

Research Statement

Denae Ford

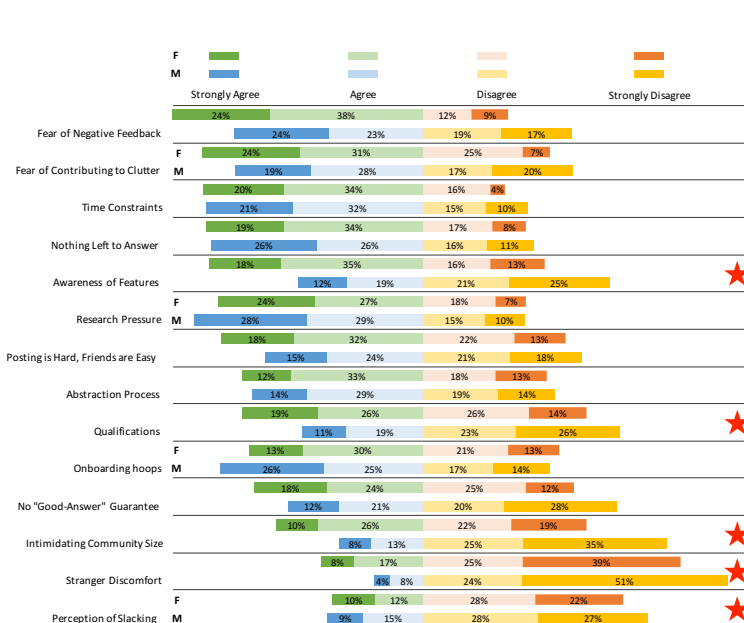
Research Overview. Online programming communities, such as Stack Overflow and GitHub, have become a fundamental resource in helping about 50 million people each month learn to code and overcome errors they encounter along the way. These spaces offer opportunities to ask questions, find answers, and review example code snippets from an international community of programmers. With their varied expertise, those who ask and answer can be novices to programming, experts learning a new language, and those in between interested in contributing to the greater body of programming knowledge. However, the global interest in learning to code has shined a light on the increased demand of diverse professional programmers and their decreasing interest to use the well-known help resources. In short, many programmers do not feel comfortable using them. At the intersection of *human-computer interaction (HCI)* and *software engineering*, my research investigates why programmers underutilize these help resources.

My research has evolved to answer one question: **How can we create a more inclusive help experience for programmers online?** To answer this question, I apply psychology principles to reshape online programming community mechanisms and expand these spaces to a broader cohort of users. Specifically, my work considers both *quantitative* and *qualitative* approaches to create an inclusive programming community on Stack Overflow and GitHub by doing the following: 1) identifying barriers underrepresented group of programmers face when attempting to participate, 2) determining approaches users take to maximize their usage of a community in its current state, and 3) design new embedded interventions, integrated into the platform, where programmers can help them engage.

My research contributions and future research agenda align well with the Research in Software Engineering group’s goal to push the state-of-the-art in software productions process. With the recent acquisition of GitHub and my expanding research in that space, I can continue to have global impact on how software engineers engage and seek resources in online programming communities.

Research Contributions

Identifying barriers to participation: Women in online programming communities



Each online programming community has a primary function that attracts users to their platform. On Stack Overflow for example, users often come to the Q&A site to find answers to programming questions. Although many users visit for the helpful programming answers, low participation rates have been observed, especially in underrepresented group such as women, as well as complaints of an unwelcoming community for novices. Identifying the barriers that exist across online programming communities allows practitioners to be more aware of who their platforms are discouraging not by intention—but by design.

I interviewed women and conducted a survey that included response from men, women, and programmers across the gender spectrum about their positive and negative experiences when posting to Stack Overflow. Specifically, we asked them about what motivated and discouraged

Figure 1: After identifying barriers by interviewing women, we conducted a survey of programmers of men and women to identify how much barriers hindered their participation on Stack Overflow. Stars indicate barriers that significantly hindered women.

them from posting questions, answers, voting, and participating on the site in any other capacity. **I found 14 barriers that resonated with programmers across the gender spectrum.** Of those, five significantly impede women more than men including: not feeling qualified enough to answer any questions (*Qualifications*) and having a fear of posting in a large, intimidating community (*Intimidating Community Size*) [1]. In summary, this work helped identify barriers that exist for women, but have the capacity to hinder all users. Using this conceptual framework of barriers, each can be used to design interventions to increase participation in online programming communities.

Determining existing approaches to engagement: Identity-based participation

After we identified barriers, we investigated existing protocols used by users to overcome the mechanisms that can discourage others. Several studies have identified reasons for low participation rates in underrepresented groups in online programming communities. However, few have identified how identity can play a role. In the previous study on Stack Overflow participation among women, participants mentioned that one reason they do not post on Stack Overflow is that, *“They are just not even on the same race track”* as the majority of programmers who post there; most presenting as men [1]. Expanding on behavioral science theories of social facilitation, I define this notion of observing technical peers on the same “race track” or having similar individuals to compare oneself to as *peer parity*. Peer parity is when an individual can identify with at least one other peer when interacting in a community. Research external to programming communities suggest the presence of peers can increase activity from underrepresented users in unfamiliar spaces.

In this work, I sought to find out if the same held true in online programming communities with the following research question: *How does peer parity influence how women post on Stack Overflow?* Building on studies of identity and peer interactions, I hypothesized that differences in exposure of peer parity may influence participation in online programming communities. To investigate this question we sampled identifiable women on Stack Overflow based on display names, reviewed the excerpt of user data displayed when a questions is posted and answered, and analyzed differences in activity for Stack Overflow parity (>1 woman) and non-parity (1 woman) threads. **We found a significant difference in the time between posts for women who asked a question on parity threads in comparison to non-parity threads** ($p = 1.83e-05$) [2]. These results demonstrate that women who ask questions on peer parity threads re-engage sooner in additional Stack Overflow activities. Thus, supporting the theory that identity-based signals, in this case gender, can increase participation in online programming communities.

Designing embedded interventions: Just-in-time mentoring

From the projects mentioned above, programmers of different experience levels and genders face barriers —*reputation-gated permissions* and *being overwhelmed by the large community*—that inhibit them from asking questions. On Stack Overflow, many questions from novices are ill-received: downvoted, left unanswered, or deleted. Actualizing design claims of online communities, I proposed that dismantling barriers with varying approaches, such as guiding novices through onboarding hoops or reducing the feeling of an intimidating community size with a mentor, can help users feel more comfortable participating in online programming communities.

For this study, we used theories of mutual engagement and formative feedback in order to learn how to create a better onboarding experience and identify improvements for the design

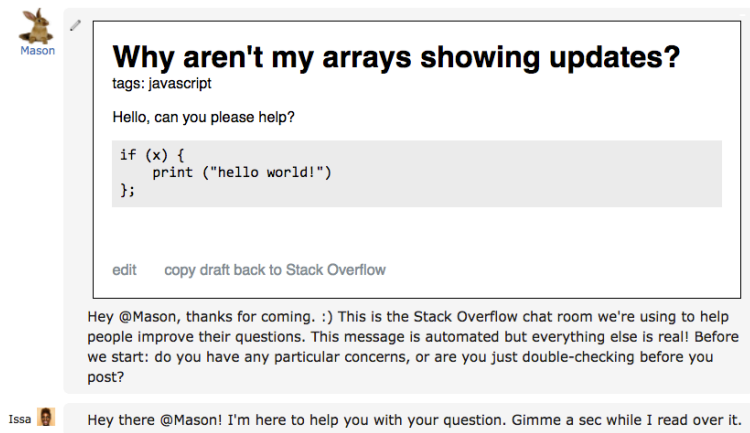


Figure 2: I designed a just-in-time mentoring program embedded into the Stack Overflow experience. In this experience, experienced Stack Overflow users and novices to the community can engage in collaborative editing to shape an upcoming Stack Overflow question.

of a more scalable system of asking questions. My colleagues and I recruited experienced Stack Overflow users as mentors and allowed novices, less experienced users to the platform, to opt-in to receive feedback on their question precisely when needed. In the one-month online study I conducted, we implemented our mechanism for mentored question-asking on Stack Overflow, and enabled 271 novices to receive help with their questions. After this just-in-time mentoring experience, we compared the question asking process of novice received feedback to those that did not. As a result, we found that mentored questions were substantially improved over non-mentored questions: **Average scores increased by 50%, resulting in fewer off-topic, deleted, and poor questions** [3]. Our empirical evaluation demonstrates that with just-in-time mentoring, we can reduce negative experiences for participants, improve community receptiveness to novice contributions, and increase a sense of belonging among new contributors.

Future Research Agenda

My future work continues to study the scope of technical workers, but evolves to consider trust of contributions, identity presentation across help spaces, and empowering marginalized people when communicating technical knowledge. The findings from my previous research projects offer reasonable transitions into this work as I 1) continue the ‘ask for one, solve for all’ approach allowing us to integrate the findings for a specific sub-community into the community-at-large and 2) successfully transfer implications of this work to industry. With these ideals at the core of my approach, I will continue to conduct research that enhances online communities to increase their utility to a diverse set of users. The following research directions guide this agenda.

How do we use identity cues to trust code from strangers?

Sharing code online has evolved to include more social characteristics of a potential contributor including larger profile images, more emphasis on their name, and highlights of affiliations to organizations. However, limited research demonstrates how these signals—beyond the code itself—effect how the code submitted is evaluated. How do we know which signal is being used to make a decision? Is it the country listed, perceived gender identity, or similarities between the identities of the submitter and reviewer that aid the decision making process? In order to study this phenomenon, I will analyze how programmers review code contributions on GitHub using eye-tracking. Using an eye-tracking methodology allows us to capture the nonverbal cues that programmers extract but cannot always vocalize in retrospective interviews. As a first step, I would like to develop a theory of what elements of code contributions are most valuable. Given my prior work [4], are these signals the best approach to communicate the value of code contributions? How should the interface programmers use to review changes better reflect these contributions?

How can we more naturally transition between public and private help spaces?

As more online programming communities are created, we see more variety in levels of privacy and control over how content is broadcast to specialized technical audiences. From previous studies, I have seen that programmers are more likely to ask questions when they are in private, protected settings, even if that they do not know if the people they are asking can help them achieve a resolution. But how does the process of asking change when programmers ask questions to the specialized audience? Should it matter how familiar they are with their work offline? How does this effect their professional world of their discipline? How do technical professionals show their knowledge but remain open enough to receive help? How does anonymity become more helpful in these spaces? We may find a similar resistance of sharing publicly to sharing privately. I am interested in investigating how the complexities of these spaces exist to connect colleagues in conjunction with strategies used to balance reputation with help-seeking behavior.

How do we better empower marginalized programmers in virtual technical spaces?

Previous work demonstrated that being seen and identifiable to similar marginalized individuals in communities can play a role in empowerment, but what about the role it plays in enriching the experiences of being a marginalized person in a technical community? My colleagues and I found that connecting with individuals through a shared identity, whether online or offline, can encourage that experience of “being the only one” [5]. Open research questions to explore include: If underrepresented users are not using or engaging in popular online programming communities of the majority, like Stack Overflow, then where are they seeking technical help? How can we make our communities more like those spaces? Expanding even further, how do we then appropriately assess these technical worker’s knowledge as candidates for careers in software?

Broader Impact

My work was been published at various ACM SIGSOFT, SIGCHI and IEEE conferences including Foundations of Software Engineering Conference (FSE), Conference on Human Factors in Computing Systems (CHI), Conference on Computer Supported Cooperative Work (CSCW), and Visual Languages and Human Centered Computing (VL/HCC). By writing brief blogs to share the research with those outside of academia, I have connected with industry programmers and discussed how the findings of the work impact their daily work flow. My work on broadening participation with Stack Overflow has been acknowledged by Jeff Atwood, founder of Stack Overflow, and has fostered research collaborations with Stack Overflow Developer and Growth Team. This work has resulted in new features to help users ask questions on the site. As a faculty member, I plan to continue these collaborations. I have also had considerable success in funding these projects through fellowships from Microsoft Research, the National Science Foundation, and the National GEM Consortium.

References

- [1] **Ford, Denae**, J. Smith, P. Guo, and C. Parnin, “Paradise unplugged: Identifying barriers for female participation on stack overflow,” *Proceedings of the 24th International Symposium on the Foundations of Software Engineering*, 2016.
- [2] **Ford, Denae**, A. Harkins, and C. Parnin, “Someone like me: How does peer parity influence participation of women on stack overflow?” in *IEEE Symposium on Visual Languages and Human Centric Computing*. IEEE, 2017.
- [3] **Ford, Denae**, K. Lustig, J. Banks, and C. Parnin, “We don’t do that here: How collaborative editing with mentors improves engagement in social q&a communities,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018.
- [4] **Ford, Denae**, M. Behroozi, A. Serebrenik, and C. Parnin, “Beyond the code itself: How programmers *really* look at pull requests,” 2019, manuscript submitted for publication.
- [5] F. Liu, **Ford, Denae**, C. Parnin, and L. Dabbish, “Selfies as social movements: Influences on participation and perceived impact on stereotypes,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW ’18. ACM, 2018.