

Gender, Feedback, and Learners' Decisions to Share Their Creative Computing Projects

EMILIA F. GAN, University of Washington, USA

BENJAMIN MAKO HILL, University of Washington, USA

SAYAMINDU DASGUPTA, University of North Carolina at Chapel Hill, USA and University of Washington, USA

Although informal online learning communities are made possible by users' decisions to share their creations, participation by females and other marginalized groups remains stubbornly low in technical communities. Using descriptive statistics and a unique dataset of shared and unshared projects from over 1.1 million users of Scratch—a collaborative programming community for young people—we show that while girls share less initially, this trend flips among experienced users. Using Bayesian regression analyses, we show that this relationship can largely be attributed to differences in the way boys and girls participate. We also find that while prior positive feedback is correlated with increased sharing among inexperienced users, this effect also reverses with experience or with the addition of controls. Our findings provide a description of the dynamics behind online learners' decisions to share, open new research questions, and point to several lessons for system designers.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**; *Collaborative and social computing design and evaluation methods*; • **Social and professional topics** → **Informal education**; **Gender**; *Computer science education*;

Keywords: broadening participation; gender differences; social learning; creative learning; online communities; social computing and social navigation; computer mediated communication; Scratch

ACM Reference Format:

Emilia F. Gan, Benjamin Mako Hill, and Sayamindu Dasgupta. 2018. Gender, Feedback, and Learners' Decisions to Share Their Creative Computing Projects. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, CSCW, Article 54 (November 2018). ACM, New York, NY. 23 pages. <https://doi.org/10.1145/3274323>

1 INTRODUCTION

Research in social computing and the social and learning sciences has shown that by sharing creative artifacts, youth participation in informal online communities can lead to meaningful learning experiences [33, 53, 60]. Although the production of unshared or unpublished artifacts may include important learning experiences for their creators, private creations will never elicit feedback, comments, and remixes from peers and can never introduce their creators to the type of rich social milieu that both theory and empirical research suggest can promote continued participation and learning. Research has shown that there are systematic reasons why some users

Authors' addresses: Emilia F. Gan, University of Washington, Paul G. Allen School of Computer Science & Engineering, Seattle, WA, 98195, USA, efgan@uw.edu; Benjamin Mako Hill, University of Washington, Department of Communication, Seattle, WA, 98195, USA, makohill@uw.edu; Sayamindu Dasgupta, University of North Carolina at Chapel Hill, School of Information and Library Science, Chapel Hill, NC, 27599, USA, University of Washington, eScience Institute, Department of Communication, Seattle, WA, 98195, USA, sayamindu@unc.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2018 Copyright held by the owner/author(s).

2573-0142/2018/11-ART54

<https://doi.org/10.1145/3274323>

will share more than others and reap greater benefits from doing so. Despite the promise of lower barriers to participation in informal learning communities, females and groups that are marginalized in the broader field of computing are relatively underrepresented participants in informal online computing communities as well [19, 56]. The missed opportunities represented by unpublished artifacts created by members from marginalized groups—potentially kept private at higher rates—reflect both potential sources of inequality as well as avenues toward broadening participation. How might designers encourage more users—especially members of marginalized groups like females—to participate by sharing their creative products?

Using a unique dataset of 5,677,206 multimedia programs created by youth in the Scratch online community, including those that are created but unshared, we characterize the dynamics that drive users' decisions to share. Drawing from a rich body of theory and prior empirical work in social computing and the learning sciences, we test hypotheses related to users' gender and history of receiving positive feedback. Using a novel stage-based stratified Bayesian regression strategy, we report results from a series of empirical models that test these hypotheses for users across a range of experience levels.

In non-parametric analysis of the data, we see that although projects created by girls are initially less likely to be shared than those created by boys, this relationship is flipped among more experienced users. We also find some evidence that a history of receiving positive feedback is correlated with sharing among inexperienced users. In both cases, these relationships are reversed in models with controls for users' social behavior and project characteristics. In these models, we find that the marginal effect of gender on sharing is near zero among users' early projects and that, with everything else equal, girls share more than boys as they gain experience. These models also suggest that the marginal effect of having received positive feedback is negatively associated with sharing among users of all experience levels and this relationship is particularly strong among boys.

Our work contributes to the literature on participation in informal learning environments by testing theories about users' decisions to share their creations. Additionally, it contributes to a growing body of research in social computing focused on broadening participation by mapping the longitudinal processes through which gender inequality in participation shifts over time. Finally, it makes a methodological contribution to social computing research through our application of a stage-stratified Bayesian regression framework that allows us to model how predictors of engagement shift over users' life cycles.

2 BACKGROUND

In this section, we engage with prior research to define the terms of, and provide rationales for, a set of three hypotheses about when users will choose to share their creative artifacts. In each case, our hypotheses are framed in terms of dynamic processes where relationships may shift over time [45]. We adopt this approach because we believe that theory suggests that strong predictors of sharing among new users might be poor predictors among those with more experience. Drawing from Monge's [45] advice on describing dynamic processes, each of our hypotheses is split into two parts: first, (a) we hypothesize about the effect for inexperienced users; second, (b) we hypothesize about how this relationship will (or will not) change among users with more experience.

Often focused on public communication and framed in terms of "self-censorship," a body of research in social computing has sought to measure and understand users' decisions to publish social media content or keep it hidden [13, 40]. We build on this work to unpack the dynamics of individuals' decisions to share in informal computing learning environments. Although definitions of "informal learning" have been contested, the term is typically used to describe learning in the absence of a fixed curriculum and evaluation that takes advantage of self-directed exploration,

incidental experiences, and socialization [62]. Although informal learning occurs over individuals' lives in many contexts, learning scientists have suggested that social computing systems are particularly appropriate environments for its support [12].

In studies at the intersection of informal learning and computing among children, Papert's constructionism framework [48], built on Piaget's theory of constructivist learning [51], is the most widely used theoretical tool. Although Papert never formally defined constructionism, he described his philosophy as based on the idea that learning "happens especially felicitously in a context where the learner is consciously engaged in constructing a public entity" [49]. In constructionism, a well-designed learning environment should provide a social context where learners are able to discuss their ideas with others, present their works as public entities, and observe their peers' works and reactions [48, 49]. Although early constructionist learning environments deployed software like Logo [1] in settings like schools and after-school programs [48], researchers and designers working in informal settings have taken up Papert's approach most enthusiastically. Many influential projects to support youth learning in computing are situated in informal constructionist environments—both physical environments like *Computer Clubhouses* [55, 75] and online environments such as MOOSE Crossing [5, 6] and Scratch [54]. Sharing is seen as one of several important pathways for supporting the learning process in constructionism and is a central feature of these environments.

Sharing is seen as important to learning in these settings for two key reasons. First, it creates opportunities to engage socially with others in a community in ways that support a virtuous cycle of social participation. In turn, participation is understood to promote, or even constitute, learning. Jenkins et al. discussed important new literacies involving social skills developed through collaboration and networking, skills that can only be learned through full engagement in a participatory culture [30]. In a number of quantitative studies, it has been argued that increased participation in informal learning communities will, on average, support increased learning [15, 61, 74]. Sharing is thought to support learning by opening the sharer to opportunities for receiving constructive critique. This perspective finds support in a large body of qualitative studies of both offline and online informal learning communities [4, 6, 55, 60, 69]. In a case study from the Scratch community, Brennan and Resnick [3] described how sharing a game made with Scratch allowed a young programmer to not only get valuable feedback, but also provided an opportunity for him to "think about what it was like for someone else to experience playing his game." In his earlier work, Resnick's [52] spiral process of learning includes the sharing of ideas and creations in a key position. Second, sharing can support learning indirectly among others by creating valuable material from which others can learn. A growing body of research has documented how learners in informal learning environments can learn by building on and extending each others' work [14] with improvements in the quality of artifacts constructed [16, 25].

One of the most important critiques of informal constructionist environments is that its interest-driven and opt-in approach does not do enough to address inequality in participation along dimensions that include race, class, and gender [23, 50, 56]. The gender gap in technology interest is the most widely studied of the three and is present as early as elementary school [9, 41]. Even when girls are performing at equivalent or higher levels than boys, their confidence and interest in technical or math-intensive subjects lag [29]. In that women learn that computing is "not for them" at a very young age, new female learners may approach informal online learning communities that are centered around technical topics (e.g., programming) with suspicion or trepidation. Sharing work publicly involves permitting oneself and one's creations to be vulnerable to the judgment and censure of others. This brings us to our first hypothesis (H1): that (a) *projects created by girls will be less likely to be shared than those created by boys*. We further hypothesize that (b) *differences in sharing behavior associated with gender will remain similar in magnitude between less and more experienced users*.

As described in detail in the paragraphs above, a critical component of the learning process afforded by public participation is the potential for constructive feedback. Learner motivation is intimately tied to learner self-identity [18, 68] and learners' visions of what they might become [24, 39]. Feedback can potentially help learners overcome the discrepancy between where they see their current self and their ideal future self [18]. Research has shown that learners respond better to feedback on small tasks that are performed well than to feedback on identical tasks that were not completed as competently [11] and has also shown that positive feedback leads to increased goal commitment and motivation [72].

Successful informal online sharing communities are often explicitly supportive of participation [7, 20, 59]. Many afford giving motivational (i.e., lacking in specific useful critiques) feedback which users find inspiring and contributing to an overall positive and supportive environment [20]. In survey-based research on users' decisions to share photos that they had already taken, Nov et al. [47] found that both intrinsic and extrinsic motivations drive sharing, and they pointed to the users' desire for self-development and reputation attainment within their community. Motivational feedback mitigates the perceived risk to reputation that publicly sharing created work poses to learners [28, 36], helping participants feel safe when sharing their creative efforts [35, 36]. This leads us to our second hypothesis (H2): (a) *projects created by learners who have received more positive feedback are more likely to be shared than those created by learners who have received less*. As suggested, feedback is important in that it can build confidence in users which, in turn, can lead them to share. In that we believe that confidence is also going to be a function of experience, we hypothesize that (b) *the effect of positive feedback on sharing will remain positive, but will be lower in magnitude among more experienced users than among less experienced users*.

Stereotype threat is the term used to describe the experience of being in a situation where one may be judged in a negative light solely as a consequence of belonging to a particular stigmatized group [63]. There exists a considerable body of research showing the detrimental effects of being the target of these negative stereotypes. Thoman et al. [66] conducted a meta-analysis describing how stereotype threat not only has a direct negative effect on goals and motivation, but also decreases a marginalized learner's sense of belonging and identification within a given domain. Other work has shown that interventions that restore or create a sense of belonging can counteract the effects of even persistent stereotype threat [70, 73]. Rusk showed how participation in a supportive informal online learning environment can help all learners feel accepted and valued [60].

As a result, although many learners will become more comfortable as they become more experienced in engaging with a community, this effect is likely to be more important to learners from marginalized groups. This need to be accepted can be considered from the viewpoint of social attraction theory. Guillen et al. described how being competent is not sufficient for women at a software development company to appear confident to coworkers and supervisors; they must also be liked by those judging their confidence [21]. Wang et al. explored similar issues in their study of contributions by women in open software projects on GitHub [71]. The decision of whether or not to participate is thus not a simple one for women, as their acceptance into the communities that they seek to join is tied to not only the quality of their contributions, but also their social attraction. This brings us to our final hypothesis (H3): *the relationship between positive feedback and sharing will be stronger for projects created by girls than for those created by boys*. Because we know of no theoretical reason to believe that this effect will be mitigated with experience, we hypothesize that (b) *this effect will be consistent across less and more experienced users*.

3 EMPIRICAL SETTING: SCRATCH

Our empirical setting is the Scratch online community. Scratch is a visual, block-based programming language that is designed for children aged 8–16 [54]. With roots in constructionism, Scratch

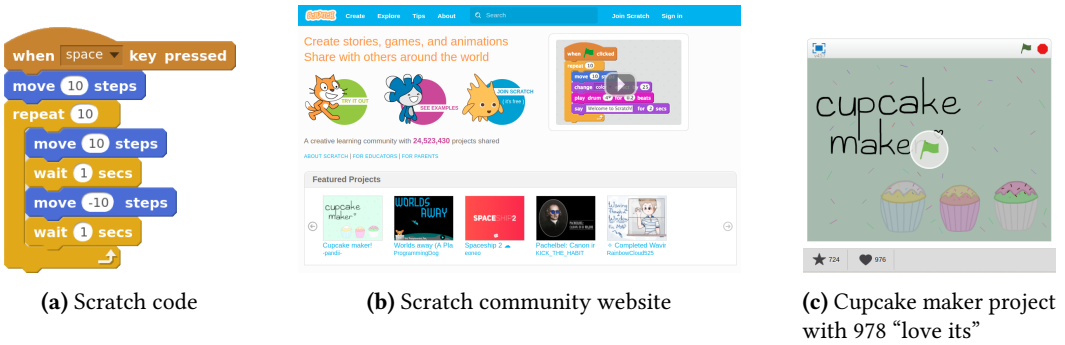


Fig. 1. Scratch programming language and community. Learners use the block-based Scratch programming language (a) to create games and animations. Shared projects are visible to all members of the Scratch community, and may be featured on Scratch’s homepage (b). An example of a project that received positive feedback from 978 other users who clicked “love it” (indicated by the heart-shaped icon) is shown in (c).

was designed to appeal to people who had never viewed themselves as programmers and to be “more tinkerable, more meaningful, and more social” than other programming environments [54]. Computer programs in Scratch are authored by dragging and dropping visual blocks of code. Each of these blocks represents instructions that determine the behavior of on-screen characters called sprites. An example of Scratch code to make a sprite go back-and-forth on the screen is shown in Figure 1a.

The Scratch online community website¹ allows children using Scratch to create and share their projects online and is shown in Figure 1b. Members of the community can remix each others’ Scratch projects, comment on each others’ work and user profiles, curate projects in galleries, bookmark projects that they find interesting, and show appreciation for projects that they like by clicking on the “love it” button shown in Figure 1c. Scratch and its online community were launched publicly in 2007. The Scratch online community is enormously popular, with more than 30 million registered users (called Scratchers). It hosts more than 33 million shared projects. All projects hosted on Scratch are openly remixable and about a quarter of these projects are remixes themselves. The Scratch community is deeply social and hosts more than 167 million comments.² Scratch is not only one of the most popular sites in the world for children to learn programming, but also one of the largest platforms for youth cooperative creative production of any kind.

There is evidence that Scratch supports learning among users who participate in a sustained way. Maloney et al. observed urban youth at a Computer Clubhouse over an 18-month period and found that self-directed novice Scratch users learned to program when given support and encouragement in a setting of mutual learning [38]. Unfortunately, Scratch has struggled to elicit sustained participation among the large majority of its users. Although quantitative studies have shown that the skills of Scratch users tend to grow with participation [14, 42, 74], attrition results in few users becoming experienced enough to register meaningful increases in quantitative measures of learning [42, 61].

A growing body of research has sought to understand the dynamics and designs that cause users of informal communities like Scratch to participate. Fields et al. have pointed to sharing as the

¹<https://scratch.mit.edu/> (<https://perma.cc/U4AP-3FWF>)

²All statistics are taken from the Scratch statistics page [65].

critical behavior to focus on because “sharing a project on the Scratch site defines the baseline of all other active participation beyond viewing” [19]. In qualitative work, Kafai et al. have described participants in Scratch after-school programs who expressed reluctance to share their projects owing to the sense that they lacked the expertise required to create projects that were of comparable quality to those created by others [31]. However, due to the fact that non-participation has generally been invisible to those studying online learning communities, we still know very little about this process.

One thing we do know is that participation in Scratch continues to be unequal in terms of gender. Using a random sample of 5,000 users of the Scratch website during January 2012, Fields et al. found out that more than half had no observable engagement and that just over two-thirds of the users who had created at least one project during the observation period were males [19]. Furthermore, when they analyzed the projects created by these users, they found that users could be separated into four classes, determined by the variety of programming concepts used in their projects. Girls were found to be over-represented in the most basic class and under-represented in the two more highly skilled classes [19].

Increased positive feedback has been cited as one useful way to increase engagement. For example, Scratch created a “welcoming committee” to promote positive newcomer experiences [17]. While running *Collab Camp*, an organized online collaborative event, Roque et al. found that encouraging positive and constructive feedback among participants was effective and appreciated [58]. Owing to the fact that girls’ projects tend to receive more positive ratings than similarly popular projects by boys [25], this may reflect a promising avenue toward addressing gender inequality as well.

In addition to actively promoting a supportive and welcoming environment, the Scratch Community Guidelines actively discourage the posting of negative feedback and instruct users to report any project or comment that they consider to be “mean, insulting, too violent, or otherwise inappropriate” [44]. This is a responsibility that Scratch users take very seriously and carry out with remarkable efficacy [37, 67]. In an informal qualitative analysis of 100 randomly selected comments, we identified only two comments as having potential negative valence—one of which had been flagged and removed by the community.³

Historically, quantitative studies of Scratch have struggled to speak directly to the issue of participation because users’ decisions to not participate were never visible. In part, this is because the initial version of the Scratch programming environment was a desktop application that uploaded projects to the website. Due to the fact that unshared projects were never observed by anybody but their creator, much of what we know about nonparticipation is from smaller qualitative studies of workshops and after-school programs. In May 2013, Scratch launched a new major revision that allowed for in-browser authoring and editing [43]. An implication of this change was that projects created on the website by community members are private by default but accessible to researchers.

4 DATA AND MEASURES

Observations in our dataset correspond to individual Scratch projects. In Scratch, users typically create projects from an empty template as “*de novo*” projects. They can also create projects by opening, modifying, and then republishing an existing project as a “remix.” Because the dynamics around users’ decisions to share remixes may differ from decisions to share individual work, our analysis is restricted to *de novo* projects. We also excluded projects uploaded by the Scratch offline client, because projects uploaded by the client are shared automatically on the website without further user interaction.

³This analysis and the text of the negative comments are described in detail in our supplement.

Variable	Range	\bar{x}	\tilde{x}	σ
<i>Create Count</i>	[1, 3247]	20.9	5	77.7
<i>Autosave Count</i>	[2, 55542]	26.0	14	53.6
<i>Love Count</i>	[0, 11198]	6.3	0	89.2
<i>Share Count</i>	[0, 1461]	8.0	0	33.6
<i>Follow Count</i>	[0, 6304]	9.3	0	62.1
<i>Block Count</i>	[0, 68952]	42.7	12	211.9
<i>Unique Media Asset Count</i>	[1, 3012]	5.8	5	8.0
<i>Age (in years)</i>	[4, 90]	15.6	12.6	9.5

\bar{x} = Mean, \tilde{x} = Median, σ = Standard deviation

Table 1. Summary statistics of variables used in our regression models.⁵

Access to our dataset came in the form of permission to query a copy of the SQL database that runs the Scratch online community and was allowed by having our team members added as researchers on an IRB protocol that covers the post hoc analysis of this database. Our dataset consists of projects by every user who registered on Scratch's website in the period between July 1st 2014 and January 31st 2015. Data on 18 users were corrupt, and we excluded all projects from this group. Additionally, we excluded users who self-reported their age to be above 90 or below four. We collected *de-novo* projects that were created using the browser-based editor through January 31st 2016 and were not subsequently deleted by their creators. We did not include projects that appeared to have been created automatically and unintentionally by users who opened the editor but had not set out to create projects.⁴

Next, we collected project-level variables that included the sequence number of the project in its creator's trajectory (*Create Count*) which we use as a categorical variable upon which we stratified our dataset. We constructed our dependent variable as a dummy variable (*Is Shared?*), coded 1 if the project was ever published and 0 otherwise. Published projects that were subsequently unpublished by their authors were still coded as 1.

Our two key predictors are qualities of the user who created and shared each project. The first is a self-reported measure of gender that we coded using a dummy variable that is 1 if the user self-reported being female and 0 if the user reported being male (*Is Female?*). Of course, not all Scratch users identify themselves as male or female and the Scratch website has allowed users to report their gender using a free-form text field since 2013. In our dataset, approximately 8% of Scratch users self-report their gender as something other than "male" or "female." Although many of these appear to be legitimate non-binary genders, a large portion appear to be playful or humorous (e.g., "robot"). With deep respect to the importance of representing the reality of non-binary gender, we restricted our sample to the large majority of users along the male/female gender binary as a way of drawing statistical generalizations about average effects among the only two genders in our sample that are preponderate and theoretically well-understood enough to support doing so. This involved dropping data on projects by 105,438 users.

⁴Formally, we excluded all projects with less than two "autosaves." Due to how the Scratch website was implemented, these projects would typically be created unintentionally when a user who is logged in opened the editor. More details on autosaves are provided below.

⁵We were puzzled by the high upper-limit of *Autosave Count*. Manual inspection revealed that some of the high values were recorded on rare occasions where the Scratch editor failed to parse a response back from the server, causing projects to be repeatedly auto-saved.

Our second key predictor is a time-varying measure of the total number of “love its” each user’s *de novo* projects received prior to the creation of the project in question (*Love Count*). “Love its” in Scratch are very similar to “Likes” on Facebook. The total number of “love its” is displayed on each project without attribution to the person giving them (see Figure 1c). We use *Love Count* as our measure of prior positive feedback.

We also include two sets of control variables. The first set attempts to control for factors that theory suggests will be unrelated to gender. Our first control in this set is a project-level variable that attempts to capture the time that users spent on projects as the total number of “autosaves” (*Autosave Count*). Under normal circumstances, the Scratch code editor automatically saves a project every 2 minutes if there has been any change to the project. Because projects can be worked on indefinitely without publication but can be shared at any point, *Autosave Count* captures a count of autosaves in the 30-day period after each project was created.⁶ Our second control reflects previous sharing experience as measured by the number of *de novo* projects shared by the creator prior to the project in question (*Share Count*). We include this control because receiving “love its” (or any type of feedback on a project) is conditional on sharing projects in the first place. Our final control in this set is a user-level variable capturing users’ self-reported age in years at the point of project creation (*Age*).

Our second set of controls attempts to capture qualitative ways that boys and girls may differ in how they use Scratch that might be correlated with projects’ shared status. Our first control in this set is a measure of social activity in the form of a count of the number of users on Scratch that have been “followed” by the creator of the project at the moment the project was created (*Follow Count*). “Following” on Scratch is akin to “following” on social media websites like Twitter and causes the Scratch website to surface the followed users’ creations. This second set of controls also seeks to capture project complexity in the form of a measure of the number of programming blocks used in the project (*Block Count*) as well as the number of distinct image and sound elements that the project contains (collectively referred to as “assets” in Scratch) (*Unique Media Asset Count*). Both of these variables are measured at the point of the final autosave that we observe. Previous work has shown that these measures are conceptually distinct and are associated with both different types or genres of projects and different modes of social interaction [25]. We expect that all our controls will be consistently positively related to a project’s shared status.

Our final dataset includes 5,677,206 projects shared by 1,182,793 users. Summary statistics for the nondichotomous variables from our final dataset are described above in Table 1. For our two dichotomous variables, 39% of *Is Female?* and 29% of *Is Shared?* are equal to 1. Detailed plots included in our supplement show the distributions and medians for all our variables broken down by stratum and gender.

5 ANALYTIC PLAN

Our analytic strategy is based on *stage-based stratified Bayesian regression models*, a term we use to describe testing an identical set of hypotheses across a series of strata. In particular, we stratify our sample into stages by *Create Count* in order to test our hypotheses at different points along users’ experience levels. Doing so effectively splits our dataset into 25 strata that contain all users’ 1st, 2nd, 3rd, etc. projects, through 26 projects. Our models are estimated on a dataset that incorporates data beginning from the 2nd project shared by each user because users must have created at least one project in order for there to be any variation in users’ number of previously shared projects or received “love its”. We choose an upper limit of 26 because this corresponds to the 98th percentile

⁶Although some of these autosaves to published projects may have occurred after publication, we believe that this measure allows for a fair comparison of the effort spent on both shared and unshared artifacts.

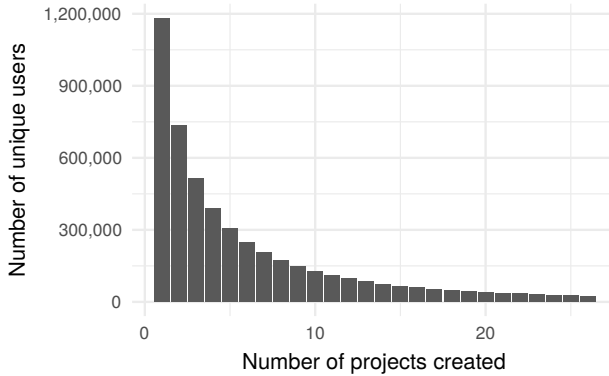


Fig. 2. Histogram of *Create Count*. Each bar reflects the number of users that have shared that many projects (i.e., the number of projects, in each stratum).

of the maximum value of *Create Count* for users in our data collection window. The fact that each statistical model contains one project per user alleviates concerns of repeated observations of individual users common in longitudinal analysis.

Our approach allows us to estimate how certain predictors are associated with the probability of an individual project being shared and how those associations vary across levels of creators' experience. Critically, our approach does not tell us how an individual's sharing behavior changes with experience. In prior social computing research, a similar "stage model" approach has been used for prediction tasks by Cheng et al. [10] to study information cascades (e.g., mass resharing of photos) and has shown that the relative importance of the predictors of cascades can vary across stages.

Our nondichotomous measures are each counts. Like many variables drawn from social media data, our counts follow long-tailed distributions similar to Figure 2. For example, given high rates of attrition in Scratch [14, 42, 61], there is a rapid drop-off in the number of projects in each stratum as *Create Count* increases (see Figure 2). As a result, we log-transform each of our count measures before adding them to our models, adding 1 when our measure's range includes 0.

Thus, we test our hypotheses using 25 separate logistic regression models fitted using the 25 stratified datasets. As a logistic regression, our outcome is the log-odds that a project will be shared. Our predictors are a sum of a linear combination of our parameter estimates and measures for gender, our log-transformed measure of "love its", and an interaction between the two variables, plus our controls. The interaction term attempts to capture hypothesis H3 that the log-linear effect of "love its" will differ across gender. This formal model is represented by the following equation (the left side reflects the logistic link function):

$$\log \left(\frac{\hat{p}(\text{Is Shared?})}{1 - \hat{p}(\text{Is Shared?})} \right) = \beta_0 + \beta_1 \text{Is Female?} + \beta_2 \log \text{Love Count} + \beta_3 \text{Is Female?} \times \log \text{Love Count}$$

Controls are added to this model in two stages. In our first model (M1), we add our first set of controls for our log-transformed measures of *Share Count*, *Autosave Count*, and *Age*. In our second model (M2), we add our controls for log-transformed *Follow Count*, *Block Count*, and *Unique Media Asset Count*. M1 attempts to reflect the marginal effects of gender and feedback on sharing, controlling for important confounding variables that theory and previous research does

not suggest will differ systematically by gender. M2 incorporates measures that previous research has suggested will correspond to systematic differences in the ways that boys and girls use Scratch [19, 56]. For example, girls may be more social and create projects with more images and sound [19]. We see M1 as reflecting the nature of gender differences in regard to feedback and sharing more descriptively, while M2 begins to unpack potential explanations for these differences.

In both models, our test of the relationship between the gender of the creator and the likelihood of sharing (H1) is reflected by the parameter estimate for β_1 which we hypothesize will be negative and similar in magnitude throughout our strata. Our test of the relationship between prior positive feedback and sharing (H2) is reflected by the estimates for β_2 . We hypothesize that the parameter estimates will be positive for lower strata and then decrease in magnitude but remain positive among more experienced users. Finally, our test that the effect of feedback will be stronger among girls (H3) is reflected in the parameter estimate for β_3 , which we hypothesize will be consistently positive.

Although much of the approach outlined above could be carried out with the frequentist statistical models that are traditionally used in HCI research, there is no appropriate statistical test to compare the results of one frequentist model fit using one dataset to the results from the same model fit on a different one. Thus, after fitting our models, we would be unable to compare the estimates and the confidence intervals for β_1 , β_2 , etc. from models associated with different strata. A common solution to this challenge is to build a single model that includes information on strata in a categorical variable that is interacted with every other model parameter. Due to the fact that our analysis contains so many strata, such a model would be difficult to estimate and interpret.

As an alternative, we adopt a Bayesian regression framework that allows this type of comparison across models [46]. Conceptually, adopting a Bayesian approach means that our model fitting process estimates distributions called “posterior distributions” instead of point estimates for each parameter. Rather than reporting estimates and standard errors, Bayesian analyses typically report the mean and 95% “credible” intervals from a series of repeated draws from the estimated posterior distributions. These intervals describe the range of values around the estimated means, within which 95% of our draws from the posterior fall. They can be compared with the intervals that result from fitting the same model on a different dataset.

In recent work, Bayesian analysis has been suggested as a promising tool for HCI research [34]. That being said, due to the fact that this approach will likely be unfamiliar to at least some, we report results from 25 traditional frequentist logistic regression models in our supplement. With very few exceptions, the means of the posteriors from our Bayesian models are identical to the point estimates in the corresponding frequentist models. In other words, while our Bayesian approach puts us on a firmer epistemological footing when comparing estimates across models, our modeling choice is not driving our estimates or conclusions.

We estimate our models using the software package *Stan* [8] and the R front-end package *rstanarm* [64]. All of our models use weakly informative priors appropriate for logistic regression as recommended by *Stan*’s developers,⁷ and are each fit using four Markov chains with 2,000 iterations per chain. The first 1,000 samples from each chain are discarded. Model fit diagnostics suggest that all the chains converged for each of our models. The potential scale reduction factor statistics (\hat{R}) for the models are included in our supplement. Although the data cannot be shared because of privacy concerns and IRB rules, complete source code for all of our analyses is provided in our supplement.

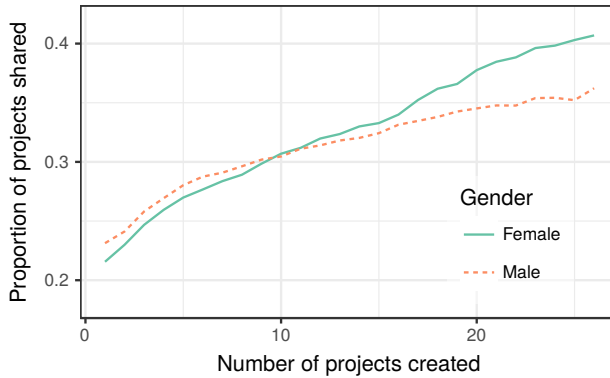


Fig. 3. Proportion of projects shared by gender across experience levels, measured as the number of projects created, for 1.1 million Scratch users. Projects created by girls are less likely to be shared than those by boys until about the 9th project is created. The relationship is subsequently reversed.

6 RESULTS

Figure 3 presents a non-parametric and descriptive depiction of the proportion of projects shared by girls and boys across different experience-based strata of users. The graph shows that when users are less experienced (i.e., number of projects created < 10), the proportion of projects shared is slightly higher among boys than it is among girls: for example, approximately 21.6% of girls share their first projects whereas 23.1% of boys do so. After 10 projects, this relationship flips and projects created by girls become more likely to be shared. The vast majority of Scratch projects are created by inexperienced users, meaning that a combined dataset of projects from the strata that we analyze would suggest that boys share projects more often than girls. As the average effect hides this dynamic, Figure 3 demonstrates the value of exploring the data longitudinally.

Figure 4 visualizes the posterior distributions of the parameters associated with our key question predictors from both of our models. The x -axis of every plot represents the *Create Count* (i.e., the stratum), and the y -axis of each plot represents the values of a specific parameter (i.e., β) from the fitted model. The distribution of each parameter is visualized as a “violin” plot [27], where the dot within the violin plot indicates the mean of the parameter distribution and the lines attached to the dot show the 95% credible interval range.

The top left panel of Figure 4 depicts the binary predictor *Is Female?* for model M1. This graph shows that the estimated posterior distributions for the lower strata are in line with our stated hypothesis for H1a. The distribution mean remains near or below 0 until the *Create Count* is 7. This suggests that, accounting for the controls in our model, projects made by girls in Scratch are less likely to be shared compared to those created by their male counterparts until users have created around seven projects. As foreshadowed in Figure 3, new projects by girls are more likely to be shared than those by boys in strata containing users who have created eight or more projects. Again, consistent with Figure 3, model M1 provides some support for H1a that girls will share less than boys initially and contravening evidence for hypothesis H1b that this trend will also apply to more experienced users.

As can be seen in the top right panel, model M2 shows a similar relationship between sharing and *Create Count*, in that more experienced girls are more likely to share. This indicates that the

⁷<https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations> (<https://perma.cc/3NKV-GWAU>)

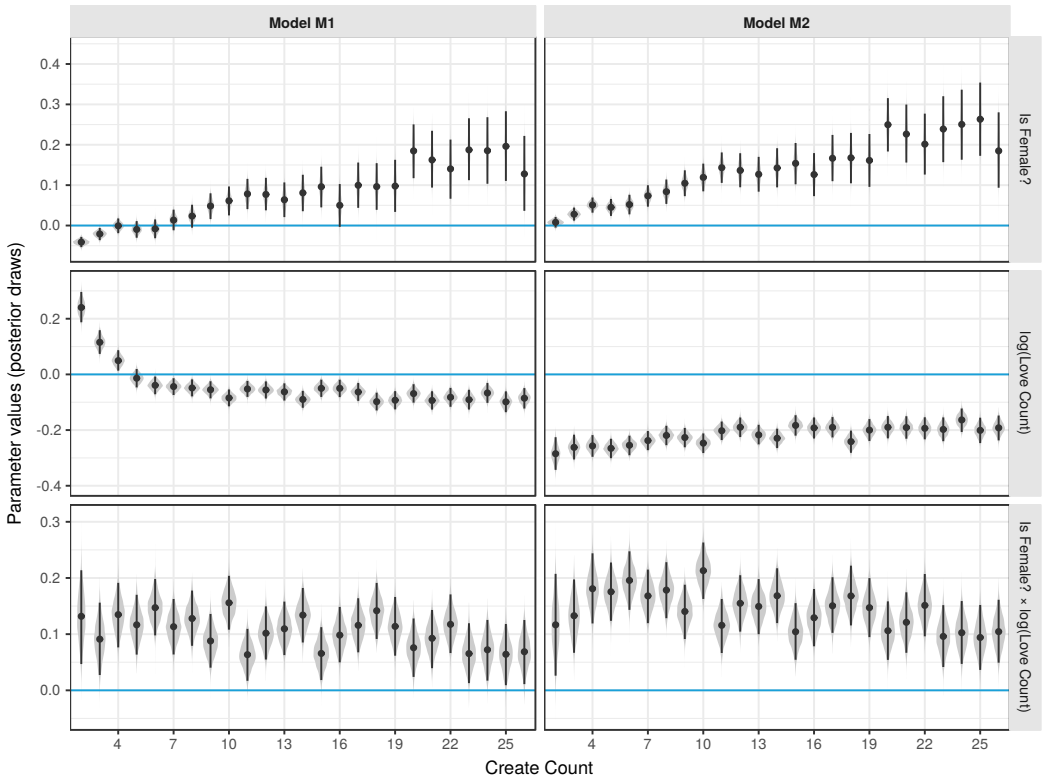


Fig. 4. Graphs showing the posterior distributions of the parameters across values of *Create Count*. Each of the three panels in a given column shows the estimates for a single parameter (i.e., β) associated with an independent variable or an interaction in our models. Columns reflect our two models. Each estimate is represented as a “violin” plot denoting the distribution of the parameter and is a representation of uncertainty. Each “violin” contains a dot representing the mean of the distribution, and a line representing the 95% credible interval. The line representing a null effect ($\beta = 0$) is indicated by the horizontal blue line in each panel.

introduction of controls for social activity and project characteristics in M2 does not alter our finding that more experienced girls are more likely to share. That being said, each of our estimates from M1 is shifted upward in M2 and we conclude that girls are at least as likely to share projects as boys at every stratum. For the stratum containing all users’ 2nd projects, M2 shows that the odds of a project being shared by a female Scratch user are not distinguishable from those of a male user. Among the stratum of users’ 12th projects, the odds are 1.15 times as high. We believe that these estimates are distinct because the 95% credible intervals for the two parameter distributions ($[-0.01, 0.02]$ and $[0.09, 0.18]$) do not overlap. M2 suggests that initial differences in the sharing rate between boys and girls can be explained by differences in the qualities of their projects and social interaction. The upward trend in both top panels of Figure 4 suggests heartening news for proponents of broadening participation.

H2 hypothesizes that positive feedback will be positively correlated with the likelihood that a subsequent project will be shared and that the effect will be consistent. The parameter estimates for *Love Count* in M1, shown in the right middle panel of Figure 4, suggest that our hypothesis is
















<i>Create #</i>	<i>Is Female?</i> (β_1)	$\log(\text{Love Count})$ (β_2)	<i>Is Female?</i> \times $\log(\text{Love Count})$ (β_3)
2	 0.01 [-0.01, 0.02]	 -0.28 [-0.34, -0.23]	 0.12 [0.03, 0.21]
6	 0.05 [0.03, 0.08]	 -0.25 [-0.29, -0.22]	 0.20 [0.14, 0.25]
12	 0.14 [0.09, 0.18]	 -0.19 [-0.22, -0.15]	 0.15 [0.10, 0.20]
18	 0.17 [0.10, 0.23]	 -0.24 [-0.28, -0.20]	 0.17 [0.12, 0.22]
24	 0.25 [0.16, 0.34]	 -0.16 [-0.21, -0.12]	 0.10 [0.05, 0.16]

Table 2. Posterior distributions of the parameters (β -s) of interest for models (M2) with *Create Count* set to 2, 6, 12, 18, and 24. Each row shows results from one fitted model (M2). Each cell contains a histogram that shows the distribution of the parameter, followed by the mean value of the parameter distribution and the corresponding 95% credible interval. The range and position of the x-axes of the histograms are held constant over a given column, so that the relative shift in the distributions can be easily discerned (e.g. the right-ward shift of the parameter distribution of *Is Female?*).

supported for lower values of *Create Count*. Surprisingly, the opposite seems to be the case from the 5th created project onward. In other words, M1 estimates that users who have created four projects or fewer are more likely to share their project if they had received more “love its” in the past. However, users who have created more than four projects are less likely to share their project when they have received more positive feedback.

The introduction of controls for project qualities and social interaction (shown in M2 in the middle right panel of Figure 4) provides a picture that is consistently at odds with our predictions in H2. The 95% credible intervals for this parameter distribution ([-0.34, -0.23]) overlap with distributions estimated at other strata (e.g., [-0.29, -0.22] for *Create Count* = 6), suggesting that we cannot confidently state that these estimates are distinct. Although the introduction of a stronger set of controls results in stable parameter estimates in support of H2b, our estimates are consistently negative and opposite in sign to what we had predicted. M2 suggests that for the 2nd project created by a user, each log-unit increase in the number of “love its” is associated with odds of being shared that are only 0.75 times as high. This result is counter-intuitive and surprising. In the discussion section of this paper, we discuss a possible explanation at length.

In terms of H3, we find that the parameter distributions for our interaction term (shown in the bottom panel in Figure 4) are positive across our strata. We also find that the introduction of stronger controls in M2 does little to change these estimates. This indicates that the association between prior “love its” and the likelihood of sharing is stronger for projects created by girls. This reflects support for both H3a and H3b.

The posterior distributions from one of our fitted models (M2) are shown in a different format in Table 2. Due to space constraints, the table only includes the posterior parameter distributions that are pertinent to our hypotheses across a subset of our strata (five different values of *Create Counts*). Our supplement includes full results for all 25 regressions for both M1 and M2.

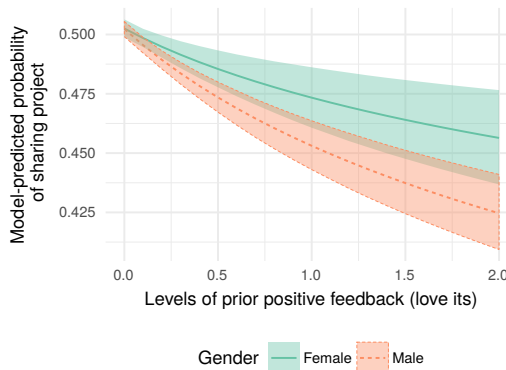


Fig. 5. Predicted probabilities from model M2 (with 95% prediction intervals) of sharing a project for prototypical female and male learners who have each created two projects and shared one project. The x-axis shows the range of 0 to 2 “love its”.

The estimates for our controls are not reported in this paper but are included in full detail in our supplement. With the exception of *Age*, which has a varied but small effect, the effects of our controls are positive, as predicted. We find very high parameter estimates for *Share Count*, suggesting that a creator’s past sharing behavior is a very strong predictor of whether a given project is going to be shared. The distribution of the parameter associated with *Autosave Count* remains positive throughout our models, suggesting that more time spent on developing projects is associated with an increased chance of projects being shared. The parameter distributions of *Age* remain close to 0 throughout the observed range of *Create Count*. Our estimates for *Follow Count* are large but decrease in magnitude in higher strata and our measures for both *Block Count* and *Unique Media Asset Count* are consistently large and positive and relatively stable in magnitude across strata. Overall, we find that more socially active users on Scratch share their projects more, and that projects that include more code and more media are shared more frequently.

We understand that the presence of interactions, especially across dozens of models, can make interpreting these results challenging. In order to aid in interpretation, we provide two visualizations of model M2 in Figures 5 and 6. These visualizations show simulated data produced by our model for certain prototypical boys and girls. These prototypical users have values of *Is Female?*, *Create Count* and *Share Count* that are fixed in ways that are described in the figures’ labels, axes and captions. Our control variables are all held at their sample medians—computed separately for boys and girls—for users in each stratum. In each case, shaded areas describe prediction intervals that contain 95% of the predicted values from our models. Lines reflect the mean value of our predictions. Overlapping shaded areas describe how often our models will make similar predictions for a type of prototypical boy and girl, but they should not be interpreted as a lack of statistical significance in terms of the marginal effects of our predictors (these are shown in Figure 4).

In Figure 5, we show the model-predicted probability of sharing for the 2nd project created by prototypical male and female Scratch users over a range of 0 to 2 “love its”. Since it is impossible to receive a “love it” without having shared a project, we hold the value of *Share Count* to 1. We also hold *Autosave Count* and *Age* at their median values. We see from the figure that, with no “love its,” the predicted probability of a project by a girl being shared is extremely similar to one created by a boy. However, the receipt of even a single “love it” leads to predictions that will nearly always be higher for girls.

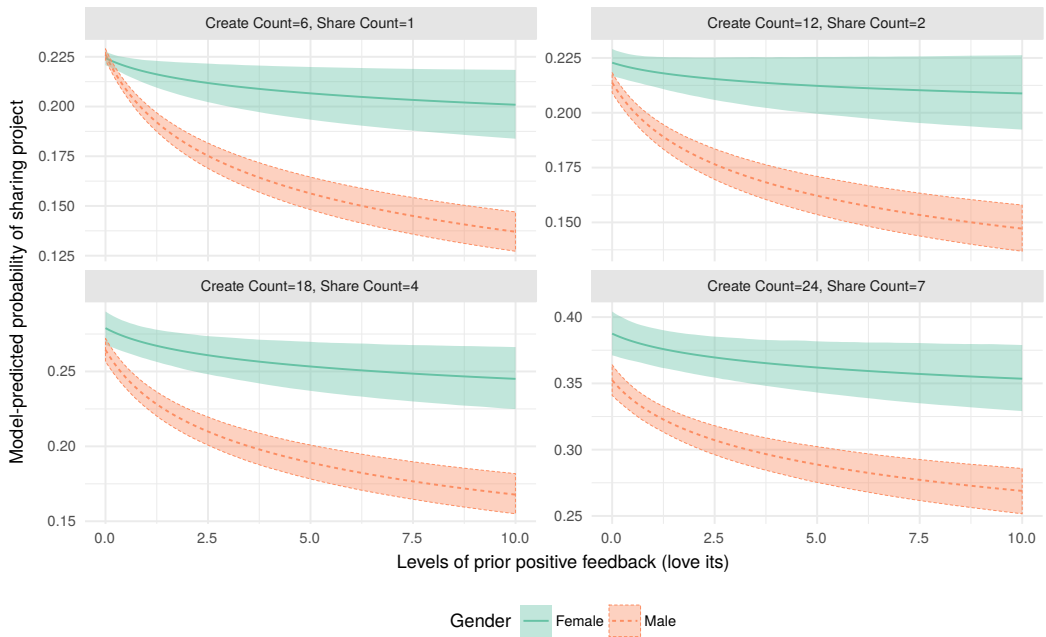


Fig. 6. Predicted probabilities from model M2 (with 95% prediction intervals) of sharing a project for female and male learners who have created 6, 12, 18, and 24 projects. The x-axis shows the range of 0 to 10 “love its”. For each set of plots, the share count is set to the median of that stratum.

In Figure 6, we use *Create Counts* of 6, 12, 18, and 24 to reproduce this model-predicted probability plot across a range of strata. These plots suggest that the relationship between “love its” and sharing is consistently negative and that the relationship is stronger (i.e., steeper) among boys than among girls. The plots also show how the relationship between gender and sharing shifts in higher strata in two ways. First, the initial gap on the left side of the plots shows that as users share more projects, projects made by girls with no previous “love its” are increasingly likely to be shared, relative to projects made by otherwise similar boys. Second, attention to the scale of the *y*-axes across the panels shows that the difference in the effect of “love its” on the sharing behavior between boys and girls widens with experience.

The results from our analysis provide mixed support for our hypotheses. We find only qualified support for our hypotheses about inexperienced users (H1a, H2a, and H3a). Our results in M2 suggest that taking project and social characteristics into account can explain most of the differences in the sharing rate across gender in users’ earliest projects (H1a) and that the marginal effect of feedback on sharing is initially negative (H2a). Our hypotheses about how the relationships will shift among users with more experience find even less support. As hypothesized, we find support for a consistent effect of our interaction between gender and feedback (H3b). However, we failed to anticipate the nature of other shifts with H1b and H2b both diverging from our predictions. For H1b, we find that boys are *less* likely to share than girls although, in the absence of controls for project characteristics and social interaction, this effect is conditional on having created more than seven projects. For H2b, we find that boys who have received more “love its” in the past are *less* likely to share than those who have received less positive feedback, although, once again, this is conditional on having shared at least four projects in the absence of stronger controls.

7 THREATS TO VALIDITY

The validity of our findings may be affected by a number of threats and limitations. The most important threat is that our data become thin as *Create Count* increases (see Figure 2). For example, this means that although M1 suggests that the early-stage inequality in sharing rates of projects across gender flips at seven projects, this reflects the experience of a relatively small proportion of users. Of course, this threat is not unique to our study or to Scratch in that levels of attrition in Scratch seem to be typical of most online learning platforms [22].

Another concern is that although it is tempting to interpret our findings as suggesting that decreasing attrition might also decrease gender inequality, our results cannot demonstrate this directly, neither do our results imply that individual users' sharing patterns change. Although within-user changes in behavior may contribute to our results, so may processes of attrition or selection. At every stratum of the process that we model, there are reasons to believe that users who share one more project are different from the users who share one fewer.

There are several limitations in our work related to the validity of measures. For example, projects may be shared in an order that is different from the order in which they were created. This might make comparisons between strata noisier or more biased. Projects may be shared at any point in time. Although most sharing happens very soon after projects are created, and although we track all projects for at least seven months after their creation, the fact that unshared projects could be published at any time means that projects created at the later end of our data collection window may be slightly less likely to be shared. This might also add error or bias.

It is also important to emphasize that *Love Count* is a limited way to measure feedback. Previous work has looked into the enormous variation in the quality, content, and tone of comments left on projects in Scratch [26, 69]. Although we do not attempt to measure feedback in this way, it is possible that the content of comments received could provide a distinct source of feedback. We might also expect to see the effects of other types of feedback like "favoriting" (similar to bookmarking in Scratch) or adding projects to studios (community curated collections of projects).

Our measure of gender is another possible threat to validity because we cannot know whether users are reporting this information truthfully. Importantly, our work ignores the experience of Scratch users with non-binary genders. Our measure of age is also self-reported and suffers from similar threats. An additional threat related to age is that older users, including adults like teachers, might persist longer. The wide range of ages in our data creates a related threat to our broader framing in terms of youth.

In order to address these threats related to age, we can point to the distributions of users by age as reported in aggregate on the Scratch website [65]. As an additional check, we have broken this data down by stratum for our dataset and published this material in our supplement. We can draw comfort from the fact that the distribution of self-reported ages among new users in Scratch is centered at the age of 12 with the large majority falling between 8 and 16. The analysis in our supplement shows that these distributions are remarkably stable across strata as well. Although it seems unlikely that the very small number of adults in our data is driving our results or threatening our ability to draw conclusions about youth, we included a control for the linear effect of age in all of our models. Its inclusion does not substantively affect our results. We ran our analysis with a dataset where the upper age-limit was restricted to 18.⁸ The substantive results remained the same.

A final threat is the external validity of our findings and the degree to which our work will generalize to other informal learning communities. Beyond the general challenge of generalizing from a study of a single community, this threat stems from the fact that although Scratch was designed as an informal environment, it is often used in settings such as schools where learners

⁸We have included the parameter plots from this analysis in our supplement.

may be expected or required to share their creations. Although we expect the proportion of projects created in such settings to be limited, our ability to generalize to informal environments may be impeded to the extent that our results reflect work in formal settings.

8 DISCUSSION AND CONCLUSION

Even though Scratch was designed to appeal to diverse tastes and interests and to be a supportive learning environment where sharing and collaboration are encouraged, we find that its users share only 29% of the projects that they create. As predicted by theories suggesting that girls have less confidence in their ability than boys, our bivariate descriptive statistics suggest that girls' early projects are less likely to be shared than similar projects by boys, but that this relationship is flipped among more experienced users. Our regression models suggest that the difference in sharing between inexperienced boys and girls can largely be explained by differences in users' social behaviors and project qualities, and that experienced girls share at a substantially higher rate than boys do.

Although we find some evidence that new users who received more positive feedback on their initial projects were more likely to share than similar users who received less, our models suggest that prior positive feedback is a negative predictor of sharing among experienced users. We find that controlling for social connection and project characteristics means that this relationship is consistently negative. We also find that the relationship between positive feedback and sharing is much more positive (i.e., less negative) in girls than in similar boys, and that this relationship was largely consistent across user experience levels.

As shown in Figures 5 and 6, the sizes of our estimated effects are mostly small—i.e., the visualized differences in the sharing rates are several percentage points at most. We draw some comfort from the fact that these small differences might still represent tens or even hundreds of thousands of projects and a substantial proportion of inequality in participation associated with gender. Although we believe that these estimated effects are meaningful, our effects' modest size reminds us that our variables are only one source contributing to inequality in participation in informal learning communities.

Our findings contribute to a body of work that is situated at the intersection of social computing and informal computer-science education research. Specifically, we contribute to the understanding of factors that are associated with sharing of artifacts in online communities, in creative computing in particular. As sharing is the central act of participation in many online communities, our findings point to factors that are predictors of participation and engagement.

Our stage-based stratified modeling approach suggests that these predictors are dynamic and that they shift and change in relative importance depending on the stratum of users whose projects are being observed. Weak predictors of sharing among early-stage users can be strong predictors among more experienced groups. As discussed in our threats section, one possible explanation may be an endogenous process of attrition. For example, girls who do not share or do not receive positive feedback early on may simply stop creating new projects at a higher rate. Through that process, girls will fall out of the dataset used in our higher strata.

For researchers studying social computing communities, our result illustrates how differences in gender and positive feedback are associated with differences in participation. Although we find some support for prior findings in social computing research which indicate that motivational feedback can support member participation, these results must be heavily qualified. Positive feedback is associated with increased odds of participation only for projects that are made by early-stage users, and this is only true in our models that do not include controls for project characteristics and social activity. All of our predicted probabilities from model M2 for prototypical users, male and females

across a range of experience levels, suggest that the marginal effect of positive feedback is lower levels of sharing.

One explanation is a phenomenon described in the music industry as “sophomore album syndrome” or “second album syndrome” (SAS). SAS occurs when a musician has an acclaimed first album and then struggles to complete a second. SAS is generally attributed to the paralyzing effect of pressure caused by a larger audience and increased expectations. We find some theoretical support for the idea that SAS might explain our results in H2 in research that suggests that imagined audiences online can lead to self-censorship [13, 40]. Research on Facebook has suggested that boys with predominantly male peer groups will self-censor more frequently [13]. Research by Ridgeway [57] described both how SAS might happen and why it might affect boys more: starting in childhood, gender frames social relations through which people compare their performance to others’ and make judgments on where their own talents lie; contemporary stereotypes describe men as more competent, whereas women are considered nicer. Boys, therefore, may both hold higher expectations for receiving community validation of their competence and feel more threatened when this community’s approval is not received. In other words, boys may be more susceptible to SAS because the stakes in terms of feedback are higher.

Brennan’s [2] ethnographic study on audience in Scratch suggested that the audience-awareness and approval-seeking behavior necessary for SAS is present in Scratch. Brennan’s subjects considered audience frequently, were frustrated at their inability to reach it, and sometimes described the attention of others as unwanted. Brennan suggested that the designers of Scratch could help individuals find smaller communities within the larger population of users. It is possible that experienced girls have been more successful at forming smaller communities in which to assess their achievements on Scratch and, thus, feel more rewarded by receiving relatively fewer “love its”. This might mitigate the effect of SAS among girls.

Of course, it is difficult for us to gather direct evidence on users’ beliefs and reasoning given the observational nature of our study and our dataset based on digital traces. One weak piece of evidence that we can present is a consistently positive correlation that we estimated between our measure of feedback (*Love Count*) and our measure of effort (*Autosave Count*) across all strata of our data (details are included in our supplement). This suggests that users with more positive feedback spend more time on their projects, something that we would expect if SAS were occurring. Unpacking this effect further will likely require complementary interpretive methods like interviews or vignette studies to understand users’ thought processes, beliefs, and feelings.

For designers of online informal learning communities and educators interested in informal STEM education, our research shows how stereotype threat may play out. Although our dependent variable is a measure of participation, not learning, prior work has argued that participation is deeply related to learning in communities like Scratch [32]. Our findings suggest that gender is, as believed, associated with the decision to participate but that this association is complex and dynamic. Our work points to a more precise locus of unequal participation as well as a potential means to combat it. In particular, our results suggest that girls might benefit from more structured support early on in their engagement with such communities. It is possible that social support gained by girls who engage with the Scratch community effectively counters some of the negative messages that young girls receive that tend to turn them away from being interested in technology.

Our work serves as a cautionary finding for designers of interventions for fostering greater engagement. What we find suggests that positive feedback mechanisms such as “love its” are not uniformly associated with increased odds of participation. Indeed, for many groups of users, the association, at least along one dimension of participation (i.e., sharing), may be negative. That being said, we do not know whether positive feedback is associated with other forms of increased

participation—e.g. it may be possible that an increased number of “love its” on Scratch is associated with a decreased likelihood of attrition.

Finally, at a methodological level, our results demonstrate how stage-based stratified regression models can allow social computing researchers to uncover dynamic statistical relationships that change over time, and that would be missed with an overall average. A single frequentist version of M2 fit on a dataset combining all our analyzed strata identified that average effect sizes do not capture the dynamic processes that we uncover in this paper. For example, the parameter estimate for *Is Female?* ($\beta = -0.1, p < 0.001$), implies that the odds of a project created by a girl being shared is 0.89 times the odds of a similar project made by a boy. This average hides the way that this relationship shifts with experience.

We present our study with the hope that future work will look more deeply into the factors that influence sharing behaviors. We also hope that future studies will extend our work's encouraging suggestion that female participation becomes more robust once an initial barrier to engagement is overcome as well as examining our surprising result that motivational feedback such as “love its” is associated with less future sharing of creative artifacts. Although our study is restricted to one community, we hope that future researchers will apply and extend our questions and methods in the context of other online communities, leading to a more general understanding of gender disparity in technology and in learner engagement in informal online communities more broadly.

ACKNOWLEDGMENTS

This work was facilitated by the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. The authors would like to acknowledge Aaron Shaw, members of the Scratch team at MIT, and members of the Community Data Science Collective for their feedback and support. Co-author Dasgupta was supported by a Moore-Sloan-WRF Innovation in Data Science postdoctoral fellowship at the University of Washington while working on this study. Portions of the project were completed at the Helen R. Whitely Retreat Center at the University of Washington's Friday Harbor Laboratories. Financial support for this work came from the National Science Foundation (grants IIS-1617129, IIS-1617468, DRL-1417663, and DRL-1417952), the University of Washington, and the University of North Carolina at Chapel Hill.

REFERENCES

- [1] Hal Abelson, Nat Goodman, and Lee Rudolph. 1974. *LOGO Manual*. Memo 313. Massachusetts Institute of Technology, Cambridge, MA. <http://dspace.mit.edu/handle/1721.1/6226>
- [2] Karen Brennan. 2014. Audience in the service of learning: how kids negotiate attention in an online community of interactive media designers. *Learning, Media and Technology* 41, 2 (2014), 193–212. <https://doi.org/10.1080/17439884.2014.939194>
- [3] Karen Brennan and Mitchel Resnick. 2013. Imagining, Creating, Playing, Sharing, Reflecting: How Online Community Supports Young People as Designers of Interactive Media. In *Emerging Technologies for the Classroom: A Learning Sciences Perspective*, Chrystalla Mouza and Nancy Lavigne (Eds.). Springer New York, New York, NY, 253–268. https://doi.org/10.1007/978-1-4614-4696-5_17
- [4] Karen Brennan, Amanda Valverde, Joe Prempeh, Ricarose Roque, Michelle Chung, Karen Brennan, Amanda Valverde, Joe Prempeh, Ricarose Roque, and Michelle Chung. 2011. More than code: The significance of social interactions in young people's development as interactive media creators. In *Proceedings of ED-MEDIA 2011—World Conference on Educational Multimedia, Hypermedia & Telecommunications*, Vol. 2011. Association for the Advancement of Computing in Education (AACE), Lisbon, Portugal, 2147–2156. <http://www.editlib.org/p/38158/>
- [5] Amy Bruckman. 1997. *MOOSE crossing: Construction, community and learning in a networked virtual world for kids*. Thesis. Massachusetts Institute of Technology. <http://dspace.mit.edu/handle/1721.1/33821>
- [6] Amy Bruckman. 1998. Community support for constructionist learning. *Computer Supported Cooperative Work (CSCW)* 7, 1 (March 1998), 47–86. <https://doi.org/10.1023/A:1008684120893>

- [7] Julie Campbell, Cecilia Aragon, Katie Davis, Sarah Evans, Abigail Evans, and David Randall. 2016. Thousands of Positive Reviews: Distributed Mentoring in Online Fan Communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 691–704. <https://doi.org/10.1145/2818048.2819934>
- [8] Bob Carpenter, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A Probabilistic Programming Language. *Journal of Statistical Software, Articles* 76, 1 (2017), 1–32. <https://doi.org/10.18637/jss.v076.i01>
- [9] Stephen J. Ceci and Wendy M. Williams. 2010. Sex Differences in Math-Intensive Fields. *Current Directions in Psychological Science* 19, 5 (Oct. 2010), 275–279. <https://doi.org/10.1177/0963721410383241>
- [10] Justin Cheng, Lada Adamic, P. Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. 2014. Can cascades be predicted?. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. ACM, New York, NY, USA, 925–936. <https://doi.org/10.1145/2566486.2567997>
- [11] Suzete Chiviawosky and Gabriele Wulf. 2007. Feedback after good trials enhances learning. *Research Quarterly for Exercise and Sport* 78, 2 (March 2007), 40–47. <https://doi.org/10.1080/02701367.2007.10599402>
- [12] Nada Dabbagh and Anastasia Kitsantas. 2012. Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and Higher Education* 15, 1 (Jan. 2012), 3–8. <https://doi.org/10.1016/j.iheduc.2011.06.002>
- [13] Sauvik Das and Adam DI Kramer. 2013. Self-Censorship on Facebook. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM '13)*. AAAI, Palo Alto, California, 120–127. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6093>
- [14] Sayamindu Dasgupta, William Hale, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Remixing as a pathway to computational thinking. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM Press, New York, New York, 1438–1449. <https://doi.org/10.1145/2818048.2819984>
- [15] Sayamindu Dasgupta and Benjamin Mako Hill. 2018. How “wide walls” can increase engagement: evidence from a natural experiment in Scratch. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, New York, 361:1–361:11. <https://doi.org/10.1145/3173574.3173935>
- [16] Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer. 2010. Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-efficacy. *ACM Trans. Comput.-Hum. Interact.* 17, 4 (Dec. 2010), 18:1–18:24. <https://doi.org/10.1145/1879831.1879836>
- [17] Champika Fernando. 2014. *Online learning webs : designing support structures for online communities*. Thesis. Massachusetts Institute of Technology. <http://dspace.mit.edu/handle/1721.1/95602>
- [18] María Fernández-Toro and Stella Hurd. 2014. A model of factors affecting independent learners’ engagement with feedback on language learning tasks. *Distance Education* 35, 1 (Jan. 2014), 106–125. <https://doi.org/10.1080/01587919.2014.891434>
- [19] Deborah A. Fields, Yasmin B. Kafai, and Michael T. Giang. 2017. Youth Computational Participation in the Wild: Understanding Experience and Equity in Participating and Programming in the Online Scratch Community. *ACM Trans. Comput. Educ.* 17, 3 (Aug. 2017), 15:1–15:22. <https://doi.org/10.1145/3123815>
- [20] Deborah A. Fields, Katarina Pantic, and Yasmin B. Kafai. 2015. “I have a tutorial for this”: the language of online peer support in the Scratch programming community. In *Proceedings of the 14th International Conference on Interaction Design and Children (IDC '15)*. ACM, New York, NY, USA, 229–238. <https://doi.org/10.1145/2771839.2771863>
- [21] Laura Guillén, Margarita Mayo, and Natalia Karelai. 2016. The competence-confidence gender gap: Being competent is not (always) enough for women to appear confident. (2016). <https://margaritamayo.com/wp-content/uploads/2016/07/The-competence-confidence-gender-gap.pdf> (working paper).
- [22] Christian Gütl, Rocaél Hernández Rizzardini, Vanessa Chang, and Miguel Morales. 2014. Attrition in MOOC: Lessons Learned from Drop-Out Students. In *Learning Technology for Education in Cloud. MOOC and Big Data: Third International Workshop, LTEC 2014, Santiago, Chile, September 2-5, 2014. Proceedings*, Lorna Uden, Jane Sinclair, Yu-Hui Tao, and Dario Liberona (Eds.). Springer International Publishing, Cham, 37–48. https://doi.org/10.1007/978-3-319-10671-7_4 DOI: 10.1007/978-3-319-10671-7_4
- [23] Eszter Hargittai and Gina Walejko. 2008. THE PARTICIPATION DIVIDE: Content creation and sharing in the digital age. *Information, Communication & Society* 11, 2 (March 2008), 239–256. <https://doi.org/10.1080/13691180801946150>
- [24] E. Tory Higgins. 1987. Self-discrepancy: A theory relating self and affect. *Psychological Review* 94, 3 (1987), 319–340. <https://doi.org/10.1037/0033-295X.94.3.319>
- [25] Benjamin Mako Hill and Andres Monroy-Hernandez. 2013. The cost of collaboration for code and art: evidence from a remixing community. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. ACM, New York, NY, USA, 1035–1046. <https://doi.org/10.1145/2441776.2441893>
- [26] Benjamin Mako Hill, Andrés Monroy-Hernández, and Kristina Olson. 2010. Responses to remixing on a social media sharing website. In *Proceedings of the 4th International Conference on Web and Social Media (ICWSM '10)*. AAAI, Palo

- Alto, California, 74–81. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1533>
- [27] Jerry L. Hintze and Ray D. Nelson. 1998. Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician* 52, 2 (1998), 181–184. <https://doi.org/10.1080/00031305.1998.10480559>
- [28] Sarah Hurlburt. 2008. Defining Tools for a New Learning Space: Writing and Reading Class Blogs. *MERLOT Journal of Online Learning and Teaching* 4, 2 (June 2008), 182–189.
- [29] Nicole M. Hutchins, Ningyu Zhang, and Gautam Biswas. 2017. The role gender differences in computational thinking confidence levels plays in STEM applications. In *The International Conference on Computational Thinking Education (CTE '17)*. The Education University of Hong Kong, Hong Kong, 34–38. <http://www.eduhk.hk/cte2017/doc/CTE2017%20Proceedings.pdf#page=45> Archived at: <https://perma.cc/4FJA-WM2V>.
- [30] Henry Jenkins, Ravi Purushotma, Margaret Weigel, Katie Clinton, and Alice J. Robison. 2009. *Confronting the challenges of participatory culture: media education for the 21st century*. The MIT Press, Cambridge, MA. <https://mitpress.mit.edu/books/confronting-challenges-participatory-culture>
- [31] Y. Kafai, Deborah Fields, and W. Burke. 2010. Entering the Clubhouse: Case Studies of Young Programmers Joining the Online Scratch Communities. *Journal of Organizational and End-User Computing* 2, 22 (Jan. 2010), 21–35. http://digitalcommons.usu.edu/itls_facpub/181
- [32] Yasmin B. Kafai. 2016. From Computational Thinking to Computational Participation in K–12 Education. *Commun. ACM* 59, 8 (July 2016), 26–27. <https://doi.org/10.1145/2955114>
- [33] Yasmin B. Kafai and Quinn Burke. 2014. *Connected code: Why children need to learn programming*. MIT Press, Cambridge, Massachusetts.
- [34] Matthew Kay, Gregory L. Nelson, and Eric B. Hekler. 2016. Researcher-centered design of statistics: Why Bayesian statistics better fit the culture and incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4521–4532. <https://doi.org/10.1145/2858036.2858465>
- [35] Joy Kim, Maneesh Agrawala, and Michael S. Bernstein. 2017. Mosaic: Designing Online Creative Communities for Sharing Works-in-Progress. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 246–258. <https://doi.org/10.1145/2998181.2998195>
- [36] Chen-Chung Liu, Kuan-Hsien Lu, Leon Yufeng Wu, and Chin-Chung Tsai. 2016. The Impact of Peer Review on Creative Self-Efficacy and Learning Performance in Web 2.0 Learning Activities. *Educational Technology & Society* 19, 2 (2016), 286–297.
- [37] Andres Lombana-Bermudez. 2017. Moderation and Sense of Community in a Youth-Oriented Online Platform: Scratch's Governance Strategy for Addressing Harmful Speech. <https://medium.com/berkman-klein-center/moderation-and-sense-of-community-in-a-youth-oriented-online-platform-scratches-governance-eaac6941e9c9>. (2017).
- [38] John H. Maloney, Kylie Peppler, Yasmin Kafai, Mitchel Resnick, and Natalie Rusk. 2008. Programming by Choice: Urban Youth Learning Programming with Scratch. *SIGCSE Bull.* 40, 1 (March 2008), 367–371. <https://doi.org/10.1145/1352322.1352260>
- [39] Hazel Markus and Paula Nurius. 1986. Possible selves. *American Psychologist* 41, 9 (1986), 954–969. <https://doi.org/10.1037/0003-066X.41.9.954>
- [40] Alice E. Marwick and danah boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13, 1 (Feb. 2011), 114–133. <https://doi.org/10.1177/1461444810365313>
- [41] Allison Master, Sapna Cheryan, Adriana Moscatelli, and Andrew N. Meltzoff. 2017. Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology* 160 (Aug. 2017), 92–106. <https://doi.org/10.1016/j.jecp.2017.03.013>
- [42] J. Nathan Matias, Sayamindu Dasgupta, and Benjamin Mako Hill. 2016. Skill progression in Scratch revisited. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM Press, New York, New York, 1486–1490. <https://doi.org/10.1145/2858036.2858349>
- [43] MIT Media Lab. 2013. Kids coding in the cloud. (May 2013). <http://news.mit.edu/2013/scratch-two-released-0514>
- [44] MIT Scratch Team. 2018. Scratch Community Guidelines. (2018). <https://scratch.mit.edu/>
- [45] Peter R. Monge. 1990. Theoretical and Analytical Issues in Studying Organizational Processes. *Organization Science* 1, 4 (Nov. 1990), 406–430. <https://doi.org/10.1287/orsc.1.4.406>
- [46] Richard D. Morey, Rink Hoekstra, Jeffrey N. Rouder, Michael D. Lee, and Eric-Jan Wagenmakers. 2016. The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review* 23, 1 (Feb. 2016), 103–123. <https://doi.org/10.3758/s13423-015-0947-8>
- [47] Oded Nov, Mor Naaman, and Chen Ye. 2009. Motivational, Structural and Tenure Factors that Impact Online Community Photo Sharing. *International AAAI Conference on Web and Social Media; Third International AAAI Conference on Weblogs and Social Media* (2009). <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/206>
- [48] Seymour Papert. 1980. *Mindstorms: Children, computers, and powerful ideas*. Basic Books, New York, New York.

- [49] Seymour Papert and Idit Harel. 1991. Situating constructionism. In *Constructionism*. Vol. 36. Ablex Publishing, New York, NY, US, 1–11.
- [50] Thomas M. Philip, Sarah Schuler-Brown, and Winmar Way. 2013. A Framework for Learning About Big Data with Mobile Technologies for Democratic Participation: Possibilities, Limitations, and Unanticipated Obstacles. *Technology, Knowledge and Learning* 18, 3 (April 2013), 103–120. <https://doi.org/10.1007/s10758-013-9202-4>
- [51] Jean Piaget. 1970. *Genetic epistemology*. Trans. E. Duckworth. Columbia University Press, New York, NY, US.
- [52] Mitchel Resnick. 2007. All I Really Need to Know (About Creative Thinking) I Learned (by Studying How Children Learn) in Kindergarten. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition (C&C '07)*. ACM, New York, NY, USA, 1–6. <https://doi.org/10.1145/1254960.1254961>
- [53] Mitchel Resnick. 2017. *Lifelong kindergarten: Cultivating creativity through projects, passion, peers, and play*. The MIT Press, Cambridge, Massachusetts.
- [54] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, and Yasmin Kafai. 2009. Scratch: Programming for all. *Commun. ACM* 52, 11 (Nov. 2009), 60–67. <https://doi.org/10.1145/1592761.1592779>
- [55] Mitchel Resnick, Natalie Rusk, and Stina Cooke. 1999. The Computer Clubhouse: Technology fluency in the inner city. In *High technology and low-income communities: Prospects for the positive use of advanced information technology*, Donald A. Schön, Bishwapriya Sanyal, and William John Mitchell (Eds.). MIT Press, Cambridge, MA, 263–286.
- [56] Gabriela T. Richard and Yasmin B. Kafai. 2016. Blind Spots in Youth DIY Programming: Examining Diversity in Creators, Content, and Comments Within the Scratch Online Community. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1473–1485. <https://doi.org/10.1145/2858036.2858590>
- [57] Cecilia L. Ridgeway. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford University Press, New York, NY.
- [58] Ricarose Roque, Yasmin Kafai, and Deborah Fields. 2012. From Tools to Communities: Designs to Support Online Creative Collaboration in Scratch. In *Proceedings of the 11th International Conference on Interaction Design and Children (IDC '12)*. ACM, New York, NY, USA, 220–223. <https://doi.org/10.1145/2307096.2307130>
- [59] Ricarose Roque, Natalie Rusk, and Amos Blanton. 2013. Youth roles and leadership in an online creative community. In *Proceedings of the 10th International Conference on Computer-Supported Collaborative Learning*, Vol. 1. International Society of the Learning Sciences (ISLS), Madison, WI, 399–405. <https://www.isls.org/cscl/2013/Volume%201%20Final%20CSCL%202013%20Proceedings.pdf>
- [60] Natalie Rusk. 2016. Motivation for Making. In *Makeology: Makers as Learners*, Kylie Peppler, Erica Rosenfeld Halverson, and Yasmin B. Kafai (Eds.). Routledge, New York, NY, 85–108.
- [61] Christopher Scaffidi and Christopher Chambers. 2011. Skill progression demonstrated by users in the Scratch animation environment. *International Journal of Human-Computer Interaction* 28, 6 (June 2011), 383–398. <https://doi.org/10.1080/10447318.2011.595621>
- [62] Daniel Schugurensky. 2000. *The forms of informal learning: Towards a conceptualization of the field*. NALL Working Paper 19. Centre for the Study of Education and Work (OISE/UT), Toronto, Canada. <https://tspace.library.utoronto.ca/handle/1807/2733>
- [63] Steven J. Spencer, Christine Logel, and Paul G. Davies. 2016. Stereotype Threat. *Annual Review of Psychology* 67, 1 (Jan. 2016), 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235>
- [64] Stan Development Team. 2017. RStanArm: Bayesian applied regression modeling via Stan. (2017). <http://mc-stan.org>
- [65] MIT Scratch Team. 2018. Scratch - Imagine, Program, Share. (2018). <https://scratch.mit.edu/statistics/>
- [66] Dustin B. Thoman, Jessi L. Smith, Elizabeth R. Brown, Justin Chase, and Joo Young K. Lee. 2013. Beyond Performance: A Motivational Experiences Model of Stereotype Threat. *Educational psychology review* 25, 2 (June 2013), 211–243. <https://doi.org/10.1007/s10648-013-9219-1>
- [67] Douglas Thomas and John Seely Brown. 2011. *A New Culture of Learning: Cultivating the Imagination for a World of Constant Change*. CreateSpace Independent Publishing Platform.
- [68] Ema Ushioda. 2011. Why autonomy? Insights from motivation theory and research. *Innovation in Language Learning and Teaching* 5, 2 (July 2011), 221–232. <https://doi.org/10.1080/17501229.2011.577536>
- [69] N. F. Velasquez, D. A. Fields, D. Olsen, T. Martin, M. C. Shepherd, A. Strommer, and Y. B. Kafai. 2014. Novice Programmers Talking about Projects: What Automated Text Analysis Reveals about Online Scratch Users' Comments. In *2014 47th Hawaii International Conference on System Sciences*. IEEE, Waikoloa, HI, 1635–1644. <https://doi.org/10.1109/HICSS.2014.209>
- [70] Gregory M. Walton, Geoffrey L. Cohen, David Cwir, and Steven J. Spencer. 2012. Mere belonging: the power of social connections. *Journal of Personality and Social Psychology* 102, 3 (March 2012), 513–532. <https://doi.org/10.1037/a0025731>
- [71] Zhendong Wang, Yi Wang, and David Redmiles. 2018. Competence-confidence Gap: A Threat to Female Developers' Contribution on Github. In *Proceedings of the 40th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS '18)*. ACM, New York, NY, USA, 81–90. <https://doi.org/10.1145/3183428.3183437>

- [72] Robin L. West, Dana K. Bagwell, and Alissa Dark-Freudeman. 2005. Memory and Goal Setting: The Response of Older and Younger Adults to Positive and Objective Feedback. *Psychology and Aging* 20, 2 (June 2005), 195–201. <https://doi.org/10.1037/0882-7974.20.2.195>
- [73] Anna Woodcock, Paul R. Hernandez, and P. Wesley Schultz. 2016. Diversifying Science: Intervention Programs Moderate the Effect of Stereotype Threat on Motivation and Career Choice. *Social psychological and personality science* 7, 2 (March 2016), 184–192. <https://doi.org/10.1177/1948550615608401>
- [74] Seungwon Yang, Carlotta Domeniconi, Matt Reville, Mack Sweeney, Ben U. Gelman, Chris Beckley, and Aditya Johri. 2015. Uncovering trajectories of informal learning in large online communities of creators. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale (L@S '15)*. ACM Press, New York, New York, 131–140. <https://doi.org/10.1145/2724660.2724674>
- [75] José Pablo Zagal and Amy S. Bruckman. 2005. From Samba Schools to Computer Clubhouses: Cultural institutions as learning environments. *Convergence* 11, 1 (2005), 88–105. <http://journals.sagepub.com/doi/pdf/10.1177/135485650501100107>

Received April 2018; revised July 2018; accepted September 2018