Voices from the open source revolution*. Sebastopol, CA: O'Reilly Media.

Dietz, T., Ostrom, E., & Stern, P. C. (2003). The struggle to govern the commons. *Science*, *302*(5652), 1907–1912.

Economides, N., & Katsamakas, E. (2006). Two-sided competition of proprietary vs. open source technology platforms and the implications for the software industry. *Management Science*, *52*(7), 1057–1071.

Edelman, L. B., Krieger, L. H., Eliason, S. R., Albiston, C. R., & Mellema, V. (2011). When organizations rule: Judicial deference to institutionalized employment structures. *American Journal of Sociology*, *117*(3), 888–954.

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science*, *22*(5), 1224–1239.

Faraj, S., von Krogh, G., Monteiro, E., & Lakhani, K. R. (2016). Online community as space for knowledge flows. *Information Systems Research*, *27*(4), 668–684.

Forte, A., Larco, V., & Bruckman, A. (2009). Decentralization in Wikipedia governance. *Journal of Management Information Systems*, *26*(1), 49–72.

Foss, N. J., Frederiksen, L., & Rullani, F. (2016). Problem-formulation and problem-solving in self-organized communities: How modes of communication shape project behaviors in the free open-source software community. *Strategic Management Journal*, *37*(13), 2589–2610.

Glaser, B. G. (2008). *Doing quantitative grounded theory*, Mill Valley, CA: Sociology Press.

Gousios, G., Vasilescu, B., Serebrenik, A., & Zaidman, A. (2014). Lean ghtorrent: GitHub data on demand. In *11th Working Conference on Mining Software Repositories, MSR 2014 - Proceedings* (pp. 384–387). New York, NY: Association for Computing Machinery, Inc.

Grewal, R., Lilien, G. L., & Mallapragada, G. (2006). Location, location, location: How network embeddedness affects project success in open source systems. *Management Science*, *52*(7), 1043–1056.

Gulati, R., Lawrence, P. R., & Puranam, P. (2005). Adaptation in vertical relationships: Beyond incentive conflict. *Strategic Management Journal*, *26*(5), 415–440.

Gulati, R., Puranam, P., & Tushman, M. (2012). Meta-organization design: Rethinking design in interorganizational and community contexts. *Strategic Management Journal*, *33*(6), 571–586.

Gulati, R., & Singh, H. (1998). The architecture of cooperation: Managing coordination costs and appropriation concerns in strategic alliances. *Administrative Science Quarterly*, *43*(4), 781–814.

Hardin, S. E. (2005). Book review: Sports tourism: Participants, policy, and providers. *Journal of Travel Research*, *43*(3), 320–321.

Heckscher, C. C., & Adler, P. S. (2006). *The firm as a collaborative community: Reconstructing trust in the knowledge economy*, New York: NY: Oxford University Press.

Helms, W. S., Oliver, C., & Web, K. (2012). Antecedents of settlement on a new institutional practice: Negotiation of the ISO 26000 standard on social responsibility. *Academy of Management Journal*, *55*(5), 1120–1145.

Jehn, K. A., & Bendersky, C. (2003). Intragroup conflict in organizations: A contingency perspective on the conflict-outcome relationship. *Research in Organizational Behavior*, *25*, 187–242.

Jensen, C., & Scacchi, W. (2005). Collaboration, leadership, control, and conflict negotiation in the Netbeans.org open source software development community. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. Big Island, HI: IEEE.

Kalliamvakou, E., Gousios, G., Blincoe, K., Singer, L., German, D. M., & Damian, D. (2016). An in-depth study of the promises and perils of mining GitHub. *Empirical Software Engineering*, *21*(5), 2035–2071.

Kane, G. C., Johnson, J., & Majchrzak, A. (2014). Emergent life cycle: The tension between knowledge change and knowledge retention in open online coproduction communities. *Management Science*, *60*(12), 3026–3048.

Karambayya, R., & Brett, J. M. (1989). Managers handling disputes: Third-party roles and perceptions of fairness. *Academy of Management Journal*, *32*(4), 687–704.

Kellogg, K. C. (2009). Operating room: Relational spaces and microinstitutional change in surgery. *American Journal of Sociology*, *115*(3), 657–711.

Kittur, A., & Kraut, R. E. (2010). Beyond Wikipedia: Coordination and conflict in online production groups. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pp. 215–224.

Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. (2007). He says, she says: Conflict and coordination in Wikipedia. In *Conference on Human Factors in Computing Systems – Proceedings*, pp. 453–462.

Klapper, H., & Reitzig, M. (2018). On the effects of authority on peer motivation: Learning from Wikipedia. *Strategic Management Journal*, *39*(8), 2178–2203.

Kolb, D. M., & Putnam, L. L. (1992). The multiple faces of conflict in organizations. *Source: Journal of Organizational Behavior*, *13*(3), 311–324.

Lakhani, K. R., Lifshitz-Assaf, H., & Tushman, M. L. (2013). Open innovation and organizational boundaries: Task decomposition, knowledge distribution and the locus of innovation. In *Handbook of Economic Organization: Integrating Economic and Organization Theory* (pp. 355–382). Northampton, MA: Edward Elgar Publishing Ltd.

Laughlin, P. R. (2011). Social choice theory, social decision scheme theory, and group decision-making. *Group Processes and Intergroup Relations*, *14*(1), 63–79.

Lee, G. K., & Cole, R. E. (2003). From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organization Science*, *14*(6), 633–649.

Lee, M. Y., & Edmondson, A. C. (2017). Self-managing organizations: Exploring the limits of less-hierarchical organizing. *Research in Organizational Behavior*, *37*, 35–58.

Leonardi, P. M. (2015). Ambient awareness and knowledge acquisition: Using social media to learn 'who knows what' and 'who knows whom'. *MIS Quarterly*, *39*(4), 747–762.

Lerner, J., & Schankerman, M. (2010). *The demand side: Assessing trade-offs and making choices*, Cambridge, MA: MIT Press.

Lerner, J., & Tirole, J. (2002). Some simple economics of open source. *The Journal of Industrial Economics*, *50*(2), 197–234.

Lerner, J., & Tirole, J. (2005). The economics of technology sharing: Open source and beyond. *Journal of Economic Perspectives*, *19*(2), 99–120.

Löhr, K., Hochmuth, C., Graef, F., Wambura, J., & Sieber, S. (2017). Conflict management programs in trans-disciplinary research projects: The case of a food security project in Tanzania. *Food Security*, *9*(6), 1189–1201.

Lorge, I., & Solomon, H. (1955). Two models of group behavior in the solution of eureka-type problems. *Psychometrika*, *20*(2), 139–148.

Maciejovsky, B., & Budescu, D. V. (2013). Verbal and numerical consumer recommendations: Switching between recommendation formats leads to preference inconsistencies. *Journal of Experimental Psychology: Applied*, *19*(2), 143–157.

Mahony, S. O., & Lakhani, K. R. (2011). Organizations in the shadow of communities. In *Communities and organizations* (Vol. 33, pp. 3–36). Bingley, UK: Emerald Group Publishing.

March, J. G. (1962). The business firm as a political coalition. *The Journal of Politics*, *24*(4), 662–678.

March, J. G., & Olsen, J. P. (1976). Organizational choice under ambiguity. *Ambiguity and Choice in Organizations*, *2*, 10–23.

March, J. G., & Simon, H. A. (1958). *Organizations*. Oxford, England: Wiley.

Markus, M. L. (2007). The governance of free/open source software projects: Monolithic, multidimensional, or configurational? *Journal of Management and Governance*, *11*(2), 151–163.

Marlow, J., Dabbish, L., & Herbsleb, J. (2013). Impression formation in online peer production: Activity traces and personal profiles in github. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pp. 117–128.

Menkel-Meadow, C. (2001). Negotiating with lawyers, men, and things: The contextual approach still matters. *Negotiation Journal*, *17*(3), 257–293.

Miller, K. D., Fabian, F., & Lin, S. J. (2009). Strategies for online communities. *Strategic Management Journal*, *30* (3), 305–322.

Moon, J. Y., & Sproull, L. (2002). Essence of distributed work: The case of the Linux kernel. In *Distributed work* (pp. 381–404). Cambridge, MA: The MIT Press.

Morrill, C., & Rudes, D. S. (2010). Conflict resolution in organizations. *Annual Review of Law and Social Science*, *6*, 627–651.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*, 87–106.

Murray, F., & O'Mahony, S. (2007). Exploring the foundations of cumulative innovation: Implications for organization science. *Organization Science*, *18*(6), 1006–1021.

Nemeth, C. J. (1986). Differential contributions of majority and minority influence. *Psychological Review*, *93*(1), 23–32.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

O'Mahony, S., & Bechky, B. A. (2008). Boundary organizations: Enabling collaboration among unexpected allies. *Administrative Science Quarterly*, *53*(3), 422–459.

O'Mahony, S., & Ferraro, F. (2007). The emergence of governance in an open source community. *Academy of Management Journal*, *50*(5), 1079–1106.

O'Mahony, S. (2003). Guarding the commons: How community managed software projects protect their work. *Research Policy*, *32*(7), 1179–1198.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge, UK: Cambridge University Press.

Ostrom, E. (1995). Designing complexity to govern complexity. In *Property rights and the environment: Social and ecological issues* (pp. 33–45). Washington DC: The World Bank.

Peirce, C. S. (1878). Deduction, induction and hypothesis: Popular science monthly, v. 13.

Piskorski, M. J., & Gorbatâi, A. (2017). Testing coleman's social-norm enforcement mechanism: Evidence from Wikipedia. *American Journal of Sociology*, *122*(4), 1183–1222.

Preece, J. (2000). Online communities: Designing usability, supporting sociability. *Industrial Management and Data Systems*, *100*(9), 459–460.

Puranam, P., Shrestha, Y. R., He, V. F., & von Krogh, G.. (2020). Algorithmic induction through machine learning: Opportunities for management and organization research. *INSEAD Working Paper No. 2018/11/STR*.

Raymond, E. S. (1999). The cathedral and the bazaar. *Knowledge, Technology and Policy*, *12*(3), 23–49.

Ren, Y., Chen, J., & Riedl, J. (2016). The impact and evolution of group diversity in online open collaboration. *Management Science*, *62*(6), 1668–1686.

Roche, W. K., & Teague, P. (2012). Human resource management and ADR practices in Ireland. *International Journal of Human Resource Management*, *23*, 528–549.

Roche, W. K., Teague, P., & Colvin, A. J. (2014). *The Oxford handbook of conflict management in organizations*. Oxford: Oxford University Press.

Ross, W. H., & Conlon, D. E. (2000). Hybrid forms of third-party dispute resolution: Theoretical implications of combining mediation and arbitration. *Academy of Management Review*, *25*(2), 416–427.

Roth, A. E., Murnighan, J. K., & Schoumaker, F. (1988). The deadline effect in bargaining: Some experimental evidence. *American Economic Review*, *78*(4), 806–823.

Saeed, A. R., McDermott, C., & Boyd, E. (2017). Are REDD+ community forest projects following the principles for collective action, as proposed by Ostrom? *International Journal of the Commons*, *11*(1), 572–596.

Schelling, T. C. (1960). *The strategy of conflict*. Oxford, England: Harvard University Press.

Selznick, P. (1948). Foundations of the theory of organization. *American Sociological Review*, *13*(1), 25.

Shah, S. K. (2006). Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science*, *52*(7), 1000–1014.

Sherif, M. (1958). Superordinate goals in the reduction of intergroup conflict. *American Journal of Sociology*, *63* (4), 349–356.

Simons, H. (1947). *Administrative behavior: A study of decision-making processes in administrative organizations*. New York, NY: The Macmillan Company.

Singh, P. V., & Phelps, C. (2013). Networks, social influence, and the choice among competing innovations: Insights from open source software licenses. *Information Systems Research*, *24*(3), 539–560.

Spaeth, S., von Krogh, G., & He, F. (2015). Research note—Perceived firm attributes and intrinsic motivation in sponsored open source software projects. *Information Systems Research*, *26*(1), 224–237.

Sproull, L., & Arriaga, M. (2007). Online communities. In H. Bidgoli (Ed.), *Handbook of computer networks: Distributed networks, network planning, control, management, and new trends and applications* (Vol. 3). New York: John Wiley & Sons.

St. Laurent, A. M. (2004). *Understanding open source and free software licensing: Guide to navigating licensing issues in existing and new software*, Sebastopol, CA: O'Reilly Media.

Stanko, M. A. (2016). Toward a theory of remixing in online innovation communities. *Information Systems Research*, *27*(4), 773–791.

Stewart, K. J., & Gosain, S. (2006). The impact of ideology on effectiveness in open source software development teams. *MIS Quarterly*, *30*(2), 291–314.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.

Strauss, A. L., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA: Sage Publications.

Subramaniam, C., Sen, R., & Nelson, M. L. (2009). Determinants of open source software project success: A longitudinal study. *Decision Support Systems*, *46*(2), 576–585.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, *58*(1), 267–288.

Tidhar, R., & Eisenhardt, K. M. (2020). Get rich or die trying... finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal*, *41*, 1245–1273.

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

von Krogh, G., Haefliger, S., Spaeth, S., & Wallin, M. W. (2012). Carrots and rainbows: Motivation and social practice in open source software development. *MIS Quarterly*, *36*, 649–676.

Wang, J., Shih, P. C., Wu, Y., & Carroll, J. M. (2015). Comparative case studies of open source software peer review practices. *Information and Software Technology*, *67*, 1–12.

Weber, M. (1946). The sociology of charismatic authority. In *From Max Weber: Essays in sociology* (pp. 248–249). New York: Oxford University Press.

Weber, S. (2004). *The success of open source*, Cambridge, MA: Harvard University Press.

Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications: A study in the economics of internal organization*. New York, NY: Free Press.

Wilson, D. S., Ostrom, E., & Cox, M. E. (2013). Generalizing the core design principles for the efficacy of groups. *Journal of Economic Behavior and Organization*, *90*, S21–S32.

Wittenbaum, G. M., & Stasser, G. (1996). Management of information in small groups. In *What's social about social cognition? Research on socially shared cognition in small groups* (pp. 3–28). Thousand Oaks, CA: SAGE Publications.

Zhou, Z. H., & Jiang, Y. (2003). Medical diagnosis with C4.5 Rule preceded by artificial neural network ensemble. *IEEE Transactions on Information Technology in Biomedicine*, *7*(1), 37–42.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: He VF, Puranam P, Shrestha YR, von Krogh G. Resolving governance disputes in communities: A study of software license decisions. *Strat. Mgmt. J.* 2020;41:1837–1868. https://doi.org/10.1002/smj.3181