

Theoretical and Empirical Modeling of Identity and Sentiments in Collaborative Groups

White Paper / Final Report

Project THEMIS.COG, funded by the Digging Into Data Initiative, Trans-Atlantic Platform

March 2017 – August 2021

Overall Project Coordinator: Jesse Hoey, University of Waterloo, Canada

Grant Holders:

University of Waterloo, 200 University Avenue West, Waterloo On N2L 3G1, Canada

PIs: Jesse Hoey (jhoey@cs.uwaterloo.ca) & Meiyappan Nagappan (mei.nagappan@uwaterloo.ca)

SSHRC grant # 869-2016-0005 & NSERC grant # 5660-501727

Department of Sociology, Dartmouth College, 6104 Blunt Hall, Room 302, Hanover, NH 03755 USA

PI: Kimberly B. Rogers (kr Rogers@dartmouth.edu), NSF grant # 1723608

Potsdam University of Applied Sciences, Kiepenheuerallee 5, 14469 Potsdam, Germany

PI: Tobias Schröder (tobias.schroeder@fh-potsdam.de), DFG grant # SCHR1282/3-1

Additional authors of this report (in alphabetical order):

Diego Dametto (Potsdam University of Applied Sciences), Nalin De Zoysa (University of Waterloo), Rahul Iyer (University of Waterloo), Jonathan H. Morgan (Potsdam University of Applied Sciences), Deepak Rishi (University of Waterloo), Antonio D. Sirianni (Dartmouth College), Seonghu Yun (University of Waterloo), Jun Zhao (currently at Georgia State University, previously at Dartmouth College), Nikolas Zöller (Potsdam University of Applied Sciences)

Copyright Notice: This content is licensed under the Creative Commons Attribution 4.0 International license (CC BY 4.0). Exceptions may apply to specific figures reproduced from separate publications, as described in the figure captions.

Summary

THEMIS.COG was an interdisciplinary research collaboration of computer scientists and social scientists from the University of Waterloo (Canada), Potsdam University of Applied Sciences (Germany), and Dartmouth College (USA). Funded by the Trans-Atlantic Platform's Digging Into Data initiative, the project aimed at theoretical and empirical modeling of identity and sentiments in collaborative groups. Understanding the social forces behind self-organized collaboration is important because technological and social innovations are increasingly generated through informal, distributed processes of collaboration, rather than in formal organizational hierarchies or through market forces. Our work used a data-driven approach to explore the social psychological mechanisms that motivate such collaborations and determine their success or failure. We focused on the example of GitHub, the world's current largest digital platform for open, collaborative software development.

In contrast to most, purely inductive contemporary approaches leveraging computational techniques for social science, THEMIS.COG followed a deductive, theory-driven approach. We capitalized on affect control theory, a mathematically formalized theory of symbolic interaction originated by sociologist David R. Heise and further advanced in previous work by some of the THEMIS.COG collaborators, among others. Affect control theory states that people control their social behaviours by intuitively attempting to verify culturally shared feelings about identities, social roles, and behaviour settings. From this principle, implemented in computational simulation models, precise predictions about group dynamics can be derived. It was the goal of THEMIS.COG to adapt and apply this approach to study the GitHub collaboration ecosystem through a symbolic interactionist lens.

In many ways, the project was more challenging than we had anticipated. Specifically, current computational techniques for natural language processing and sentiment analysis did not allow us to match GitHub data to affect control theory models in the precise or even automatic ways we had hoped. Nevertheless, we were able to develop various novel theoretical and empirical computational approaches to the study of group collaboration. Among them were a classification of socio-emotional versus task-oriented group behaviours and other ways of extracting emotional information from group discussions, a demonstration of correlations between developer personality/identity and the likelihood to get code changes accepted, development of a typology of interaction patterns on GitHub, and various approaches to simulating group dynamics among software developers with affect control theory. THEMIS.COG thus contributed substantially to the novel endeavour of theory development in social science based on large amounts of naturally occurring digital data.

Keywords: Computational Social Science, Group Dynamics, Software Development, Virtual Teams, GitHub, Identity, Affect Control Theory, Sentiment Analysis, Social Simulation

List of Tables

1. One-Versus-All classifications: (a) IPA Categories, (b) Emotions, (c) Aggregated classes.
2. Possible GitHub reactions with emojis.
3. Percentage of positive reactions to comments on GitHub, inferred from on emojis.

List of Figures

1. EPA sentiment (inferred from personality scores) and pull-request acceptance.
2. Typology of interaction patterns in GitHub projects.
3. Geographical distribution of GitHub projects in Germany.
4. Geographical distribution of GitHub projects in China.
5. Frequency distribution of programming languages used in German vs. Chinese GitHub projects.
6. Co-occurrence of programming languages used in German vs. Chinese GitHub projects.
7. Example distribution of empirical sentiment ratings.
8. Cross-cultural differences between the U.S. and China when interacting with stigmatized group members.
9. Effects of hierarchical vs. egalitarian group cultures on experienced affective deflection (tension).
10. Constraints on interactions in stylized egalitarian (E) vs. hierarchical (H) groups.
11. Distribution of group behaviors as a function of evaluation (status) vs potency (power) distance in collaborative groups.
12. Simulation of gendered status differences in task groups emerging from affect control mechanisms.
13. Distinct GitHub projects employing the Julia language by use type for each month from January 2010 to June 2019.
14. Network of programming languages in the GitHub community in 2017.

1. Background and Goals of THEMIS.COG

Collaborative open-source software development is a good example of major changes in economic production and social problem-solving in the 21st century. The field is characterized by strong meritocratic norms, often accompanied by a rejection of traditional hierarchical and bureaucratic ways of organizing work. Developers can contribute to projects hosted on digital platforms from almost anywhere in the world. They often contribute to a commons rather than just private wealth, and it is easy to move on from one project to the next as opposed to committing oneself to a specific corporation for one's entire career or substantial parts thereof.

The goal of the THEMIS.COG project was to contribute to a better understanding of the social psychological dynamics underlying the described socio-economic changes, ranging from questions about the internal functioning of collaborative groups – how their communication patterns can be understood, who gains status or power in them, what groups are more successful than others, etc. – to those about the larger emerging ecosystem sustained by such patterns of collaboration – who is attracted to what kind of group, how sustainable are different types of groups, how do innovative ideas spread from one group to the next, etc.

In line with the goals of the Trans-Atlantic Platform's Digging Into Data (TAP-DiD) Initiative, which funded the project, THEMIS.COG sought to address the research questions with a data-science approach, capitalizing on the fact that it is possible to access large-scale data from hundreds of thousands of public collaborative software development projects on the GitHub platform. However, in contrast to most computational social science work which is rather inductive by its nature, we pursued a strict theory-driven approach by grounding the work in affect control theory, a mathematically formalized theory of social interaction established in sociological social psychology. As a result, THEMIS.COG pursued a dual computational strategy: (1) theory-driven development of generative computational models of social interaction and group dynamics, accompanied by (2) empirical models of data pulled from the GitHub platform, where the analysis was informed by the theoretical models and ultimately aimed at testing them.

Affect control theory (ACT; Heise, 2007) combines social constructionist ideas about the role of shared meanings as motivators of social action with psychological theories of cognitive-affective coherence and cybernetic principles from complex-systems science. The empirical point of departure is the measurement of meanings of identities, actions, settings, personality traits, and emotional states in a three-dimensional affective space spanned by base vectors evaluation, potency, and activity (EPA) shown in decades of research to be a universal organizing structure of the human mind (Heise, 2007, 2010; Osgood, May, & Miron, 1975). In addition, affect control theory researchers have developed statistical techniques to model culture-specific mechanisms of information integration during dynamic social interactions (e.g., Heise, 2010; Morgan, Rogers, & Hu, 2016; Schröder, 2011). Coefficients from these statistical models have been used to parameterize generative computational models of social interaction at the dyadic and small-group levels, which have been shown to yield valid predictions of behavioral and emotional dynamics (e.g., Heise, 2013; Schröder & Scholl, 2009). In a recent probabilistic generalization called BayesACT, affect control theory was merged with partially observable Markov decision process (POMDP) models from Artificial Intelligence in order to better account for meaning variance, uncertainty, and dynamic changes in realistic social interaction (Hoey, Schröder, & Alhothali, 2016; Schröder, Hoey, & Rogers, 2016).

The overarching goal of the THEMIS.COG project was to use affect control theory as a theoretical framework to study the above-mentioned questions about the social psychological mechanisms underlying the self-organized work of software development teams on digital platforms. To this end, we first sought to develop computational methods that would allow us to scale the empirical apparatus

of ACT from a classic social-science approach (surveys of people, small-scale experiments) to a large-scale digital context capitalizing on existing real-world data from the GitHub platform. Thus, the first goal consisted of developing appropriate natural-language processing, sentiment, and network-analysis methods to enable us to conceptualize data from GitHub in terms of the theoretical concepts of ACT. Correspondingly, we sought to extend the existing generative simulation models of ACT so as to be applicable to the analysis of the real-world work situation of collaborative software development on a platform. Equipped with these tools, we hoped to be able to deliver not only descriptive accounts of interaction patterns in the GitHub ecosystem, but also theoretical explanations in terms of extant theory about the role of cognitive-affective mechanisms, interaction structures, and culture as embedded in ACT's mathematical models.

2. Project Development, Problems, and Deviations from Original Plan

We were able to organise a truly integrated research project in line with the goals of the Trans-Atlantic Platform initiative, with sustained cross-site collaboration between researchers at various career stages from the three involved universities. At least before the start of the Covid-19 pandemic, we organised several physical meetings of the research team and specific longer visits (between 2 weeks and 2 months) of researchers to the partner labs (e.g., Nikolas Zöller from Potsdam to Waterloo, Jun Zhao and Antonio Sirianni from Dartmouth to Potsdam). In addition, we were able to sustain regular online meetings (before and during the pandemic) mostly between the PIs, but also partly involving the more junior researchers. Overall, we were thus able to capitalize on the unique combination of interdisciplinary expertise across the Atlantic (Waterloo: Computer Science, Potsdam: Cognitive Science and Innovation Studies, Dartmouth: Sociology) to substantially advance the work on the problems outlined in the original grant proposal.

Nevertheless, we substantially underestimated some of the scientific problems related to our approach, which led to significant setbacks and deviations from the original work plan. These changes included frequent readjustment of goals and partial reallocation of responsibilities across the three universities in a more agile style of project management than we had originally anticipated. They also resulted in substantial resource constraints, as we made management decisions to devote more time than anticipated to basic problems, and thus our inability to address some of the problems (especially in later work packages) in the comprehensive way we had intended to at the start of the project.

At the empirical level, it turned out that sentiment analysis of GitHub data was more challenging than initially thought. First, ongoing group discussions aimed at solving technical problems constitute a much more contextualized and complex structure of linguistic data than other areas of application where sentiment analysis was developed (e.g., product reviews or limited-length microblogging services such as Twitter). Second, software development is a highly structured task domain where the socio-emotional group processes we were interested in are not the focus of the conversations captured by a platform such as GitHub – we reckon these processes do happen, but likely in physical settings or on other platforms more tailored towards informal communication (e.g., Slack, for comparison see Corritore, Goldberg, & Srivastava, 2020). As a consequence, there were many null results, only small effect sizes, and needed but unanticipated workarounds. Most importantly, a direct mapping of GitHub data to the constituent components of affect control theory models (identity, behaviour, and setting meanings) could not be achieved. Hence, much of our analyses focused more on structural aspects of communication and concepts more loosely coupled to our theoretical framework such as personality, status, or general mood/tonality of group communication.

There were also challenges related to theoretical model development. Since directly mapping from empirical GitHub data to model entities proved not feasible, a more indirect approach was chosen, where we established empirical phenomena in one line of work and sought to explain them mechanistically with affect control theory through simulations in another line of work. Also, when starting the project, we overestimated the capacity of BayesACT model to represent task problems in a structured environment; therefore, theory development throughout the project split into two branches: a novel BayesACT dual-process model at the level of individual cognition, and a group process model based on classical, deterministic affect control theory, while integration of these models remains a task for the future.

Despite these setbacks and readjustments of our research strategy, we were able to significantly develop our understanding of self-organized group collaboration. We were able to not only advance computational methods for analyzing social interaction, but also to do this in a firmly theoretically grounded way by using affect control theory as a conceptual framework. We outline key results of the research throughout the project in the following section. The TAP-DiD grant also allowed us to substantially strengthen our interdisciplinary and transatlantic research collaboration, which we expect to be able to sustain well beyond the end of the funding through this program.

3. Key Results

3.1. Identifying Sentiments and Group Behaviours on GitHub

Our initial investigations centered around attempts to characterize discussions on GitHub in terms of their social purposes (the description in this section is based on Hoey et al., 2018 and Rishi, 2017, where more details on methods and results can be found). We used the Interaction Process Analysis (IPA) categories of Bales (1950) as a standard classification scheme for human actions in a group. This scheme distinguishes 12 categories of human actions, including *showing solidarity*, *giving suggestions*, *showing tension*, etc. We also looked at trying to classify 10 emotions, both positive (thanks, calm, cautious, happy) and negative (sorry, nervous, careless, aggressive, defensive, angry). These emotion words were chosen to span ACT's three-dimensional emotional space (EPA) as identified by and related to IPA categories by Heise (2013). The ability to make these classifications would have allowed us to directly feed into group process simulations of software development on GitHub similarly to Heise (2013), explained in more detail in subsequent sections.

We focused on discussions related to GitHub pull requests. We randomly selected 834 pull requests and a total of 3,000 pull request comments from GitHub in February 2017. The comments were filtered to remove sections of code, then annotated by four people for the 12 IPA labels and 10 emotions described above. One annotator was a member of the project team, while three were hired on Amazon Mechanical Turk (MTurk). The three MTurk annotators had experience in programming and were familiar with GitHub. Furthermore, they were screened according to their ratings on an initial set of 50 pull request comments. Detailed instructions on how to annotate a particular pull request comment were provided, and each pull request comment could be annotated with a maximum of three IPA categories and a maximum of three emotions. The participants were also asked to filter out any unnecessary sections of code in the comment. More details can be found in Rishi (2017). Majority voting was used to threshold all the ratings; a comment was assigned an IPA/emotion label if three out of four raters assigned it that label.

(a)	
IPA category	F1-score
Shows solidarity	56.8
Shows tension release	10
Agrees	64
Gives suggestion	33.4
Gives opinion	51.4
Gives orientation	58.6
Asks for orientation	36.2
Asks for opinion	22.9
Asks for suggestion	10.6
Disagrees	56.6
Shows tension	30
Shows antagonism	13.2

(b)	
Emotion	F1-score
Thanks	54.7
Sorry	58.7
Calm	69.3
Nervous	23.6
Careless	15.7
Cautious	69.8
Aggressive	25.2
Defensive	16.7
Happy	2.5
Angry	0

(c)	
Aggregated sets	F1-score
Positive versus negative reactions	73.2
Questions versus attempted answers	81
Positive versus negative emotions	80.5

Note. IPA = Interaction Process Analysis.

Table 1. One-versus-all classifications: (a) IPA Categories, (b) Emotions, (c) Aggregated classes. Reproduced from Hoey et al. (2018) with permission, copyright by the authors of the article.

Each word in each pull request was first mapped into a dimensional space defined by statistical patterns of words (i.e., word vectors) found in a large online corpus (the Google News corpus). The word vector space is structured so that words that statistically co-occur in the same contexts in the corpus are coded by vectors that are close to one another in the space. Word vectors can be used to numerically compare words according to their usage in context, and thus can be used as a way of characterizing the cultural relationships between words (Kozlowski et al., 2019). Word vectors from pull requests in our study were weighted by their term-frequency, inverse document frequencies (TF-IDF) for each comment, which promotes words that are more important for the comment as a whole. Finally, a linear support vector machine (SVM) was trained on the resulting weighted vectors. An SVM optimizes a linear boundary between elements of two classes such that the two classes are maximally separated. Logistic regression, metric learning, and a variety of deep learning methods yielded similar results (for details, see Rishi, 2017). We show the F1-scores (evenly weighted precision and recall) for a one-versus-all classification task of all IPA categories and all emotions in Table 1. Each score in this table shows how

easily the given category or emotion could be distinguished from all other categories if they were grouped together. Parameters for the algorithms were set by searching exhaustively over a reasonably large, evenly spaced set of possibilities. Results are for a fivefold cross-validation in which four-fifths of the data is used for training the SVM classifier and one fifth is used for testing, and this process is repeated for all five splits. From the results, it is clear that the task presents a significant challenge, which we believe can only be overcome by using more detailed emotional analysis of each comment. We also examined aggregated IPA and emotion categories by grouping IPA categories into positive vs. negative reactions and questions vs. attempted answers, and grouping emotions into positive and negative categories. The results in Table 3(c) show that this task is much simpler, and F1-scores over 80% can be achieved.

The results of this data analysis showed that the task of sentiment and interaction analysis is a major challenge in cases with more technical conversations than what is usually attempted. While at least high-level aggregate classification seems achievable, more fine-grained extraction of basic IPA categories, emotions, or – for the purposes of modeling with affect control theory – identities remains an unsolved problem.

3.2. Personality and Pull Request Acceptance

One of the major theoretical motivations of the project was to understand how human factors such as identity maintenance and, relatedly, personality influence self-organized online group collaboration. Tsay et al. (2014) had found that perceived social distance influences the evaluation of group member contributions. Social distance in this case refers to whether a group member making a request for changes in the code follows the person who evaluates the request (referred to as the closer) on GitHub. Tsay et al. had found that following closers, a history of previous interactions, and the requester's follower count all increase the odds of accepting a change request.

Using Tsay et al.'s work as a launching point, the guiding research question for this subproject was whether personality dynamics also influenced the likelihood of acceptance (for more details on the work described in this section, see Iyer et al., 2019). We first replicated Tsay et al.'s major findings. We then extracted all the comments made by the developers of 1,860 projects – these projects come from Tsay et al.'s study and from the RepoReaper dataset. We next had *IBM Watson Personality Insights*, a commercial software tool for psycholinguistics, analyze the comments to infer the personality traits of 16,935 developers (see Iyer et al., 2019, pp. 3-4 for more details) using the Big Five personality model (Norman, 1963; Goldberg, 1992; Golbeck et al., 2011).

We found that the personality of the requester, of the closer, and the difference between the two in terms of personality traits influenced acceptance of pull requests (requests to change the code). Requesters who tended to be more open and conscientious in their comments had higher odds of being accepted, while requesters who were more extroverted had lower odds. Closers whose comments reflected greater conscientiousness, extraversion, and neuroticism tended to accept more requests. Finally, we found that absolute differences in openness, conscientiousness, extraversion, and neuroticism between closers and requesters all increase the odds of an acceptance. For more detailed results, see Iyer et al. (2019).

By linking personality to action in a workplace setting based on naturally occurring data, this study made a major contribution. The personality/situation paradox, the seeming inability of personality traits to predict behavior, has hounded personality researchers for decades (Ross & Nisbett, 1991; Zayas & Shoda, 2009). Nevertheless, the study remains limited in that it is correlational only and does not provide a mechanistic, causal explanation of how personality influences the group process, which in turn influences the outcome of the work; i.e., a changed code base.

3.3. A Deeper Analysis: Personality and Affective Group Processes

To address the lack of mechanistic explanation in the Iyer et al. (2019) study, we sought to extend the work on personality by modeling group interactions using an agent based simulation developed in a different work package (see section 3.11 below), Affect Control Theory based simulations of Interaction in Networked Groups (ACTING). Specifically, the goal is to test whether identity-maintenance processes as described by affect control theory help explain the associations between pull-request acceptance and the trait/role combinations observed by Iyer et al. (2019). Do affective alignment between the roles of closer and requester and traits such as conscientiousness and extraversion, for example, help explain why some trait/role combinations increase the likelihood of acceptance while other combinations dampen it? Do some trait/role combinations result in less intra-group tension because they better align with the connotative meanings of these roles, making acceptance more likely?

Assessing these hypotheses using Iyer et al.'s data requires two data transformations: 1) inferring raw Big 5 scores from reported percentiles; and 2) transforming these scores into Evaluation, Potency, and Activity (EPA) profiles. Fortunately, Gogoladze (2020) collected Big 5 data that includes both percentile and raw scores from AGILE developer groups that was roughly contemporaneous with Iyer and colleagues' data. Using Gogoladze's data, we were able to infer the raw scores from the percentiles. We next used equations published in a series of papers in the 1990s by Albert Mehrabian to transform the Big 5 score into EPA profiles, with his 1996 paper proving particularly helpful.

The movement from Big 5 scores to EPA profiles, however, entails a reduction in dimensions. Big 5 personality traits involve combinations of sentiment dimensions. For example, extraversion loads primarily on potency (0.72) and to some extent on evaluation (0.24) – see Mehrabian, 1996, pp. 88-89 for more details. Consequently, before moving to simulation, we re-estimated the Iyer et al. (2019) model (cf. section 3.2) using EPA values in order to get a better sense of the associations between acceptance and the EPA dimensions. Like in Iyer et al., we control for social factors and technical factors. Figure 1 displays the association between acceptance and role/sentiment combinations. Each panel in the plot shows a role/sentiment pair such as closer activity vs. requester evaluation; cold colors indicate a significant negative association with acceptance, warm a positive association, white no association. The labels indicate the magnitude and direction of the effect. The panel's cells correspond to magnitude pairs (high-low, medium-low, etc.). For example, the top-left cell of the top-left panel indicates that a requester who was a bit brusque (-0.97 on evaluation) in their commit comments who made a pull-request to a closer who was also brusque (-1) was no more or less likely to have their pull request accepted than a requester whose comments were moderately nice who made a pull request to a moderately nice closer. In contrast, a requester who was brusque in their comments who made a pull request to a moderately nice closer was less likely to have their pull-request accepted.

Overall, we find that very nice and very active commit comments do not generally portend acceptance. In contrast, comments by requesters and closers that express dominance *do* tend to lead to acceptance. Our working hypothesis is that dominant comments signal confidence, whereas nice, active comments are associated with explanations and face-saving. We are currently working on simulation studies using the ACTING model described in section 3.11, attempting to replicate these associations in role/sentiments pairs, first in closer-requester pairs and then in larger groups.

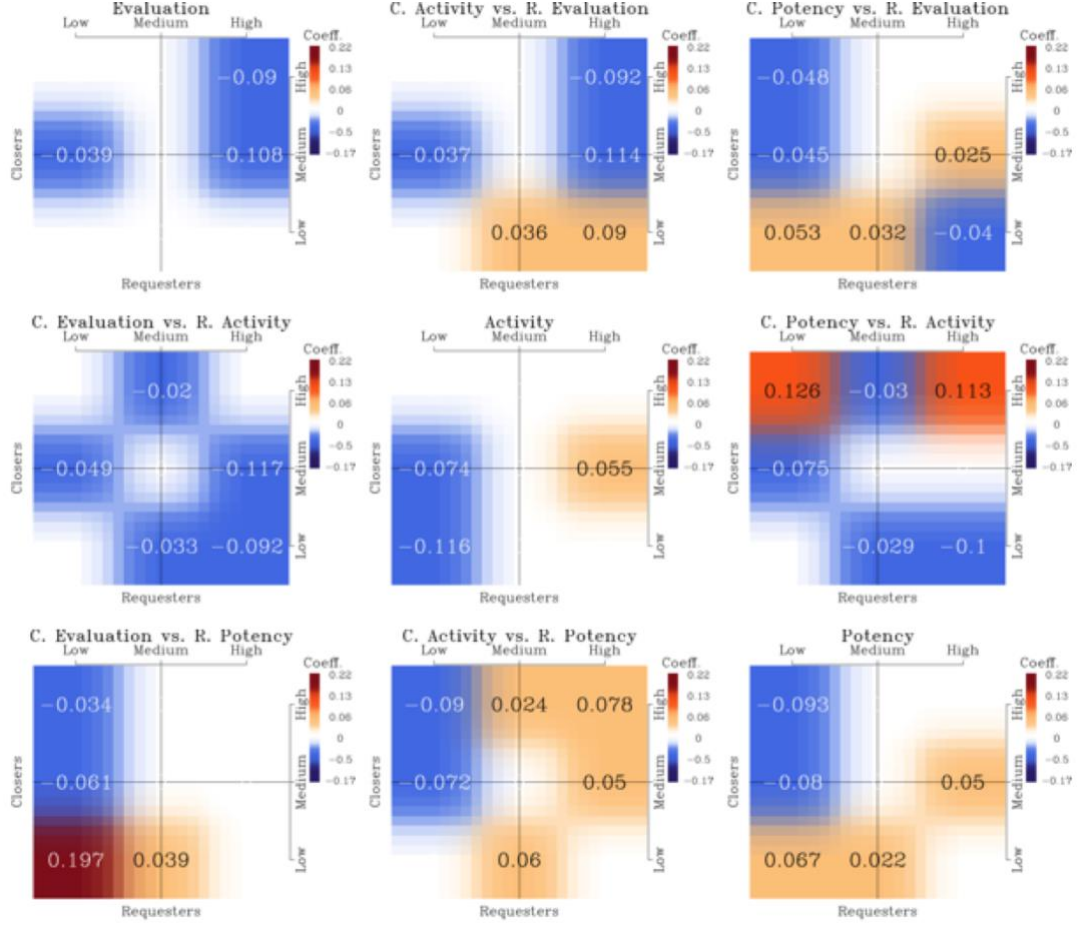


Figure 1. EPA sentiment (inferred from personality scores) and pull-request acceptance. Unpublished data as of completion of this report.

3.4. Reactions to Pull Request Comments

We also attempted to infer sentiment in pull-request comments directly from the use of emoji (for details, see DeZoysa, 2020). These can be considered a large set of pseudo labelled data. Previous work looking at such reactions was done by the DeepMoji project (Felbo et al., 2017), which attempted to link the emoji used in a tweet with the textual information. In a Twitter tweet, an emoji can be used to replace a word, or even just for the aesthetic use of the emoji rather than for the actual meaning or content of it, such as when creating a form of text art. In the case of GitHub reactions, there is only a limited set of emojis which are used for reactions, and the nature of their usage makes it less likely for them to be used in nonstandard ways (see Table 2). They are almost exclusively used to indicate the reaction towards a post.

The type of posts which can receive a reaction are a pull request, an issue, and a comment on either a pull request or issue. By using sites such as GitHub Trending and RepoReaper (Munaiah et al., 2017), we were able to extract a list of the top GitHub repositories, from which all the comments from pull requests and issues which had reactions associated with them were scraped. This resulted in a dataset consisting of 1.5 million posts with reactions, across 1700 projects with a set of 800,000 unique users (for details, see DeZoysa, 2020). With this data we tested whether prior interaction between users plays a role in influencing reactions, and more broadly if there is non-textual information that can influence what type of reaction is received.









content	emoji
+1	
-1	
laugh	
confused	
heart	
hooray	
rocket	
eyes	

Table 2: Possible GitHub reactions with emojis.

We first categorized the reactions into positive versus negative. The +1, heart, laughing, and hooray reactions are all unambiguously positive. The rocket emoji is also classified as positive as it is typically used to mean “ship it” – a common phrase in software development communities which means to deploy, launch, or merge something. The -1 and confused reactions are unambiguously negative. The remaining reaction is the eyes emoji, which is not clearly negative or positive. Based on manual assessment it is used positively, negatively, and in the sense of “take a look at this.” This is also the reaction used least often. Due to its semantic ambiguity and infrequent use, we decided to exclude the eyes emoji from the analysis.

	No Interactions		Only Seen		Mentioned		Overall	
	Percent	SE	Percent	SE	Percent	SE	Percent	SE
Positive Reactions	94.00%	0.01%	94.54%	0.03%	97.66%	0.02%	94.64%	0.01%

Table 3: Percentage of positive reactions to comments on GitHub, inferred from on emojis.

Our first question was whether or not a history of interactions between the user making a post and the user reacting to it influences what sort of reaction is created. We consider two basic forms of previous interaction. First, when two users have commented on the same pull request or issue, we can consider both users as having “seen” each other, and count it as an interaction. Second, when one user specifically includes the username of another user in their post, we consider one user as having “mentioned” the other, and count this as an interaction. Table 3 shows the percentage of positive reactions and the corresponding standard error for different subgroups.

Users who had no previous interactions apart from reacting to a post had the lowest percentage of positive reactions at 93.63%. Users who both commented in the same issue or pull request, but had not mentioned each other by username specifically, are indicated in the table as “seen but not mentioned.” They had a positive reaction ratio of 94.54%, which is higher than the set of users who had never interacted before. Users mentioning another user had the highest positive reaction ratio at 97.62%. We also performed a Mann-Whitney U-test or Wilcoxon test for each of the interaction subsets, using each project as an observation, and found that the p-values for each of the tests were below 0.05, confirming that these were different distributions. This same test was repeated for each subsequent analysis, and all were confirmed to have a p-value of less than 0.05.

These results seem to support the notion that users who have interacted with each other will have more positive reactions to one another. The first issue that we might have with this analysis is that different projects are likely to have different repository characteristics, such as the ratio of positive reactions or the amount that users interact with each other. It is therefore possible for there to be a few projects which skew the overall dataset in ways which are not representative of the standard behavior in the majority of projects. Further results investigating this are found in DeZoysa (2020). The simulation tools described in subsequent sections below can be used to study how affective tonality in a GitHub group becomes established in chained interactions.

3.5. Task Interdependence, Emotion, and Relational Cohesion

So far, the analyses described relate to the sentiments associated with the constituent elements of the groups (personality, identities, specific acts), but not the groups as an object themselves. Therefore, an additional study was aimed at investigating how the interaction patterns within small groups of 8 to 21 members on GitHub affect the collective emotional experience of developers (for details, see Dametto, 2019). Specifically, the following research question was addressed: *Under which conditions do members develop positive emotions towards a small group in which they are involved?* The project combined Lawler et al.’s (2009) theory of social commitment with vast-scale analysis of real data from GitHub – social network, statistical, and sentiment analysis – allowing us to move beyond a simplistic topological explanation of the origins of emotions by including cognitive factors derived from an established theoretical tradition.

The interactional data was modeled following the two pillars of the theory of social commitment: the theory of relational cohesion and the affect theory of social exchange. The theory of relational cohesion (Thye et al., 2002), which focuses on the mutual dependency of group members, was operationalized as the number of other small groups a developer is contributing to. The affect theory of social exchange (Lawler, 2001; Lawler, Thye, & Yoon, 2008) hypothesizes the particular conditions that make group members’ contributions to a task non-separable and, hence, foster a mutually affirming micro-social order. This theory was operationalized using the flow authority algorithm developed by Luo and Magee (2011), which conceives a network’s level of hierarchy as depending on directionality. These two indicators were aggregated into a *shared-responsibility score*, used as the independent variable in the model. A complementary analysis of the group’s discussion using a “bag of words” approach was also performed to define a sentiment score – the dependent variable – for each group.

The results suggest that mutual dependency and equal power distributions are not significantly associated with the sentiment expressed by groups in the discussions. On the contrary, hierarchical groups with a low sense of shared responsibility seemed more likely to develop positive emotions. This finding contradicts the theory of social commitment and suggests that members of dispersed small GitHub groups facing complex tasks frame hierarchy and the authority derived from professional

experience overwhelmingly positively. The theoretical consequence of this finding is that the interactional foundation of emotions could not be considered separately from framing processes related to identities and affective meanings. However, these results come with the caveat that we could not fully rule out the possibility that the sentiment analysis applied here failed to accurately measure emotion at the group level (cf. section 2).

3.6. Group Topologies on GitHub

One of the goals of the THEMIS.COG project was to better understand the GitHub ecosystem in general, in particular to describe “stylized facts” about social interaction patterns on the platform. We approached this problem by developing a typology of the groups that are present on GitHub. To this end, we studied the collaboration patterns of open source software projects by analyzing the pull request submissions and acceptances of repositories. We developed a group typology based on the structural properties of the corresponding directed graphs, and analyzed how the topology is connected to the repository’s collective identity, hierarchy, productivity, popularity, resilience, and stability. For details on this work, refer to Zöller et al. (2020).

The identified typology is displayed in Figure 2 and consists of five different types: 1) hierarchical, star-shaped graphs that are completely centered on one maintainer, 2) complex network structures that show reciprocity and clustering, 3) hierarchical, star-shaped graphs with several maintainers, 4) hierarchical graphs with a brokerage user that often decompose into two components and 5) very modular graphs that decompose into many components.

This clear evidence of patterned variation in group structures indicates that the dichotomy between free/libre open-source software (FLOSS) development as the bazaar and proprietary software development as the cathedral (Raymond, 1999), which has shaped the public perception, is no longer applicable for conceptualizing open source development or the social changes resulting from its wide adoption. The found typology partially maps to different stakeholders such as non-profit foundations, companies, educational bodies, or individuals and self-organizing groups, which are all present in the FLOSS development landscape.

This implies the hypothesis that the typology not only provides a picture of the interaction structure of the GitHub platform as such, but also (at least to some extent) of the institutions outside the digital realm. Examining a sample subset of projects from each cluster manually, we found that many projects in cluster 1 are headed by single developers, while the majority of projects in cluster 2 are run by non-profits and research collaborations and projects in cluster 5 are mainly backed by companies. However, these are mere tendencies. In particular, we found that projects backed by companies and non-profits are present in every cluster. Hence, not all companies structure their software development teams hierarchically, but different companies have different corporate identities and approaches. Likewise, although non-profits and foundations are often run by a democratic or decentralized board of stakeholders that shape the bureaucratic and strategic orientation of the organization, the wider organizational form is not necessarily reproduced on the task group level.

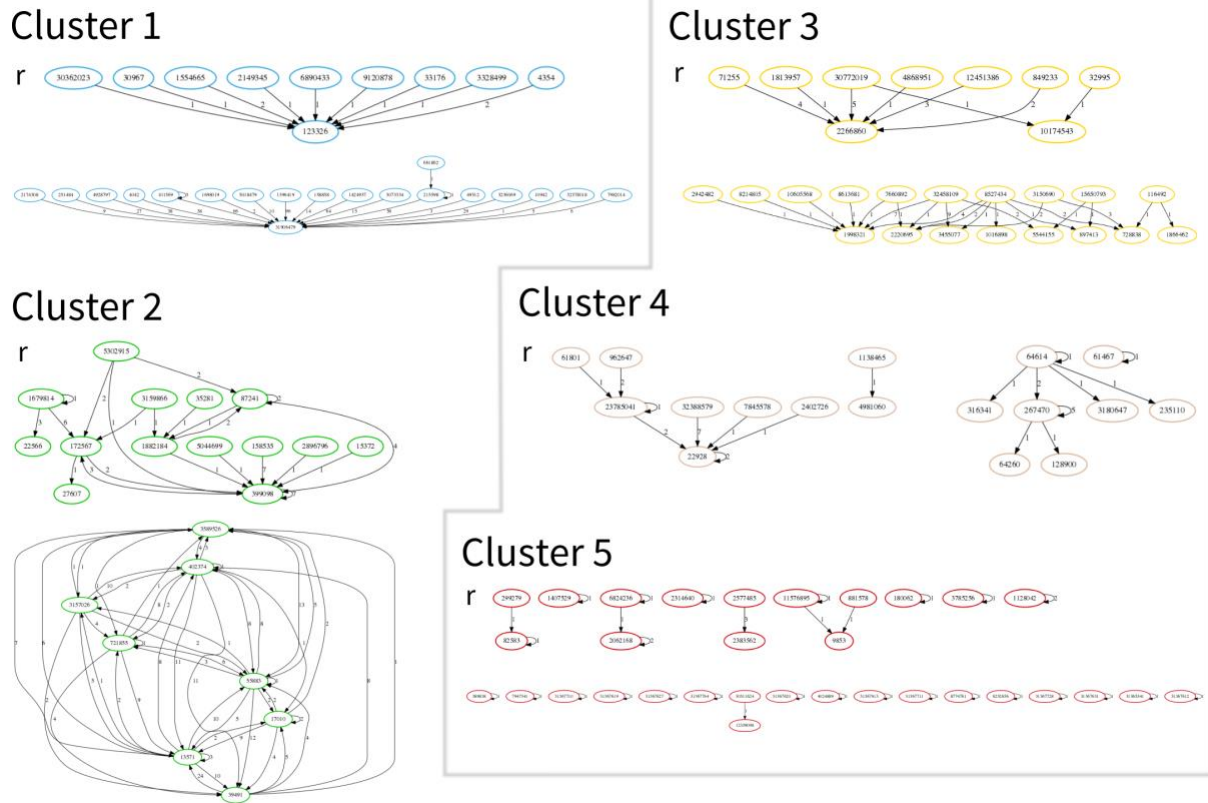


Figure 2. Typology of interaction patterns in GitHub projects. Reproduced from Zöller et al. (2020) with permission, copyright CC-BY-NC-ND 4.0.

Our findings indicate a strong correlation between the identified group structures and other outcomes of interest like the group identity, a project's popularity, productivity, resilience, and temporal stability. Group structures and outcomes are, to some extent, endogenous and self-reinforcing, with popularity and preferential attachment being the clearest examples. But often the direction of causality is ambiguous as in the case of productivity, which was significantly higher in groups with more complex interaction structures. While it might be that complex interaction structures facilitate productivity, it is equally likely that for very productive groups more complex interaction structures simply are more efficient and can become unavoidable.

Our analysis of temporal stability indicates that the majority of projects lose contributors over time, which can be attributed to the fast cycle of innovation in software development. Since the user base of GitHub grows exponentially, new competitor projects arise regularly that might use newer technology or address demands in the user base that are difficult to reconcile with the code base of an existing project. However, we found that the complex projects in cluster 2 have a higher probability to maintain their organizational structure. They also were the projects with the highest ratio of members still contributing a year after the initial analysis. Perhaps this finding reflects the fact that these projects seem to be backed by organizations, which can help manage the demands of sustaining a collaborative community. We also noticed a considerable flux to cluster 2 from projects of all clusters that were able to maintain their group size. While this suggests a tendency of resilient projects to build more complex structures over time, the results were not pronounced enough to indicate a clear progression towards one preferred structure, which supports the assumption that the group structure is influenced by other factors such as group identity, task complexity, personalities of core developers or personal preferences.

The found typology of GitHub groups based on the topology of their PR-graphs is potentially meaningful for practical, not just analytical, purposes, as it can help software developers make more informed decisions. Lead developers and repository maintainers in charge of organizing a FLOSS project can learn from the shown correlations between outcomes and group structure what kind of topology fits best with the goals they are trying to achieve and the community that is behind the software project they are trying to build. Volunteer developers can make a more informed choice on what kind of project they want to contribute to, depending on what kind of group structure they want to be part of and what roles in these projects might or might not be available to them in the future. Also, the results are relevant to users who are considering using and relying on a GitHub project. In particular, the analysis of resilience and temporal stability is helpful to weigh the benefits against the risk of a project getting stalled in the future. More detailed results and discussions can be found in Zöller et al. (2020).

3.7. Geographical Mapping of GitHub Projects

Going beyond group-internal communication structures, The purpose of these analyses is to better understand how national and regional patterns of development influence collaboration on GitHub, and whether these differences translate to differences in activities. For these analyses, we focused on Germany and China, in line with the cross-cultural analysis of social interaction patterns performed in a different work package of the THEMIS.COG project (see section 3.8 below). These polities played different roles in the science and technology sector during the 20th and the early 21st century. Nevertheless, they share some important traits. Both states are regional hubs, and both initiated major reform programs to further develop their science and technology capacity at the end of the 20th century.

To better understand the distribution of GitHub repositories at the state level, we extracted geo-spatial data for all repositories who reported their location for the years 2012 through 2018 from GHTorrent, an offline mirror of data offered through the GitHub Rest API (Gousios, 2013). With these data, we were able to generate bivariate kernel densities based on the repository owners' coordinates. We plot these densities as isolines on each map. The map for Germany is displayed in Figure 3, while Figure 4 shows the map for China. We supplemented these data with city-level population data to provide a sense of what is organizing the system. Points on the maps are scaled by city size. Because the countries differ in size, the scaling of the distances and points are relative to each map.

The partition of Germany during the 20th century continues to influence the development of Germany's science and technology sector. As shown in Figure 3, these geopolitical realities are quite visible in the structure of collaboration on the digital GitHub platform. The former West Germany exhibits a larger and more interconnected ecology organized around large and middle-sized cities. In contrast, the former East Germany consists of a primary hub, the Berlin/Potsdam area, and a smaller hub organized around Dresden and the university cities of Chemnitz, Leipzig, and Jena. Other major metropolitan areas such as Rostock and Magdeburg remain outside of the ecology. Although we are confident in the overall pattern, the difference between East and West may be exaggerated because locations are self-reports.

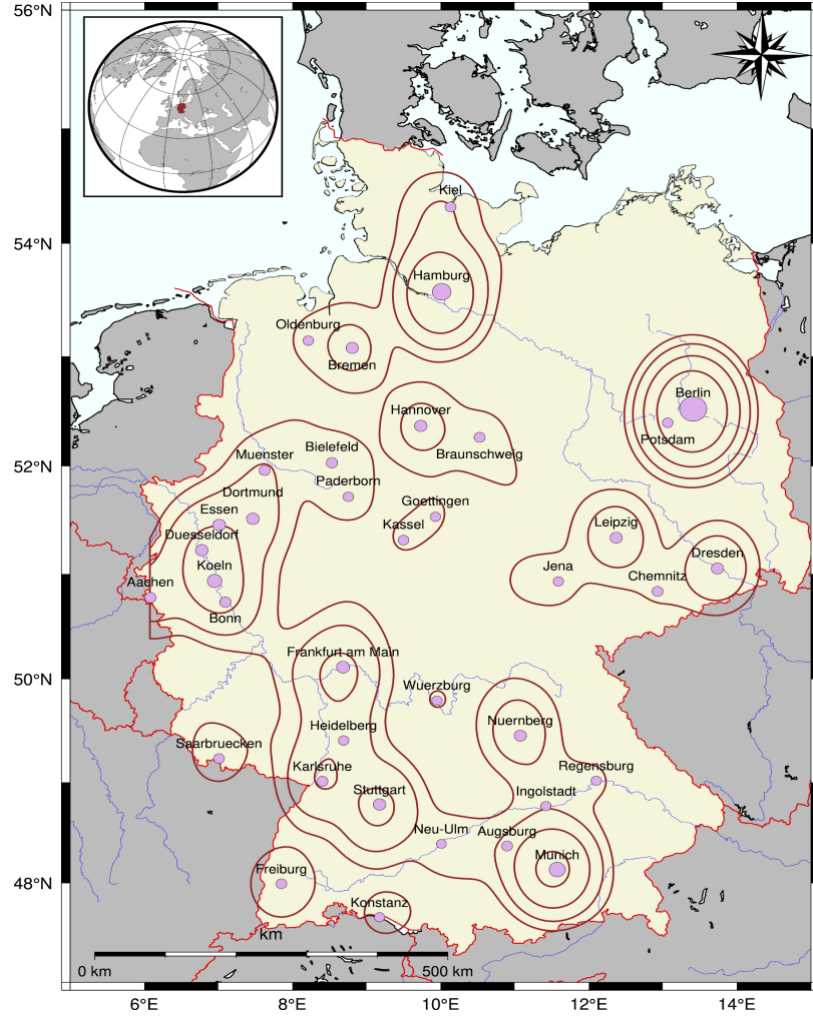


Figure 3. Geographical distribution of GitHub projects in Germany. Unpublished data.

Large cities organize China's GitHub participation to an even greater extent than in Germany. China exhibits three major hubs: 1) Beijing and Tianjin in the north, 2) the open coastal cities of Shanghai and Ningbo, and 3) the special economic zone cities of Shenzhen and its inland transportation hub Qingyuan in the south. Other major centers such as Wuhan and Harbin are important but isolated – these cities tend to be provincial capitals. The organization of China's GitHub ecology exhibits a stronger overlap between government, economic, and academic activities (i.e., there is no Chinese equivalent to the role played by smaller German cities such as Freiburg or Chemnitz).

Although we find that sub-state variation in socio-economic activity influences GitHub participation, we also find evidence of a common market when we examine activities on GitHub in terms of the computer languages developers use. To understand to what extent developers in China and Germany specialize in the same things, we calculated for each language how many repositories had code in that language (displayed in the bar plots shown in Figure 5), and for each pair of languages the number of users who wrote code in both languages (e.g., Python and C). For the co-occurrence analyses (displayed in Figure 6), we then adjusted the frequencies to account for chance by performing a Chi Square Independence test. We shade the co-occurrence matrices using the adjusted counts, where cold colors indicate counts less than expected based on chance and warm colors greater than expected. The x-axis of each matrix indicates the most frequent language in the pair for each developer, the y-axis the secondary language.

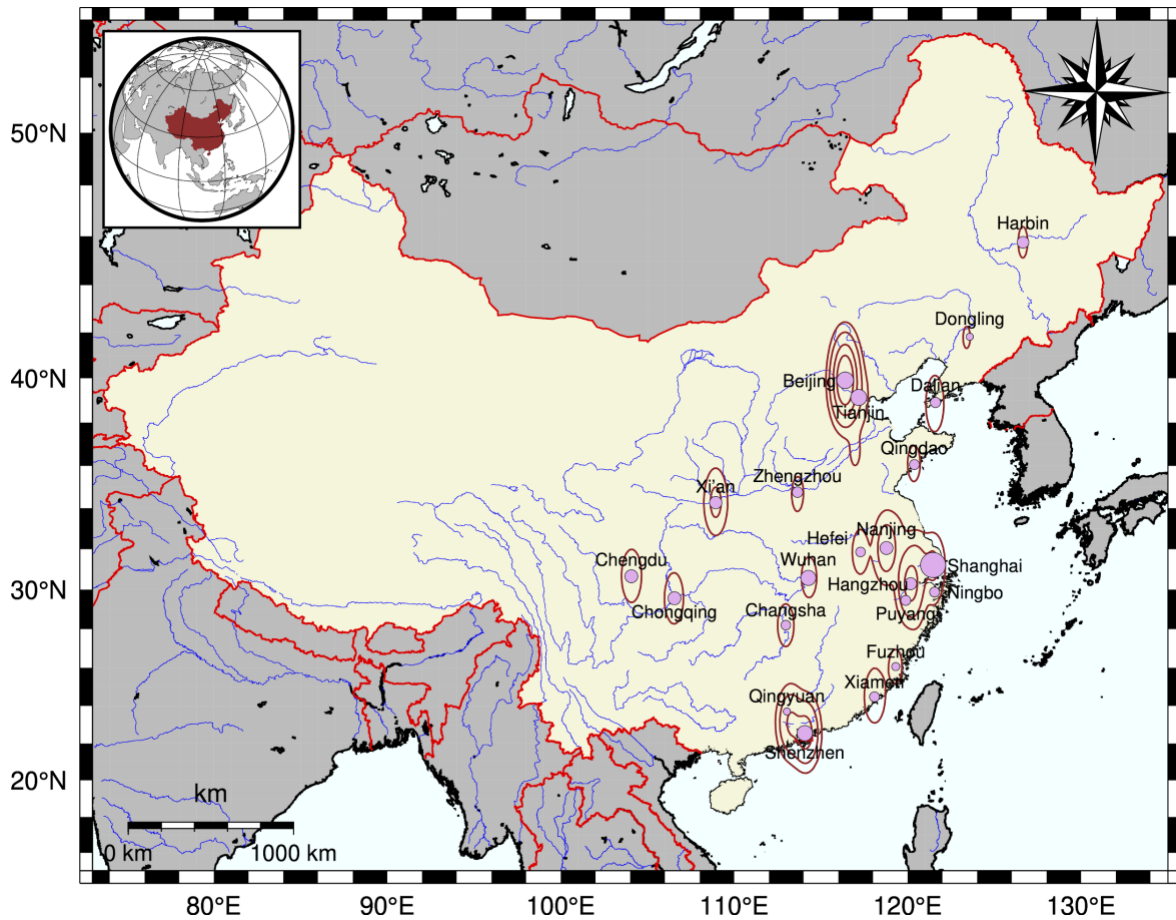


Figure 4. Geographical distribution of GitHub projects in China. Unpublished data.

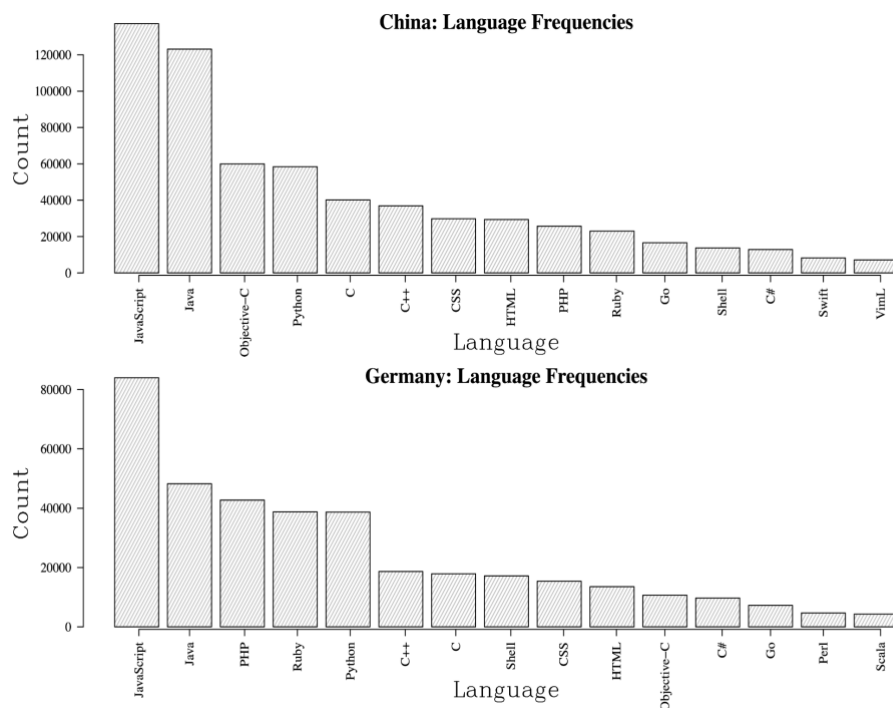


Figure 5. Frequency distribution of programming languages used in German vs. Chinese GitHub projects. Unpublished data.

We find similar rankings in terms of languages for both countries, with an overall Kendall rank order correlation of 0.64 ($p < 0.001$, two-tailed). We display here the top fifteen languages. For both China and Germany, Javascript and Java are the top two ranked languages. Although we find some differences in terms of the relative ranks, we find very similar lists. The rankings likely best reflect groups associated with institutions and organizations, as again we depend on self-reports.

When we look at the co-occurrence patterns, we find some evidence that Chinese developers are more likely to work in multiple languages than their German counterparts based on the shading of the off-diagonal cells; but overall, we find more variation in terms of languages than in terms of states. Not surprisingly, working in a language in the past is most correlated with working in the same language in the future. But, we find differences between the language communities in terms of the likelihood of their practitioners to work in other languages. For example, although Java and Javascript developers are the largest community, they are particularly unlikely to work in other languages if Java is their primary language, indicated by Java's darker blue column shading in both matrices. We also find asymmetries. For example, one of the few languages Java programmers are more likely to work in is HTML. In contrast, developers who work primarily in HTML are highly unlikely to work in Java. Unlike Java developers, Python developers are more likely to work in a number of other languages, Perl or Ruby being prominent examples. We find these patterns to be true in both countries. The commonalities between the two countries highlight common market forces and, more generally, a worldwide software market. The variation in geo-spatial patterns, however, possibly highlights the role state-level policy has on the capture and diffusion of these services (cf. Lee, 2018).

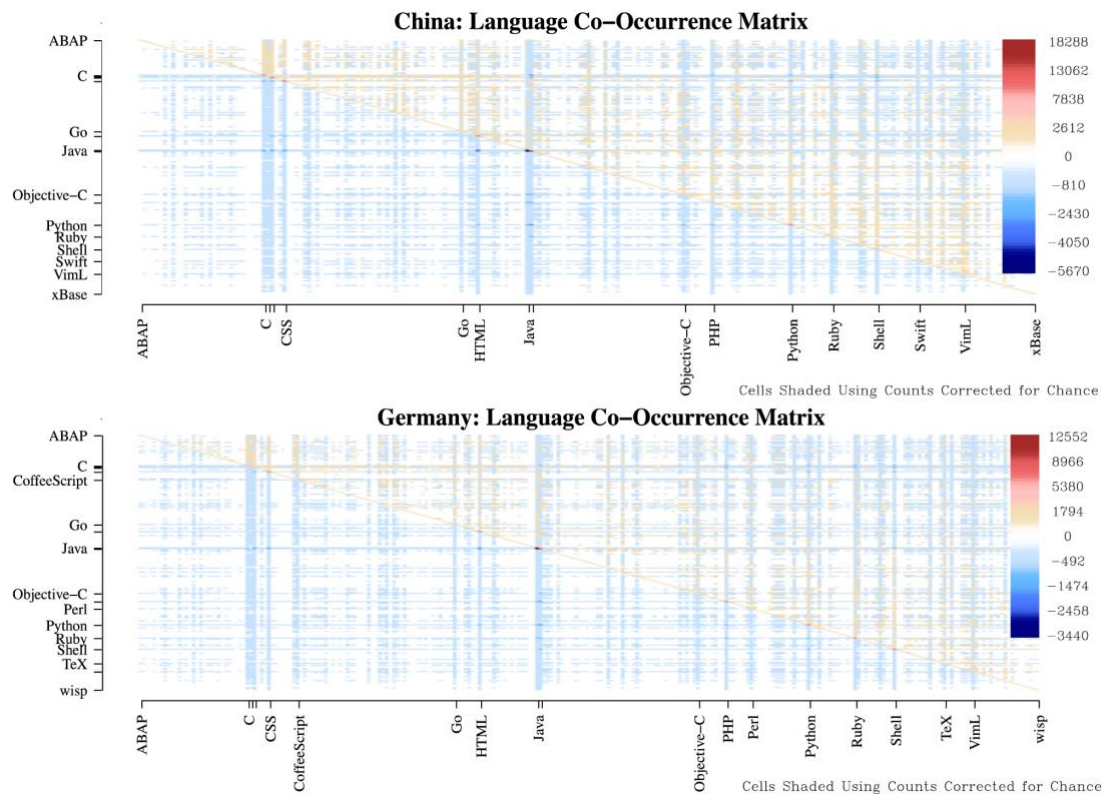


Figure 6. Co-occurrence of programming languages used in German vs. Chinese GitHub projects. Unpublished data.

3.8. Mapping Developer Subcultures

In order to apply our theoretical models of group dynamics to the study of self-organized collaborations on GitHub (see section 3.9 below), it was necessary to quantify a set of concepts relevant to these collaborations, including the identity labels (e.g., developer, project owner) and characteristics (e.g., experienced, novice) used to describe software developers and their work, and the behaviors that regularly take place between them (e.g., recruit, teach) or toward the code they have created (e.g., accept, reject). The meanings of these concepts were quantified through a sentiment study, in which GitHub users reported their sentiments for each concept along three affective dimensions: evaluation (good-bad), potency (powerful-powerless), and activity (lively-inactive). As explained in the introduction, these dimensions (collectively known as EPA) reflect universal dimensions of social perception employed by researchers across disciplines to describe cultures, and accordingly span the vector space in which affect control theory models operate (e.g., Heise, 2007, 2013; Hoey et al., 2016; Schröder et al., 2016).

503 GitHub users were recruited through Qualtrics Panels and responded to survey questions online. As GitHub users are predominantly white males, we oversampled women and non-white users to enable predictions about inequality. The sample was 50% female and 30% non-white. All respondents were North American English speakers aged 18+. We constructed a list of 524 concepts to be rated after a comprehensive review of prior work, and a quantitative analysis of the words most commonly used to describe identities and behaviors in conversations on GitHub. This list contained a high concentration of concepts we expected to have divergent meanings in the mainstream U.S. culture and GitHub subculture. We gathered sentiment ratings on the same dimensions for two other sets of items in this study: 1) vignettes describing group structures found on the site, their owners, and their contributors; and 2) vignettes describing people who display particular values and behavior, derived from Bales' Interaction Process Analysis (1950, 1999; cf. section 3.1 above). Lastly, we asked a series of questions about respondents' demographics, GitHub usage, and work on coding projects.

We supplemented this effort with a second data collection round where we collected data on a core subset of identities and behaviors. This second round was necessary because we found that behaviors such as *reject* were too underspecified in the original survey to support modeling online interactions. While rejection is rare and negative in a general social context, it is normal in a development context; rejecting substandard code is a good thing, whereas arbitrarily rejecting high quality code is not. The second data collection effort involved collecting ratings for 18 concepts from 20 individuals who we confirmed as active GitHub users.

Because we found that the concepts evoked sentiments associated with common and specialized cultural meanings, we developed a method for examining meaning consensus where the sentiment is likely to have two or more modes. We focus on modes that consist of at least 10% of respondents because these indicate potential violations of ACT's cultural consensus assumptions. Our analysis identifies modes in respondents' ratings for each concept, enabling researchers to discern which concepts carry widely consensual meanings (one mode) and which are more divisive (two or more modes). The method also enables us to determine: 1) which modes reflect majority vs. minority rating patterns by calculating the proportion of respondents whose ratings align with each mode; 2) whether each respondent's ratings most often align with majority or minority modes; and 3) whether alignment with majority vs. minority modes is associated with other respondent characteristics measured in the study. The method enables researchers to identify sources of meaningful multi-modality, which does not arise simply from rating error.

In the GitHub data for example, we identified many concepts with unimodal, normally distributed meanings; however, 163 of the 524 concepts rated had more than one mode (i.e., cultural meaning) on

the evaluation (good-bad) dimension, for instance. Respondents who use GitHub more frequently were significantly more likely to identify with the minority, subcultural mode, while GitHub tenure (e.g., when respondents joined GitHub), gender, and nationality did not predict inclusion in a minority mode.

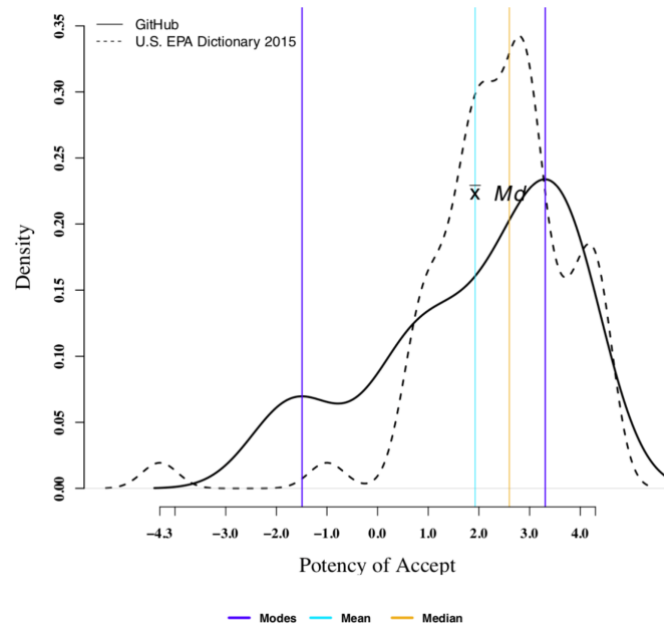


Figure 7. Example distribution of empirical sentiment ratings (here: potency ratings for the concept *accept*). Unpublished data.

As an example, Figure 7 displays one concept for which ratings were divided along the potency (powerful-weak) axis of sentiment, *accept*. While the majority of respondents rated this concept quite high in potency (3.3), a smaller contingent rated the concept quite low in potency (-1.5). This is perhaps because of the dual denotative meaning of this concept on GitHub, where acceptance can refer to a person being integrated into a group (dominant cultural meaning) or a code revision being integrated into a piece of software (subcultural meaning). Using the mean to represent the central tendency of respondent ratings would be quite misleading in this case; because it incorporates ratings from the minority mode, the mean value (2) is substantially lower than the central tendency of the dominant mode. Despite the bimodality of this distribution, the sample mean is quite similar to those found in earlier sentiment studies. Different measures can be used depending on research goals. While isolating the central tendency of the dominant mode may be most appropriate for consensus studies, comparing modes and the respondents that fall closest to them may be more appropriate for studies focusing on variability or subcultures. Mean and median are only sound choices for unimodal meaning distributions.

3.9. Modeling group Dynamics in the U.S. vs. China

Online collaborative platforms such as GitHub have become increasingly popular for advancing digital societies worldwide. As informal collaboration within geographic boundaries increases, so does the frequency of self-organized collaboration between people from different countries and cultures (for geographical analysis of collaboration patterns, cf. section 3.6). To foresee challenges and opportunities arising from these cross-cultural collaborations, researchers need to know how people interact with members of their own culture as well as cultures other than their own.

This study was theoretically grounded in affect control theory and set up to 1) identify the cultural rules that govern social interactions in two societies: China and the United States; 2) populate theoretical simulation models for group dynamics with culture-specific parameters to simulate and compare interactions in each cultural context; and 3) substantiate model predictions by comparing our simulation results to empirical data scraped from GitHub groups composed of North American English speakers and of Chinese speakers.

Using archived sentiment data from China and the classic impression change data from the United States, we first estimated impression change equations for each culture (Zhao, in press). Findings suggested a fair amount of cross-cultural commonality in the affective processing of cultural concepts, providing support for the assumption that determinants of impression change are largely universal. Both cultures display a strong tendency to reinforce normative meanings about identities and behaviors in situations. The processes of psychological consistency and congruence are also present and comparable in magnitude across the two cultures, reflected in the two cultures' impression change equations. On the other hand, results revealed significant cultural variations in identity plasticity, suggesting more room for contesting identity meanings among Chinese than among Americans. We also find differences in obtaining social status across the two societies. Americans reward kindness with status, but only if the recipients of their actions occupy culturally positive identities, perhaps due to their ingrained sense of social justice (see Figure 8). In contrast, Chinese believe that acting nicely to stigmatized others enhances, rather than sullies, the reputations of those who are socially esteemed, a belief that may be rooted in the philosophy of Taoism.

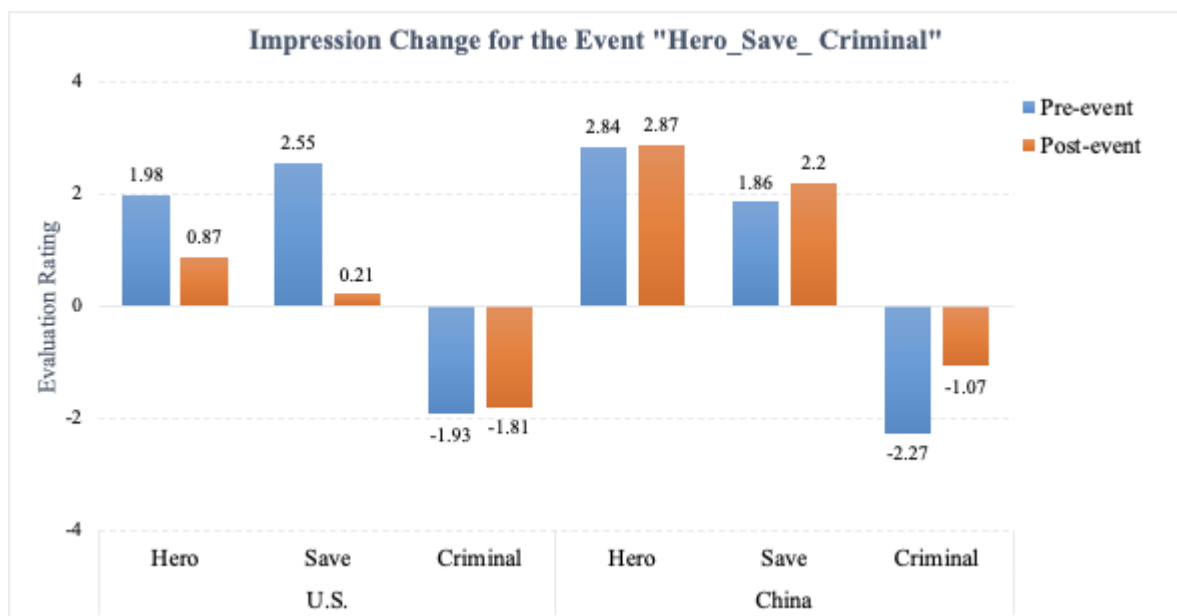


Figure 8. Cross-cultural differences between the U.S. and China when interacting with stigmatized group members. Reproduced from Zhao (in press).

Our research in progress examines the implications of these findings for cross-cultural differences in group dynamics, with a specific focus on conflict resolution. We ask, how do cultural differences in impression formation dynamics affect how intra-group conflicts are resolved in North American vs. Chinese workgroups? We address this question through a simulation experiment using our models of group interactions, populated with culture-specific impression formation parameters and linked with

data repositories that contain sentiment ratings of key identities and behaviors in each culture under study, gathered through original survey research (see section 3.7). Model predictions will be made about behavior, emotions, and affective (mis)alignment in simulated groups from each culture. We will validate these predictions by comparing them with empirical data on conflict resolution scraped from GitHub project teams composed of North American English speakers and Chinese speakers. Comment data scraped from GitHub will be analyzed with *IBM Watson Personality Insights* to infer developers' personality traits then the procedure discussed under 3.2 above will be applied to convert these into estimated sentiment profiles for each group member. These data will be linked with data on members' comments, acceptance vs. rejection of pull requests, and projected emotions based on emoji usage.

3.10. Modeling Egalitarian vs. Hierarchical Group Cultures

Culture in the workplace is a multi-faceted construct. Besides “developer (sub)cultures” (section 3.8) and national cultures (section 3.9), we sought to understand variation in group/organizational culture across GitHub projects. We focused mainly on the dimension of hierarchy (where there is an accepted “chain of command” in a group) vs. equality (where coordination relies on negotiation among equal peers). As the developed typology of groups (section 3.6) shows, this distinction is clearly visible in the communication structures of GitHub projects. The theoretical simulation work reported here was aimed at analyzing how these different structures translate into affective experiences that help developers navigate group collaboration.

In a first study (Hoey et al., 2018) employing an earlier group-dynamics model based on affect control theory developed by Heise (2013), we found that, from an identity dynamics perspective, hierarchical groups are better able to meet group member expectations, primarily because power differences encoded in the sentiments of the identities result in fewer violated expectations about who should act on whom. This is captured in the simulation as lower levels of affective deflection (tension). Employees expect to take direction from bosses, while friends who tell other friends what to do are bossy.

In the 2018 study, we kept relational norms governing how often group members addressed the group, and how often group members reciprocated an action constant. Small groups researchers, however, have found that relational norms are important moderators of group behavior. Greater intra-group reciprocity tends to weaken dominance hierarchies by fragmenting the focus of attention, while the ability to address the group tends to strengthen the leader's potency (Gibson, 2010). Consequently, we wanted to learn how changing these constants influenced our findings. We also wanted to see if identity maintenance constrains the application of relational norms, particularly norms governing the interaction opportunities of leaders. Given that both leaders and followers need opportunities to affirm their identities, is there an optimal mix of appeals to the group and exchanges with group members that meets the identity needs of both followers and leaders?

We compared the effect of addressing the group and reciprocation on the level of intragroup tension by conducting two simulation experiments (for details, see Morgan et al., 2021). We employed a 2x2x11 design, comparing two group types (status-equal and hierarchical) with respect to the proportion of actions directed towards the group and the proportion of actions reciprocated by group members across eleven settings. We conducted 200 simulations of each group type/setting combination. When we varied the address-the-group rate, reciprocity was held constant at 0.8; when varying reciprocity, actions directed at the group were held constant at 0.4. For both experiments, we held group size constant at three; all other parameters were set at their default values as per the original affect control theory model of group dynamics (Heise, 2013).

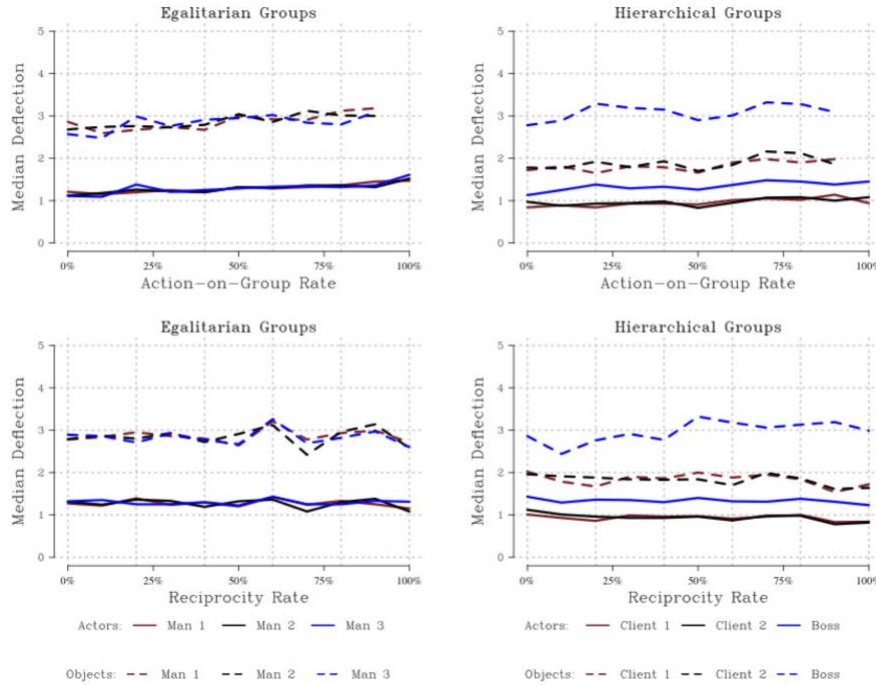


Figure 9. Effects of hierarchical vs. egalitarian group cultures on experienced affective deflection (tension). Reproduced with permission from Morgan et al. (2021).

Selected results are shown in Figure 9 (for more details on methods and results, see Morgan et al., 2021). We did not find a mix of reciprocity and appeals to the group that led to lower levels of tension across the board. We did find that hierarchical groups are better able to meet group member expectations than status-equal groups across a range of settings, corroborating the earlier tentative finding (Hoey et al., 2018). It is important to note that meeting expectations is not equivalent to happier interactions but less tension about the exercise of authority. More surprisingly, although our simulations confirmed that the majority of tension is experienced by individuals who are acted upon by others, they also highlight the tension experienced by leaders. Even when they dominate the interaction, leaders experience more tension than subordinates because they must maintain a perception of both power and goodness. In short, hierarchies where the leader must be *both* good and potent provide protections that status-equal relationships do not necessarily provide to weaker identities because hierarchies are constrained by sanctions against bullying that do not apply to nominally status-equal groups.

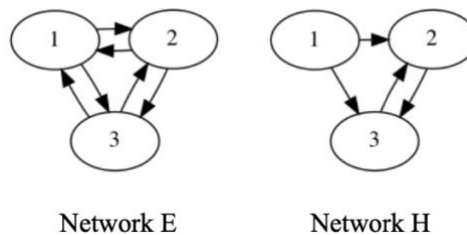


Figure 10. Constraints on interactions in stylized egalitarian (E) vs. hierarchical (H) groups. Reproduced from Sedlacek (2020).

In a further study (Sedlacek, 2020), we systematically examined the tradeoff between esteem (evaluation) and power (potency) by analyzing the tension agents experience and the actions they take in different group settings. Social psychologists have observed this tradeoff in the *warmth/competence dilemma* faced by female leaders (Cuddy, Glick, & Beninger, 2011), in affect control theory's impression change equations (Morgan, Rogers, & Hu, 2016), and perhaps most famously in Machiavelli's advice to princes about the merits of being loved versus feared. To account for the effect of institutional arrangements, we compared agents interacting in hierarchical and egalitarian groups; these groups differ from each other in terms of the ability of each group member to act on the others. In hierarchical groups, leaders can act on followers, but followers can act only on each other. In egalitarian groups, everyone can act on everyone else as shown in Figure 10.

Within each network, we ran a series of simulations examining the evaluation and potency distance between Agent 1 and Agents 2 and 3. Agent 1 occupies a structurally more powerful position in hierarchical networks (H) because Agent 1 can act on others but cannot be acted upon. In analyses of sentiment distance, Agent 1 was assigned greater evaluation or potency than Agents 2 and 3; the sentiment distance between Agent 1 and Agents 2 and 3 was increased by increments of 0.5 starting at a value of 0 for both evaluation (with potency and activity held constant at 1.0) and potency (with evaluation and activity held constant at 1.0). All simulations used sentiment ratings from the study of developer subcultures described above (3.7) and impression change equations for North American English speakers. The reciprocity rate was held constant at 0.8, the rate commonly observed in task groups. The address-to-group rate was held constant at 0.4.

We begin by discussing the effect of evaluation distance on tension. We found that increased evaluation distance between agents creates greater misalignment than increased potency distance in both networks. Who experiences the greater misalignment, however, differs in the two networks. As evaluation distance increases, Agent 1 (the "good" agent) experiences greater tension than Agents 2 and 3 (the "bad" agents) in the egalitarian networks. In contrast, Agents 2 and 3 experience greater tension than Agent 1 in hierarchical networks. When varying potency distances, we find that Agent 1 experiences greater tension than Agents 2 and 3 (the "weak" agents) in both hierarchical and egalitarian networks.

Besides the analysis of affective deflection (tension), we looked at the functional meaning of the predicted behaviors in both network settings. Again, we relied on Bales' (1999) Interaction Process Analysis as a classification scheme (cf. section 3.1). Simulation results are shown in Figure 11. In both egalitarian and hierarchical networks, negative socio-emotive behavior (e.g., disagreement, antagonism) increases and positive socio-emotive behavior (e.g., agreement, solidarity) decreases as the evaluation and potency distances between the agents grow. Negative socio-emotive behaviors dominate group interactions as evaluation distance increases, as some agents come to occupy "bad" roles. This is especially true in hierarchical networks, where negative socio-emotive behaviors comprise around 90% of actions at an evaluation distance of 4+.

Active task behaviors (e.g., give suggestion, opinion) are common in both network types at low levels of evaluation distance, but become rare in hierarchical networks at evaluation distances greater than 4. In contrast, these behaviors are common in egalitarian networks even as the evaluation distance grows. Passive task behaviors (e.g., ask for suggestion, opinion) become rare in both network types as evaluation distance grows. As potency distance increases, active task behaviors become *more* common in egalitarian networks, but remain fairly consistent in hierarchical networks. In contrast, passive task behaviors remain constant in egalitarian networks as potency distance increases, but comprise a growing proportion of the group's behavior in hierarchical networks.

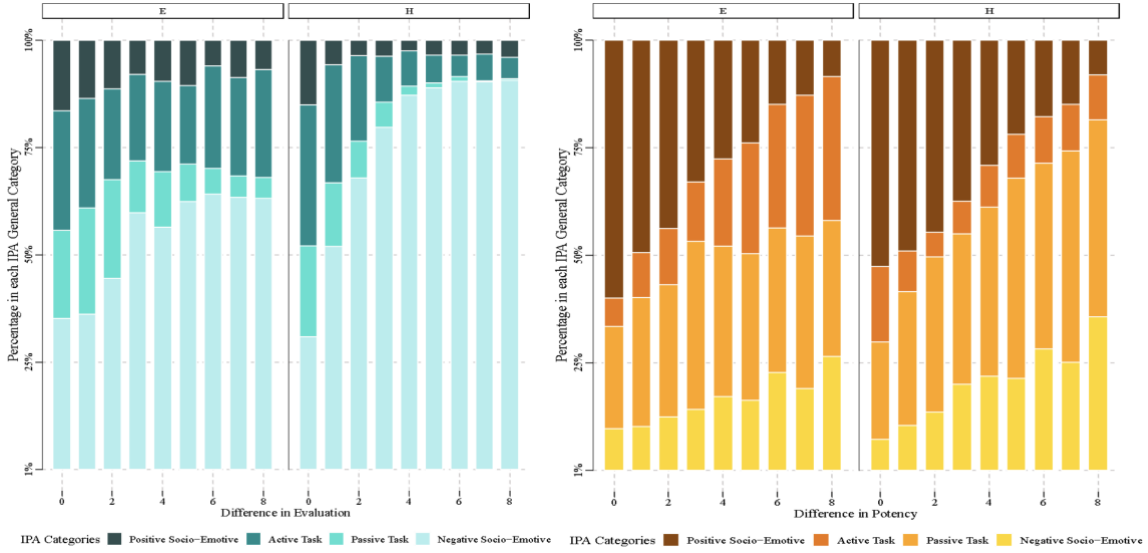


Figure 11. Distribution of group behaviors as a function of evaluation (status) vs potency (power) distance in collaborative groups. Reproduced from Sedlacek (2020).

These results suggest that the legitimacy of a status order is particularly sensitive to group members’ perceptions of their mutual goodness, while mismatches between an agent’s structural role and their perceived potency play a lesser role. Although intuitive, this result was not necessarily obvious, given that violations of a person’s agency are a major source of intragroup tension. A limitation of this work is that, given the impossibility to classify Bales’ IPA categories with actual GitHub data (see sections 2 and 3.1), we were not able to develop empirical validation of these simulations despite many efforts.

3.11. Modeling Network Structure, Status Expectations, and Task Protocols

Simulations described in the previous section were performed with Heise’s (2013) older group dynamics model based on affect control theory. The model has a variety of limitations for the purpose of simulating collaboration of software development teams. One limitation is technical, as the model is implemented in NetLogo, a specialized language for agent-based modeling which is hard to integrate with more general data science tools as employed in the THEMIS.COG project. We therefore developed a new affect control theory based group dynamics model in the more widely known Python language, using the Mesa framework (<https://mesa.readthedocs.io/en/master/overview.html>). We named the model ACTING, which is an acronym for Affect Control Theory based simulations of Interaction in Networked Groups. Details on the model can be found in Zöllner et al. (in press). ACTING also has important new features for modeling group collaboration on GitHub or in similar environments in ways that are more realistic than the highly stylized simulations described in the previous section.

First, we allow the analyst to specify a network topology that restricts possible action pathways in the group which can be used to model hierarchical role configurations. This extension was motivated by our work on the group typology (section 3.6) showing that, in online task groups, interactions occur solely along network structures. Second, we allow the analyst to specify a discrete set of actions for agents to choose from, which caters to the fact that in many group situations only a limited number of behaviors are admissible, acceptable, or available. Third, we allow the analyst to specify sequences of actions that have to be carried out in a specific order. These protocols are often observed in task groups

where the completion of a task requires several steps of role-based division of labor (e.g., in the case of GitHub, the pull-request mechanism for quality control of changes in the code base).

In a series of simulations, we found that network positions have an influence on agents' experienced affective deflection (tension) and their transient identities, particularly along the dimension of potency. This finding highlights the interconnection between structural properties of groups and group members' cognitive-affective representations of group interactions. By expressing status hierarchies in terms of node positions in a network topology and interpreting deflection as a metric for cultural acceptance, we were able to build a bridge between expectation states theory (Ridgeway, 2019) and affect control theory, contributing to theoretical unification in social psychology.

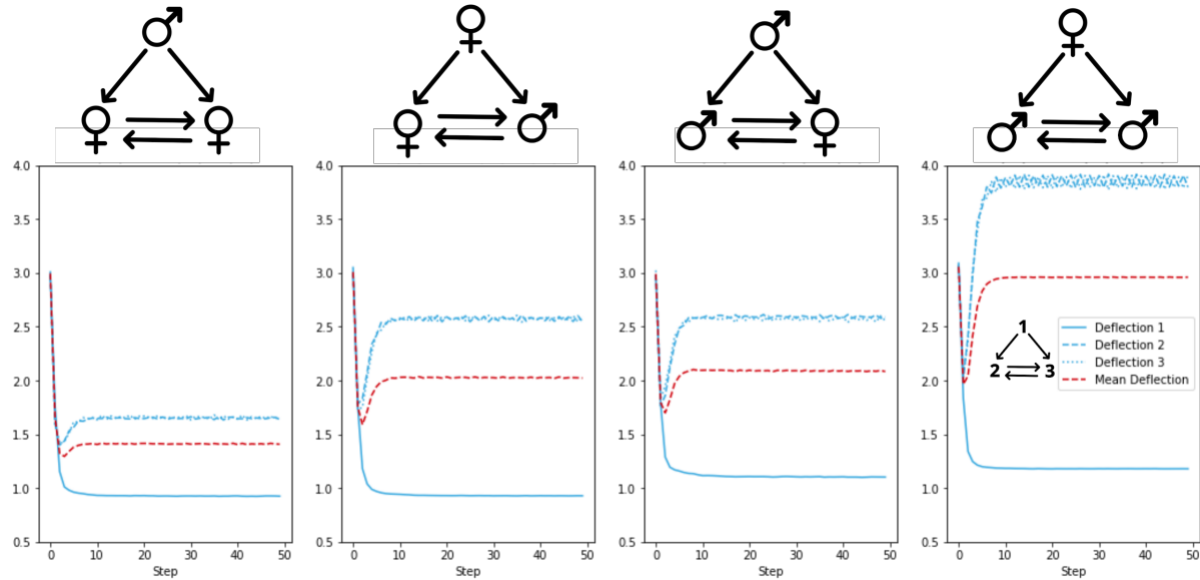


Figure 12. Simulation of gendered status differences in task groups emerging from affect control mechanisms. Reproduced with permission from Zöller et al. (in press).

In particular, we succeeded at reproducing core results about gender stereotyping from the expectation states framework (Ridgeway, 2019). The simulations of idealized, mixed-gender task groups (see Figure 12) suggested that, regardless of competence, men were on average culturally more accepted in leadership positions than women, which in the simulations could be traced back to the encoding of cultural norms in affective identity labels. Agents chose their behaviors in such a way as to align them with the cultural expectation of the group about their role identity. The average deflection in the group was significantly higher when female rather than male agents occupied leadership positions.

This result suggests that in order to promote gender equality sustainably, the cultural norms in which female identity is generally perceived to be less powerful and more positive than the male identity would have to be adjusted. From a symbolic interactionist point of view, cultural meanings arise and are consolidated through interactions. That means that by continuously promoting and normalizing female leadership and by increasing its visibility, the perception of female identity possibly can be changed in the long-term future through changes in the cultural affective meanings of female identities.

As an example of a realistic complex task group, we modeled in a series of simulation experiments groups consisting of open source software developers with their domain-specific work flows, based on the extracted network structures from GitHub in a different work package of the THEMIS.COg project (see section 3.6). We conducted a small survey among verified GitHub users in which we had the most

important role identities and task related behaviors rated on affective scales. This subculture and domain-specific data served as an empirical basis to parameterize the simulation model.

The simulations showed that, in these groups, average deflection was significantly lower if high status network positions, signifying more responsibility and the power to accept or reject proposed changes to the code base, were filled with experienced developers instead of novice developers. This result is in accordance with the presumed meritocratic value system in the open source software development community which is essential for the process of self-organization in groups that are not backed by a larger organization (more details are discussed in Zöller et al., in press).

We believe that this is a good example for how our model can be adapted and generalized to the study of groups in other domains and subcultures, which we think is a promising use case of the model for future research. An adaptation process will generally consist of the following steps. First, one will have to research the task domain and its subculture of the groups of interest. This might include the identification of an organizational structure that guides the group interactions. Furthermore, this is likely to include the identification of important role identities and behavior sequences that are characteristic or necessary for completing core tasks or conducting other focal interaction protocols in the group. Second, one can then design a mini-survey to collect affective ratings of these concepts. These empirical findings can subsequently be injected into the simulation setup which can then be used to illuminate affective group dynamics, generate hypotheses, or compare different role constellations. In this way, it is possible to systematically analyze the interplay between status characteristics, role identities, and cultural norms and expectations.

3.12. Modeling Uncertainty in Group Processes

The modeling work described so far is entirely targeted at social coordination in groups as driven by affect control processes. While this is an important angle for understanding collaboration in software development and beyond, it is also limited in that it disregards the requirements of the task and the deliberate problem-solving requirements that collaborative groups are subject to. We therefore also undertook an effort towards developing a dual process model that besides affective (or as we put it in this context: intuitive) processes can account for deliberative decision-making. Due to time and resource constraints of the THEMIS.COG project, the model currently operates only at the level of one individual, while scaling it to the level of group dynamics (e.g., through integration with the ACTING model described in section 3.11) must remain for future work.

The interplay between intuitive and deliberative processing is known to be important for human decision making. As independent modes, intuitive processes can take on many forms from associative to constructive, while deliberative processes often rely on some notion of decision theoretic rationality or pattern matching. Dual process models attempt to unify these two modes based on parallel constraint networks or on socially or emotionally oriented adjustments to utility functions. Our work developed a new kind of dual process model that unifies decision theoretic deliberative reasoning with intuitive reasoning based on shared cultural affective meanings in a single Bayesian sequential model (Hoey, MacKinnon, & Schröder, 2021). Agents constructed according to this unified model are motivated by a combination of affective alignment (intuitive) and decision theoretic reasoning (deliberative), trading the two off as a function of the uncertainty or unpredictability of the situation. The model also provides a theoretical bridge between decision-making research and sociological symbolic interactionism. Starting with a high-level view of existing models, we advanced a hierarchical Bayesian model that includes affective meanings as control mechanisms as a promising new type of dual process model that explicitly and optimally (in the Bayesian sense) trades off motivation, action, beliefs and utility. We

demonstrated a key component of the model as being sufficient to account for some aspects of classic cognitive biases about fairness and dissonance, and outlined how this new theory relates to parallel constraint satisfaction models.

These theoretical advances, borne out in analytical simulations, demonstrate the explanatory power of an affect-based model of decision making. Our long term goal is to apply the same concepts to larger scale groups than dyads that are handling more complex situations than a simple dissonance. Applying the model to GitHub may be less productive not just because of a lack of relevant data on social and affective processes (cf. sections 2 and 3.1), but also because of the structure that is “imposed” in some sense on GitHub users. For example, the role of “maintainer” explicitly puts one individual in a position of power that cannot be disputed while possible behaviors are constrained by protocols embedded in the platform design (cf. section 3.11); ergo, there is little uncertainty that needs to be dealt with through an intuitive/affective mode of reasoning as described by our model. However, in other collaborative groups, in particular those that come together for the first time, tools that are able to provide a “readout” of the emotional landscape and dynamics of the group may be very useful for promoting collaboration, particularly in social dilemmas and societal transformations such as those related to climate change or rapid technological change (Hoey & Schröder, under review; Tilbury & Hoey, 2020).

3.13. Work in Progress: Dynamics of Innovation

A final work package of THEMIS.COG was devoted to look at the interaction dynamics across different GitHub groups. Building on our understanding of the platform’s innovation ecosystem (see sections 3.6. and 3.7) and the affective mechanisms driving people’s interactions on the platform (see sections 3.10 and 3.11), we sought to deepen our understanding of how the entire GitHub graph evolves dynamically as individual developers make decisions about projects to join or abandon and thereby cause skills and ideas to spread. Due to resource constraints and unanticipated problems in prior work packages (cf. section 2) we were not able to advance this part as much as we had hoped in the beginning of the project, but we were able to zoom in on the problem of innovation dynamics on GitHub in two specific ways.

A first perspective is the study of popularity dynamics as a potential driver of the evolving ecosystem. GitHub’s user interface offers developers the possibility to star repositories. This feature is mainly used to show appreciation to projects and to bookmark projects for later retrieval. Additionally, developers might star projects that they are currently using or have used in the past. The amount of collected stars is generally seen as a measure of popularity, but users also take it as a signal for quality and a marker of a project’s success. Furthermore, maintaining or contributing to a project with a high star count provides prestige and status among peers in a culture whose value system is presumably built on meritocracy. We have started to analyze with a combination of agent-based modeling and reanalysis of the data on the GitHub typology (section 3.6) to explain the dynamic interplay between individual popularity and success of GitHub projects. This work is ongoing at the time of writing this report; therefore, results cannot be reported yet.

A second approach to understanding GitHub as an innovation ecosystem, which we developed, is to look at the emergence of programming languages as an example of innovation diffusion. The GitHub data set gives us the opportunity to look at how these languages spread, percolate, and fade across the larger user community. Social network analysts have studied how a variety of phenomena spread across populations, including but not limited to how diseases, ideas, practices, beliefs, and norms spread among connected individuals. Specifically, the data from the GitHub data set allows us to look not only at how a single language spreads out across a network of users, it gives us a system-level overview of how

competing and complementary innovations rise, spread, percolate, and fall. Languages increase (and decrease) in popularity quite rapidly. As an example, the rapid expansion of the Julia language in the GitHub community is shown (by use type) in Figure 13.

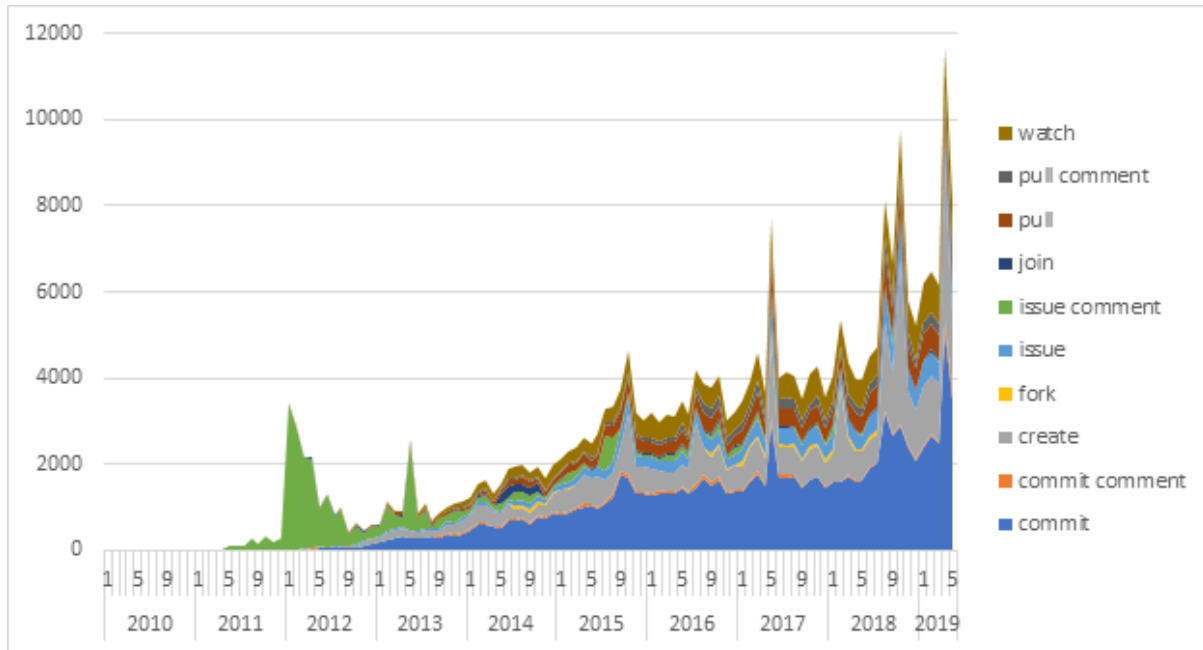


Figure 13. Distinct GitHub projects employing the Julia language by use type for each month from January 2010 to June 2019. Unpublished data.

As a result, we can simultaneously analyze two aspects of diffusion that have been considered at the level of the small group, but less so at the level of broader social networks. First, we can consider how the hierarchical dynamics of groups shape the spread of diffusion within groups. Is one new person evangelizing an innovation and bringing it to large groups, or are groups more likely to consist of users that have previously used an innovation, bringing on converts only one or two at a time? Classic work in the small groups literature has found that group characteristics such as size, structure, the position of the potential and previous adopters all influence whether a group adopts an innovation. The GitHub data set allows us to tie these local diffusion processes into more global studies of networks.

Second, we can consider how innovations are spreading with or against one another. Are certain languages “latching” on to other languages? Recent work has shown how cultural beliefs spread with and against one another in small groups, leading to polarized sets of cultural preference even when groups are fully connected (Goldberg & Stein, 2018). The GitHub dataset allows us to assess whether innovations latch on to one another, compete with one another, and allows us to consider the broader network of innovations themselves, adding a new layer of detail to our understanding of the diffusion of innovations. An example of this is shown in Figure 14, which shows how GitHub languages co-occurred with one another in the year 2017. In this network, each node is a language, and each edge is weighted by the Jaccard coefficient of the set of users of each language.

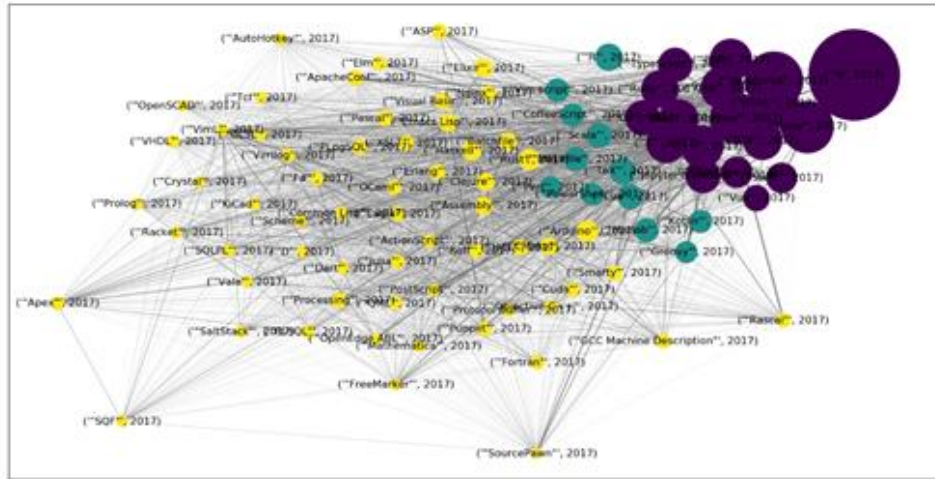


Figure 14. Network of programming languages in the GitHub community in 2017. Each node is a language, and each edge is the Jaccard coefficient of the user sets of the connected languages. Colors represent communities detected via the Louvain algorithm.

Taken together, we hope to analyze how the micro-aspects of language diffusion within and between groups can help us understand the broader patterns of innovation diffusion in the GitHub community, informing broader understanding of how cognitive mechanisms and status processes influence not only the way that diffusions spread, but the way they spread in relation to one another.

4. Outlook

The joint work on the THEMIS.COG project detailed in the previous sections has allowed us to develop and strengthen our collaborative ties across the disciplines and universities involved. We expect, and have already started, to sustain this collaboration beyond the end of the project. Some areas for continuation or extension of the work were already outlined; e.g., the cross-cultural comparison of group dynamics between the U.S. and China (see section 3.9) or the work on popularity dynamics as a driver of innovation diffusion on the GitHub platform (see section 3.13), which is part of Nikolas Zöller’s still ongoing PhD thesis work (cf. section 6 on training).

More future work ensues naturally from combining project results described above. For example, embedding the individual-level Bayesian affect control theory model of deliberative vs. intuitive decision-making (section 3.12) into the ACTING model of group dynamics (section 3.11) would yield a more comprehensive theory of task groups than either approach alone. As another example, the coefficients of the American vs. Chinese impression-change models (section 3.9) could be used to parameterize culture-specific variants of the ACTING model, in order to derive predictions about cross-cultural misunderstandings.

Another area for future work might be to apply the concepts and technology developed in THEMIS.COG to other datasets than the one from GitHub, which might prove more amenable to the affect control theory framework as they contain more social and emotional information than the predominantly technical discussions in GitHub pull-request and issue comments (cf. Corritore et al., 2020). While the project results are not mature enough for immediate commercialization, in the future,

some applications are certainly conceivable (e.g., in organizational consulting especially as more and more businesses move their operations online).

5. Publications related to THEMIS.COG as of August 2021

- Hoey, J., & Schröder, T. (under review). *Disruption of social orders in societal transitions as affective control of uncertainty*. Revised and resubmitted at *American Behavioral Scientist*.
- Kenny, J., & Morgan, J. H. (in press). Friend or foe: A Review and synthesis of computational models of the identity labeling problem. *The Journal of Mathematical Sociology*. <https://doi.org/10.1080/0022250X.2021.1923016>
- Zhao, J. (in press). Modeling Affective Basis of Morality and Justice among Chinese and Americans. *American Behavioral Scientist*.
- Zöllner, N., Morgan, J. H., & Schröder, T. (in press). Modeling interaction in collaborative groups: Affect control within social structure. *Journal of Artificial Societies and Social Simulation*. <https://www.jasss.org/24/4/6.html>
- Morgan, J. H., Zhao, J., Zöllner, N., Sedlacek, A., Chen, L., Piper, H., Beck, Y., Rogers, K. B., Hoey, J., & Schröder, T. (2021). Modeling the culture of online collaborative groups with affect control theory. In P. Ahrweiler & M. Neumann (Eds.), *Advances in social simulation* (pp. 147-169). Springer Proceedings in Complexity.
- Hofstede, G., Frantz, C., Hoey, J., Scholz, G., & Schröder, T. (2021). Artificial sociality manifesto. *Review of Artificial Societies and Social Simulation*. <https://rofasss.org/2021/04/08/artsocmanif/>
- Hoey, J., MacKinnon, N. J., & Schröder, T. (2021). Denotative and connotative control of uncertainty: A Bayesian dual-process model. *Judgment and Decision Making*, 16(2), 505-550. <http://journal.sjdm.org/20/200104/jdm200104.pdf>
- MacKinnon, N. J., & Hoey, J. (2021). Operationalizing the relation between affect and cognition with the somatic transform. *Emotion Review*, 13(3), 245-256. <https://doi.org/10.1177%2F17540739211014946>
- Hoey, J. (2020). *Structure is management of uncertainty in groups*. SocArXiv. <https://osf.io/preprints/socarxiv/rzywc>
- Joseph, K., & Morgan J. H. (2020). When do word embeddings accurately reflect surveys on our beliefs about people? *The 58th Annual Meeting of the Association for Computational Linguistics*. <https://arxiv.org/abs/2004.12043>
- Zöllner, N., Morgan, J. H., & Schröder, T. (2020). A topology of groups: What GitHub can tell us about online collaboration. *Technological Forecasting and Social Change*, 161, 120291. <https://doi.org/10.1016/j.techfore.2020.120291>
- Hoey, J., Sheikhabaee, Z., & MacKinnon, N. J. (2019). Deliberative and affective reasoning: A Bayesian dual-process model. *8th International Humaine Association Conference on Affective Computing and Intelligent Interaction*. <https://ieeexplore.ieee.org/document/8925215>
- Iyer, R. N., Yun, S. A., Nagappan, M., & Hoey, J. (2019). Effects of personality traits on pull request acceptance. *IEEE Transactions on Software Engineering*. <https://ieeexplore.ieee.org/abstract/document/8935389>
- Paryab, N., Sachs, A., Li, A., Nagappan, M., & Hoey, J. (2019). Relating values and social network structure. *5th International Conference on Computational Social Science*. <https://themis-cog.github.io/Paryab-IC2S2-2019.pdf>

- Hoey, J., Schröder, T., Morgan, J. H., Rogers, K. B., Rishi, D., & Nagappan, M. (2018). Artificial intelligence and social simulation: Studying group dynamics on a massive scale. *Small Group Research*, 49(6), 647-683. <https://doi.org/10.1177%2F1046496418802362>
- Rishi, D., Hoey, J., Nagappan, M., Rogers, K. B., & Schröder, T. (2018). Emotion and interaction processes in a collaborative online network. *4th International Conference on Computational Social Science*. <https://cs.uwaterloo.ca/~jhoey/papers/RishiIC2S2-2018.pdf>
- Hoey, J., Schröder, T., Morgan, J. H., Rogers, K. B., & Nagappan, M. (2018). Affective dynamics and control in group processes. *ICMI Workshop on Group Interaction Frontiers in Technology*. <https://cs.uwaterloo.ca/~jhoey/papers/Hoeyetal-GIFT2018.pdf>

6. Training related to THEMIS.COG

We were able to provide training to numerous students at all three universities at the undergraduate, Master's, PhD, and postdoctoral level, in some cases directly by employing them through the grant, in some cases more indirectly through thesis supervision provided by members of the THEMIS.COG research team.

University of Waterloo (Hoey and Nagappan)

- Deepak Rishi. *Affective Sentiment and Emotional Analysis of Pull Request Comments on GitHub*. Master of Mathematics thesis, University of Waterloo. 2017. <http://hdl.handle.net/10012/12728>
- Yuwei Jiao. *Affect Lexicon Induction For the GitHub Subculture Using Distributed Word Representations*. Master of Mathematics thesis, University of Waterloo. 2018. <http://hdl.handle.net/10012/14108>
- Areej Alhothali. *A Socio-mathematical and Structure-Based Approach to Model Sentiment Dynamics in Event-Based Text*. Doctoral dissertation, University of Waterloo. 2017. <http://hdl.handle.net/10012/12480>
- Rahul Iyer. *Effects of Personality Traits and Emotional Factors in Pull Request Acceptance*. Master of Mathematics thesis, University of Waterloo. 2017. <http://hdl.handle.net/10012/14952>
- Alexander Sachs. *Predicting Repository Upkeep with Textual Personality Analysis*. Master of Mathematics thesis, University of Waterloo. 2019. <http://hdl.handle.net/10012/14991>
- Nalin De Zoysa. *Analysis of Textual and Non-Textual Sources of Sentiment in GitHub*. Master of Mathematics thesis, University of Waterloo. 2020. <http://hdl.handle.net/10012/15952>
- Seonghu Yun. *Personality Traits of GitHub Maintainers and Their Effects on Project Success*. Master of Mathematics thesis, University of Waterloo. 2020. <http://hdl.handle.net/10012/16195>
- Gema Rodriguez-Perez was supervised by Dr. Nagappan and Zahra Sheikhabaee by Dr. Hoey during their terms as postdoctoral fellows at the University of Waterloo.

Dartmouth College (Rogers)

- Andrea Sedlacek. *How network structure and cultural sentiments shape affective alignment and behavior dynamics in online collaborative task groups*. Undergraduate honors thesis in Quantitative Social Science, Dartmouth College. 2020. <https://bit.ly/3i2HhTu>
- Six undergraduate students worked as research assistants on the grant for two or more terms: Andrea Sedlacek (2018-19), Annie Sherrill (2017), and Cayla Plotch (2017) were Presidential Scholars through Dartmouth's Office of Undergraduate Advising and Research (UGAR);

Lena Chen was a Sophomore Research Scholar through UGAR (2018-19); Hayley Piper (2019) and Yliana Beck (2019) were research interns through Dartmouth's Women in Science Project (WISP).

- Undergraduate research assistants Andrea Sedlacek, Lena Chen, Hayley Piper, and Yliana Beck co-authored a paper with the core research team, published in the proceedings of the *Annual Conference of the European Social Simulation Association* in 2019.
<https://par.nsf.gov/servlets/purl/10135900>
- Jun Zhao and Antonio Sirianni, who worked on the project, were supervised by Dr. Rogers during their terms as postdoctoral fellows in the Program in Quantitative Social Science at Dartmouth College (Zhao 2017-2019, Sirianni 2019-2022).

Potsdam University of Applied Sciences (Schröder)

- Nikolas Zöllner, PhD studies at Jacobs University Bremen (ongoing, completion expected in late 2021/early 2022). Supervisor: Jan Lorenz, Co-supervisor: Tobias Schröder. Working title: *The Structure and Dynamics of Groups in Open Source Software Development: A Computational Social Science Approach to Understanding Online Collaboration*. [Note: Potsdam University of Applied Sciences does not have the right to award PhD degrees; hence, this is a collaborative project with Dr. Lorenz]
- Constantin Dubyk. *Impact of collaboration support systems on group cohesion in virtual teams*. Unpublished Master of Arts thesis, Potsdam University of Applied Sciences. 2021.
- Alexander Yacine. *Forms of self-organized collaboration in digital space. An exploration of small group microcultures on GitHub*. Unpublished Master of Arts thesis, Potsdam University of Applied Science. 2018.
- Diego Dametto. *The relation between group dynamics and emotion in the context of dispersed working groups on GitHub*. Unpublished Master of Arts thesis, Free University of Berlin. 2019 (supervised by Dr. Schröder).
- Hong Mao Li (M.Sc. student in the Cognitive Science program at the University of Potsdam) worked as research assistant on the grant in 2018/2019.
- Jonathan Morgan, who worked as a postdoctoral fellow on the project at Potsdam University of Applied Sciences, was supervised by Dr. Schröder.

7. References

- Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. University of Chicago Press.
- Bales, R. F. (1999). *Social interaction systems: Theory and measurement*. Transaction Publishers.
- Corritore, M., Goldberg, A., & Srivastava, S. (2020). The new analytics of culture: What email, Slack, and Glassdoor reveal about your organization. *Harvard Business Review*, 98(1), 77-83.
- Cuddy, A. J., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior*, 31(1), 73-98.
- Dametto, D. (2019). The relation between group dynamics and emotion in the context of dispersed working groups on GitHub. Unpublished Master's thesis, Free University of Berlin.
- De Zoysa, N. (2020). Analysis of textual and non-textual sources of sentiment in GitHub. Master's thesis, University of Waterloo. <http://hdl.handle.net/10012/15952>
- Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., & Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.
- Gibson, D. R. (2010). Marking the turn: Obligation, engagement, and alienation in group discussions. *Social Psychology Quarterly*, 73(2), 132-151.
- Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from Twitter. *2011 IEEE 3rd International Conference on Privacy, Security, Risk, and Trust and IEEE 3rd International Conference on Social Computing*, 149-156.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), 26-42.
- Goldberg, A., & Stein, S. K. (2018). Beyond social contagion: Associative diffusion and the emergence of cultural variation. *American Sociological Review*, 83(5), 897-932.
- Gogoladze, T. (2020). The importance of personality traits in agile software development: A case study. Master's thesis, University of Tartu. https://comserv.cs.ut.ee/home/files/Gogoladze_ITJ_2020.pdf?reference=F596F53BF21FEE5D4EFB6C60CDA74487AAE8A3B7&study=ATILoputoo
- Gousios, G. (2013). The GHTorrent dataset and tool suite. *10th Working Conference on Mining Software Repositories*, 233-236.
- Heise, D. R. (2007). *Expressive order: Confirming sentiments in social actions*. Springer.
- Heise, D. R. (2010). *Surveying cultures: Discovering shared conceptions and sentiments*. Wiley.
- Heise, D. R. (2013). Modeling interactions in small groups. *Social Psychology Quarterly*, 76, 52-72.
- Hoey, J., MacKinnon, N. J., & Schröder, T. (2021). Denotative and connotative control of uncertainty: A Bayesian dual-process model. *Judgment and Decision Making*, 16(2), 505-550.
- Hoey, J., & Schröder, T. (under review). Disruption of social orders in societal transitions as affective control of uncertainty. Revised and resubmitted at *American Behavioural Scientist*.
- Hoey, J., Schröder, T., Alhothali, A. (2016). Affect control processes: Intelligent affective interaction using a partially observable Markov decision process. *Artificial Intelligence*, 230, 134-72.
- Hoey, J., Schröder, T., Morgan, J. H., Rogers, K. B., Rishi, D., & Nagappan, M. (2018). Artificial intelligence and social simulation: Studying group dynamics on a massive scale. *Small Group Research*, 49(6), 647-683.
- Iyer, R. N., Yun, S. A., Nagappan, M., & Hoey, J. (2019). Effects of personality traits on pull request acceptance. *IEEE Transactions on Software Engineering*, 1-12.
- Kozlowski, A. C., Taddy M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5), 905-949.
- Lawler, E. J., Thye, S. R., & Yoon, J. (2009). *Social commitments in a depersonalized world*. Russell Sage Foundation.

- Lawler, E. J., Thye, S. R., & Yoon, J. (2008). Social exchange and micro social order. *American Sociological Review*, 73(4), 519-542.
- Lawler, E. J. (2001). An affect theory of social exchange. *American Journal of Sociology*, 107(2), 321-352.
- Lee, K. F. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt.
- Luo, J., & Magee, C. L. (2011). Detecting evolving patterns of self-organizing networks by flow hierarchy measurement. *Complexity*, 16(6), 53-61.
- Mehrabian, A. (1996). Relations among political attitudes, personality, and psychopathology assessed with new measures of libertarianism and conservatism. *Basic and Applied Social Psychology*, 18(4), 469-491.
- Morgan, J. H., Rogers, K. B., & Hu, M. (2016). Distinguishing normative processes from noise: A comparison of four approaches to modeling impressions of social events. *Social Psychology Quarterly*, 79(4), 311-332.
- Morgan, J. H., Zhao, J., Zöller, N., Sedlacek, A., Chen, L., Piper, H., Beck, Y., Rogers, K. B., Hoey, J., & Schröder, T. (2021). Modeling the culture of online collaborative groups with affect control theory. In P. Ahrweiler & M. Neumann (Eds.), *Advances in social simulation* (pp. 147-169). Springer Proceedings in Complexity.
- Munaiah, N., Kroh, S., Cabrey, C. et al. (2017). Curating GitHub for engineered software projects. *Empirical Software Engineering*, 22, 3219-3253.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66(6), 574-583.
- Osgood, C. E., May, W. H., Miron, M. S. (1975). *Cross-cultural universals of affective meaning*. University of Illinois Press.
- Raymond, E. S. (1999). *The Cathedral and the Bazaar*. Sebastopol, CA: O'Reilly & Associates.
- Ridgeway, C. L. (2019). *Status: Why is it everywhere? Why does it matter?* Russell Sage.
- Rishi, D. (2017). Affective sentiment and emotional analysis of pull request comments on GitHub. Master of Mathematics thesis, University of Waterloo. <http://hdl.handle.net/10012/12728>
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. McGraw-Hill Book Company.
- Schröder, T. (2011). A model of language-based impression formation and attribution among Germans. *Journal of Language and Social Psychology*, 30(1), 82-102.
- Schröder, T., Hoey, J., & Rogers, K. B. (2016). Modeling dynamic identities and uncertainty in social interactions: Bayesian affect control theory. *American Sociological Review*, 81(4), 828-855.
- Schröder, T., & Scholl, W. (2009). Affective dynamics of leadership: An experimental test of affect control theory. *Social Psychology Quarterly*, 72(2), 180-197.
- Sedlacek, A. (2020). *How network structure and cultural sentiments shape affective alignment and behavior dynamics in online collaborative task groups*. Undergraduate honors thesis in Quantitative Social Science, Dartmouth College. <https://bit.ly/3i2HhTu>
- Thye, S. R., Yoon, J., & Lawler, E. J. (2002). The theory of relational cohesion: Review of a research program research program. *Advances in Group Processes*, 19, 139-166.
- Tilbury, K., & Hoey, J. (2020). Multi-agent reinforcement learning and human social factors in climate change mitigation. *AAAI Spring Symposium on Challenges and Opportunities for Multi-Agent Reinforcement Learning (COMARL)*, Stanford, CA.
- Tsay, J., Dabbish, L., & Herbsleb, J. (2014). Influence of social and technical factors for evaluating contribution in GitHub. *Proc. 36th International Conference on Software Engineering*, 356-366.

- Zayas, V., & Shoda, Y. (2009). Three decades after the personality paradox: Understanding situations. *Journal of Research in Personality*, 43(2), 280-281.
- Zhao, J. (in press). Modeling Affective Basis of Morality and Justice among Chinese and Americans. *American Behavioral Scientist*.
- Zöller, N., Morgan, J. H., & Schröder, T. (2020). A topology of groups: What GitHub can tell us about online collaboration. *Technological Forecasting and Social Change*, 161, 120291.
- Zöller, N., Morgan, J. H., & Schröder, T. (in press). Modeling interaction in collaborative groups: Affect control within social structure. *Journal of Artificial Societies and Social Simulation*.