Coding by Choice: A Transitional Analysis of Social Participation Patterns and Programming Contributions in the Online Scratch Community

Deborah A. Fields, Yasmin B. Kafai* and Michael T. Giang◊

Utah State University
*University of Pennsylvania
◊Mount Saint Mary's University

Abstract. While massive online communities have drawn the attention of researchers and educators on their potential to support active collaborative work, knowledge sharing, and user-generated content, few studies examine participation in these communities at scale. The little research that does exist attends almost solely to adults rather than communities to support youths' learning and identity development. In this chapter, we tackle two challenges related to understanding social practices that support learning in massive social networking forums where users engage in design. We examined a *youth* programmer community, called scratch.mit.edu, that garners the voluntary participation of millions of young people worldwide. We report on *site-wide* distributions and patterns of participation that illuminate the relevance of different online social practices to ongoing involvement in the online community. Drawing on a random sample of more than 5,000 active users of Scratch.mit.edu over a three-month time period in early 2012, we examine log files that captured the frequency of three types of social practices that contribute to enduring participation: DIY participatory activities, socially supportive actions, and socially engaging interactions. Using latent transition analysis, we found 1) distinct patterns of participation (classes) across three time points (e.g., high networkers who are generally active, commenters who focus mainly on social participation, downloaders engaging in DIY participatory activities), 2) unique migration changes in class membership across time, 2) relatively equal gender representation across these classes, and 4) importance of membership length (or age) in terms of class memberships. In the discussion we review our approach to analysis and outline implications for the design and study of online communities and tools for youth.

Introduction

A growing body of research in massive online communities, often defined by having millions of voluntary users, has sought to understand patterns of participation in online sites, games, social networking sites, and virtual worlds (e.g., boyd, 2013; Gee, 2003). Research on participation patterns and profiles in these massive communities have provided insights into how people participate and can develop collaborations within and beyond the designed structures; for instance, by developing fluid social networks for information gathering and gameplay (Williams, Contractor, Poolec, Srivastad, & Cale, 2011), by building trust in long-term relationships that promote more effective teamwork (Chen, 2012), and by engaging in knowledge sharing and problem solving in game forums (Steinkuehler & Duncan, 2009). Yet collaboration in massive sites is often more diffuse and less obvious than the clear teamwork or knowledge building described above. For instance, some studies (see boyd, 2013; Ito et al, 2009), have illustrated how younger users (children and youth) engage in social practices that are less directly "collaborative" but en masse result in distributed peer support for learning through knowledge diffusion (Fields & Kafai, 2009, 2010), praise and constructive criticism (e.g., Black, 2008), or simply engagement with an interested audience (e.g., Magnifico, 2010).

These social practices also hold great potential in an emerging genre of online communities where socializing centers around things that people create: do-it-yourself (DIY) social networking forums. These DIY social networking forums differ from the more typically thought of social network sites (SNS, see boyd and Ellison, 2007) like MySpace and Facebook where user participation focuses on reports of daily life (Grimes & Fields, 2012). Instead, DIY social networking forums are communities where participants share their own self-created DIY media and where communication, profile pages, and networking residues all focus in some way on user-created projects that range from film to fanfiction to programming (Grimes & Fields, 2015). Commonly designed social networking features such as comments, "likes," favorites, or even the simple ability to share projects form a baseline of social support and encourage young writers, programmers, and artists to pursue their personal interests (e.g., Black, 2006; Resnick et al, 2009). Yet despite these potentials, studies are rare for *youth* amateur design communities. Most of the current research has focused on adults' online activities (e.g., Benkler, 2006; Luther, Caine, Ziegler, & Bruckman, 2010), possibly because such communities are easier to access and study due to participants' age and overlapping interests with that of researchers (Kafai & Fields, 2013).

Further, with millions of participants and projects, it is not always clear who is participating and what they are contributing; thus patterns and trends that might reveal issues of equity in participation are not easily discerned by the naked eye. Participating in these sites can be a rich but also challenging experience, in particular for youth as we have observed (Kafai, Fields & Burke, 2010). Creative participation that involves highly technical expertise such as programming has only recently received more attention due to the increased interest in promoting computational thinking and access to computing (Kafai & Burke, 2014). Some ethnographic studies illustrate that many youth do not engage in activities that hold the most potential for learning through creating, perhaps socializing and "messing around" rather than creating and posting content or "geeking out" as Ito et al. (2009) express. Larger scale studies reveal differential patterns of participation (and by extension, possible collaborations) in massive online communities with often a relatively small group of established users driving the majority of the interactions and contributions, raising concerns about equity and diversity in participation (e.g., Giang, Kafai, Fields & Searle, 2012; Yee, 2014). At the same time, issues of broadening access and deepening participation are particularly critical because of technology's longstanding history of underrepresentation of women and minorities (Margolis & Fisher, 2002; Margolis et al., **2008**). At the same time, Understanding site-wide patterns of youth amateur design communities can allow us to make more informed decisions on how to design for collaborative supports and learning at a massive level, as well as identify which users (e.g., gender or newbies) may need scaffolds in participating in such large DIY communities.

In this chapter, we tackle two challenges related to understanding social practices that support learning in social networking forums where users engage in design. First we study a *youth* programmer community called Scratch.mit.edu that garners the voluntary participation of millions of young people worldwide. Second, we report on *site-wide* distributions and patterns of participation that illuminate the relevance of different online social practices to ongoing involvement in the Scratch online community. Drawing on a random sample of more than 5,000 active users of Scratch.mit.edu over a three-month time period in early 2012 we examine log files that captured the frequency of four types of social practices that contribute to enduring participation: DIY participatory activities, socially supportive actions, socially engaging interactions, and identity building activities. We apply latent transition analysis (LTA) to investigate the following questions: 1) What types of users shape the Scratch online community and what combinations (or patterns) of social practices differentiate their participation? 2) Do gender and length of membership play a role in these patterns of participation? 3) Finally, what changes in participation can we see over time? In the discussion we consider what this says about who is getting the most out of their participation and what designed-for practices may contribute to longstanding engagement in these massive online communities. We review our approach to analysis and outline implications for the design and study of online communities and tools for youth.

Background

Understanding Collaborative Learning in Massive Online Communities

We situate our study in the larger context of research conducted on *collaborative* learning, which for the most part has focused on youths' abilities to interact in and contribute to small groups inside and outside of schools. Hundreds, if not thousands, of studies have investigated various aspects of collaborative learning, including the nature of various group arrangements such as reciprocal teaching or jigsaw techniques, interactions with members of different gender, race, ability, and experience, and causes for success and failures of group work (for general overviews, see O'Donnell, 2006; Webb & Palincsar, 1996). Studies that examine collaboration in larger groups, especially with the support of computers, are only now beginning to develop such as knowledge base. Most notable is here the work on the Computer Supported Intentional Learning Environments (CSILE; now: Knowledge Forum) (Scardamalia & Bereiter, 1991) and other studies following the knowledge forum tradition, as they have examined how students' knowledge-sharing, knowledge-construction, knowledge-creation, and knowledge-assessment come into play through student-driven inquiry that builds knowledge at a community level (e.g., Ares, 2008; Eddy, Chan & van Aalst, 2006; van Aalst, 2009). Most CSILE implementations have operated within a classroom environment, sometimes connecting students from other classes or previous years through an emphasis on collective cognitive responsibility (Scardamalia, 2002; Zhang, Scardamalia, Reeve, & Messina, 2009). What becomes apparent from these studies is that productive collaborative interactions can take place at larger scales beyond small groups, through a mixture of unstructured and structured groups, concurrent asynchronous and synchronous interactions, and persistent shared virtual environments that can hold community-level knowledge. Yet even these relatively larger scale studies are quite far away in scale from the size of social media communities today, particularly communities that rely on voluntary participation outside of any classroom requirements (e.g., Rick & Guzdial, 2006).

A number of studies have identified key social practices that can support users' learning and deepening participation in massive online DIY social networking forums. Black's (2008) work on youth' fanfiction sites documents the importance of peer feedback in the form of comments on multiple iterations of written fiction projects. Users share sections of stories, solicit readers' feedback, and revise their writing based on the comments of others. Enthusiastic comments, often expressed in desires to see new or extended work, can encourage youth to stick with their writing or even their programming (see also, Brennan, Valverde, Prempeh, Roque & Chung, 2011). Related, Magnifico (2010) theorizes about this important role of audience that online communities can provide for users' work. Learning to write, program, or draw in order to gain the attention and interest of an online audience can focus youths' creative work in ways rarely available in classrooms (e.g., Lammers, Magnifico & Curwood, 2014). While comments may be the most obvious evidence of an authentic audience, other traces of users' viewing of one's work, called networking residues (Grimes & Fields, 2012), also provide feedback. These networking residues may include traces such as "love-its," friend requests, "favorites," "likes," comments, replies, downloads, and even gifts depending on what websites record and display on users' artifacts and profiles. They may even become a type of commodity as they elevate the virtual presence of a person or project through signs of popularity. Many DIY social networking websites organize their front pages by featuring "most liked" or "most viewed" designs.

We have established that these participatory practices from social networking forums focused on story and digital media productions also apply to youth software production communities such as Scratch. Our recent study of a random selection of comments about projects on the Scratch website points to the overall positive ethos of the site where 72% of these comments were positive and 14% were neutral in emotional tone (Fields, Pantic & Kafai, 2015). We also found that about half of the comments were generally constructive: 58% contained at least some minor level of detail in the feedback beyond more generic "Awesome" or "Cool!" statements. Our analysis of the purpose of these comments supports the idea that motivationally encouraging feedback is key in shaping participation on the Scratch site (58% of the comments) and that drawing an audience to one's work is also a significant felt need amongst users (23% of the comments). These outcomes support the findings of other studies of the Scratch community that have documented how members solicit and leverage networking residues to support user-created design contests, offering projects, illustrations, love-its, and friending as prizes (Nickerson & Monroy-Hernandez, 2011).

Beyond the more obvious social practices of commenting or otherwise leaving markers of audience in DIY social networking forums, Grimes and Fields (2015) point to the importance of simply sharing one's creations. Sharing projects online makes them visible to others for feedback, viewing, and remixing; this is a key feature that is often missing in website design for children. Transparency of projects is necessary in order to leave and receive feedback, even in the milder forms of a thumbs-up or a "like" button. In the Scratch community transparency of projects goes a step further than most DIY social networking forums (Grimes & Fields, 2015) in that it enables users to download, see inside, and even remix others' projects. Remixing projects involves taking someone's existing work, changing something about it (whether a minor or major change), and re-sharing it online. Remixing can provide an opportunity to learn by seeing how someone's project works and exploring what various changes do. It can also solicit a form of fandom when users post projects intended for remixing (e.g., adding a character to a dance party project) or even use remixing as a way to exchange projects in collaborative work (see Monroy Hernandez, 2012).

While there may be a range of collaborative practices available to users on DIY social networking forums, thus far it has been difficult to evaluate how widespread or distributed these practices are across a full range of users as well as whether and how these patterns shift over time. Our work aims at filling some of these gaps, in particular identifying patterns of social practices found in the Scratch youth amateur design (or do-it-yourself, DIY) community that is the focus of this paper. Here we consider less the smaller enterprises of small collaborative groups who work together on shared projects online in favor of studying broader dynamics of participation in amateur design communities. Although there are growing numbers of such communities where youth share designed artifacts such as art (e.g., Deviant Art, Bitstrips), mods of games (e.g., Little Big Planet, the Sims), or stories (e.g., Fanfiction.net, Storybird), we know little about who is participating in what practices and for how long. To contribute to a framework for understanding "mass collaboration", we analytically bring together different designed-for social practices that support participation on a massive scale, from creating to remixing to commenting to favoriting, and investigate who engages in these practices, in what combinations of activity, and for what duration. Our larger goal is to understand what the large numbers reveal about participation and collaboration that is not visible at smaller scales.

Researching Collaborative Learning in Massive Online Communities

As described above, in youth amateur design communities, many different types of activities contribute to the community and provide supports for learning to design. To study these activities at a massive scale we need to identify key types of practices that can be tracked through backend (or log file) website data. While case study and ethnographic research can illuminate the roles these practices play in learning (see above section), quantitative or analytical research must be used to understand patterns of use at a large scale. Based on the research concerning feedback, audience, networking residues, sharing, and remixing, we identified three types of social practices that may contribute to learning that are recorded and identifiable in log file data:

- **DIY participatory activities**: These activities involve *sharing* projects users have created, *remixing* projects (editing and posting changes to another's project) and *downloading* projects. They primarily involve users creating, sharing, and editing content which are innately but not obviously social. In other words, they do not involve direct social interaction with another user.
- **Socially supportive actions**: These actions include socially oriented actions that are supportive but do not directly engage a response from a user. They include simple networking residues that can be left with a simple click, such as *loving* projects (clicking "love-it" on projects one likes) and *favoriting* projects (clicking "favorite" on a project).
- **Socially engaging interactions**: Certain actions on the social networking forums are more directly interactive, namely writing comments on a user's project or submitting a friend request. We consider these more socially engaging as they invite and are more likely to

generate a response. *Comments* provide an opportunity for conversation (many users actually respond to others' comments). *Friending* another user results in a notification to that user (implicitly inviting a responsive friending action) and allows the requester to get notifications of that user's new projects.

Equally important to understanding social practices that form participation, is understanding of *who* engages in the various forms of participation designed for on DIY social networking forums. The sheer number of projects and members, often reaching millions in these massive online communities, can mask differential levels of participation amidst the seemingly endless activity on sites. The few studies that have been able to assess participants often find that a relatively small number of members, between 5-10% is highly active and generates most of the social interactions and content, while other, larger groups range from more distant involvement to simply being onlookers (see Kafai and Fields' [2013] research of a virtual world and Yee's[2014] research of gaming communities). The issue of access and participation becomes even more salient when we look at motivations, or the lack thereof, of new members joining such massive communities: not everyone is interested in becoming a central member of online communities (Kafai, Fields & Burke, 2010). While newcomer membership is an important factor in judging participation in online communities, there are unanswered questions about how diverse and open such communities are in inviting in others. In particular digital communities pertaining to gaming and computing are predominantly male, with few exceptions (e.g., Kafai, Heeter, Denner & Sun, 2008), thus replicating discrepancies found in the technology culture at large, whether the participants are adults or youth.

Whether and how users participate are of relevance when examining youth amateur communities that focus on making and sharing programming designs, as will be the case in this paper. While the Scratch community consists of one-third female users, thus far we have not been able to judge the extent to which they engage in contributing and collaborating in the online site. Further, almost no attention has been paid to whether length of membership on a site influences the types of social practices users engage in. To understand the practices that form the underlying social fabric that encourages and supports continuing participation in a youth design community, we examine whether and how users engage in different activities though an analyses of log files from a random sample of users in the Scratch online community. We look systematically at the massive scale of participation in Scratch, asking what patterns of participation users exhibit, how this changes over time, and how these patterns relate to gender or length of user membership in the Scratch site.

Scratch Online Community

Scratch.mit.edu is an online massive community where participants, mostly youth ages 11-18 years share their computer programs (Resnick et al., 2009; see also Roque & Resnick, this volume). Kids who share an interest in programming post animations, games, stories, science simulations, and the interactive art they have made in the visual programming environment of Scratch (see Figures 1 and 2). Scratch is a visual programming environment, allowing designers to create various media through a process of dragging-and-dropping command blocks of code then stacking these blocks together to form coding scripts that can become increasingly complex and nuanced depending upon a user's facility with coordinating a range of command blocks through programming concepts such as loops, synchronization, variables, conditionals, and more (Maloney et. al, 2008).

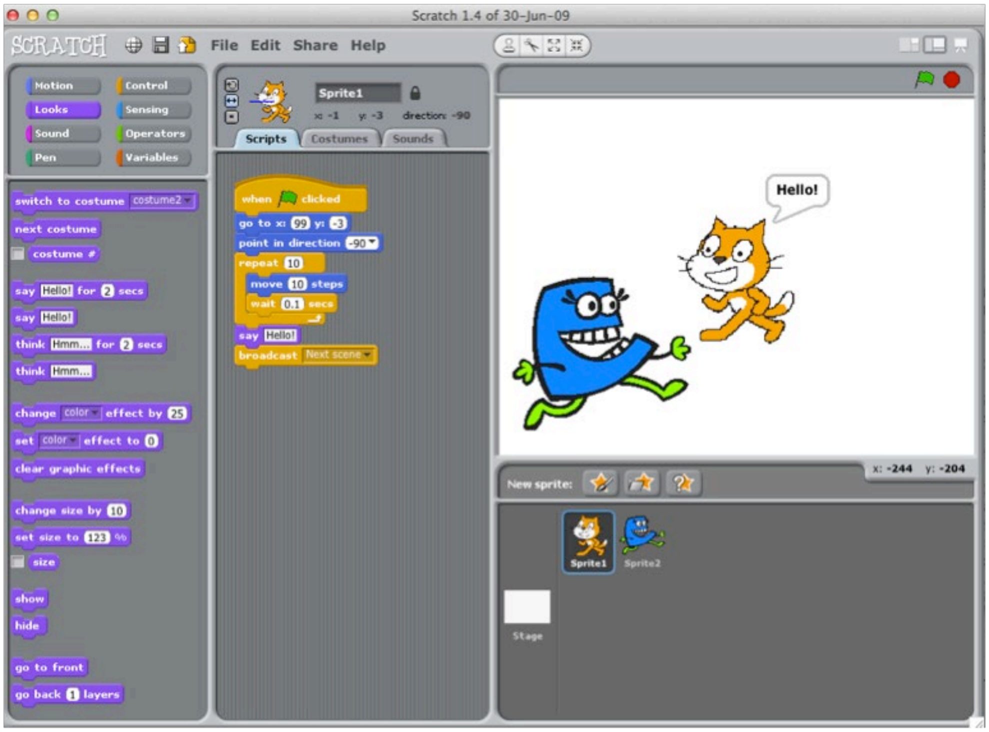

Fig. 1 Scratch programming interface (Version 1.4)

Launched in May 2007, as of May 2015 the Scratch site has grown to more than 6 million registered members with nearly 1500 Scratch projects shared everyday (a total of 9 million projects since 2007). Notably, the data from this investigation come from the version of the Scratch site that existed from May 2007 through May 2013, familiarly known as Scratch 1.0 (to 1.4). In May 2013, the Scratch Team released a new version of the site that had several new design features. Most notably, the site now allows users to program projects directly on the site and to edit (remix) others' projects without having to download them. In other words, users do not need to program offline (though they can) and subsequently share their projects online. Instead they can simply program online without having to upload. They can also "see inside" the code of others' projects without having to download them. With these changes, user participation on the site has nearly quintupled (up to 90,000 active users every month). While our study examines the data from Scratch 1.0, the social networking features and project uploads we studied continue to be key participation practices in the Scratch 2.0 online community.

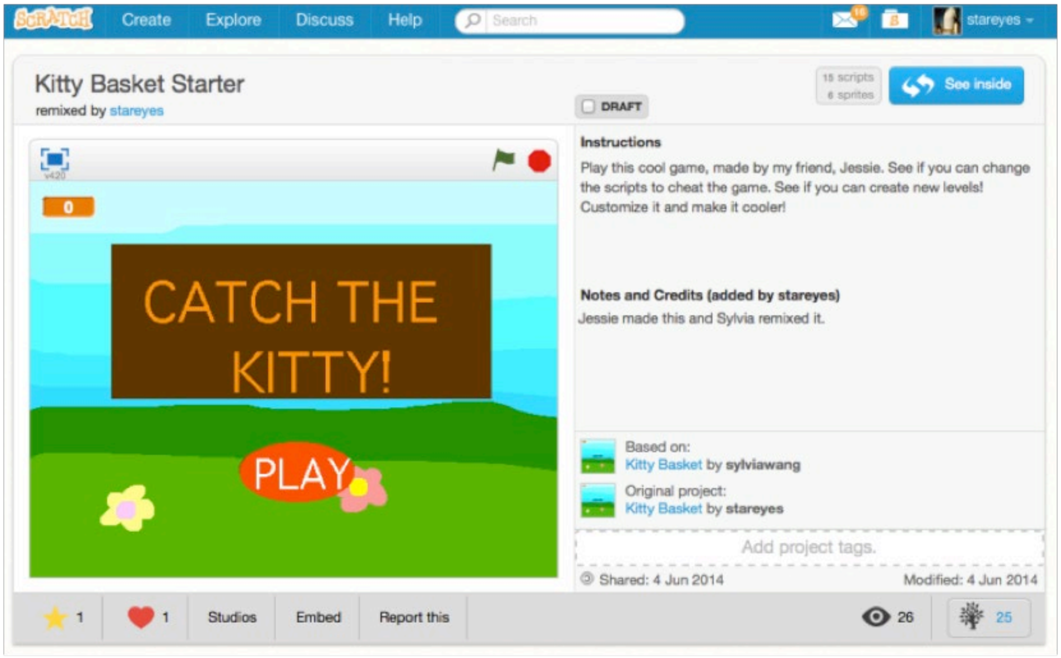

Fig. 2 User Profile on Scratch.mit.edu (of the first author)

We chose several types of social networking features to study, namely designed-for activities that shape user interaction on the Scratch site and that were also available through backend log file data. In our study these include DIY participatory activities (*sharing, downloading,* and *remixing* projects), socially supportive actions (*loving* and *favoriting* projects), and socially engaging interactions (*commenting* and *friending*). Although these activities are done by individual users, as a whole they leave traces of users' views and opinions on projects, demonstrating the presence of an audience. Accumulations of high numbers can result in a user's project being posted on the prized front page of the Scratch site (see Figure 3) through categories like "Featured Projects," "What the Community is Loving," and "What the Community is Viewing." Thus we chose these activities as a lens of social practices that shape mass participation, for both convenience of data collection, the breadth of user practices they demonstrate, and their prominence in shaping the Scratch site,

One goal of our analysis in this paper is to reveal how these networking activities are related to programming activities. In a related study we examined computational participation on the Scratch site by analyzing trends in users' posted projects, using the same sample of users as this study, (Fields, Giang & Kafai, 2014). In this case we used latent class analysis on sets of programming practices rather than social practices to understand classes of Scratch programmers. Using categories of programming blocks such as loops, variables, operators, broadcasts, and Booleans, we found four stable and cohesive classes of programmers in the Scratch community that reflect a range of experience based on their use of programming concepts. Beginners tended to create smaller, simpler projects with a small number of loops and none of the other more advanced programming concepts. Intermediate users created slightly larger (middle-sized) projects and also included variables, operators, and broadcasts. Advanced users utilized all of those concepts as well as Booleans in middle-sized projects. Experienced users were similar to the Advanced users except that their projects were much larger in terms of using increased numbers of commands in all of the concepts studied. Looking at gender and length of membership, we found that girls were disproportionately represented in the Beginner class and likewise underrepresented in the Advanced and Experienced class. We found little relationship in terms of length of membership except for a slight under-representation of the newest users (newbies) and an over-representation of the longest users (oldies) in the Experienced class. We return to these findings in the discussion, considering the relationship between programming and participation on the Scratch site and reporting how findings reported in this paper relate to analyses about programming.

Fig. 3 The front page of scratch.mit.edu (Version 1.0 2007-2012)

Methods & Analyses

Data Sample: Participants

Our analysis initially drew from a random sample of 5,004 users drawn from amongst more than 20,000 users who logged into Scratch during the month of January 2012. While Scratch usage fluctuates month to month over the course of a year, with summer months usually showing a higher usage than other months, in consultation with Scratch community managers we chose these winter months as times of typical steady participation. Our random sample reflected the broader population on Scratch in regard to self-reported gender and age. Members on the Scratch site are self-reported 33% female and 67% male; this distribution was reflected in our random sample. Age on Scratch is also only known through self-report (i.e., whatever birth year the user chooses). Of course, because age and gender are by self-report, users can choose to state otherwise. Other studies of youth online have shown that youth may lie about their age online, sometimes showing a difference between age and grade reporting. This may be because kids gain some social status from being older (see Kafai & Fields, 2013) and because of national laws governing what data websites can collect from youth under 13 years of age (e.g., COPPA in the United States). Many youth may increase their reported age so that they are allowed to participate in social networking sites like Facebook that refuse access to youth under 13 because of these regulations (Grimes & Fields, 2015; Grimes & Fields, 2012). In our sample, the mean age was 20 years old, the median 14, and the mode 12. Since there were a surprising number of individuals (more than 70) who were over 100 years old or under 4 years old (more than 50), we view the averages with great skepticism. (Similarly there are a surprising number of individuals reporting their home country as Antarctica or Aruba). In this paper we focus more on the length of membership of Scratch users rather than their reported age, though the generally accepted age range of the majority of participants on Scratch is 11-18 years of age.

We collected data on this sample of users for three months. During these three months, 1379 users shared an original project in one of the three months (January – March 2012), 533 created a project in each of two months, and 313 created a project in all three months. These 2225 users (67% boys and 33% girls, reflective of the broader Scratch population) who created at least 1 project across a three-month period formed *the new subsample from which all further analyses reported in this paper are drawn*. This sub-sample represents about 44.5% of the initial random sample of users (Fields, Giang & Kafai, 2013). At this moment we do not have access to participation information from other related youth programming sites that can serve as a benchmark for whether this participation rate is standard or unusual. Data collected from the backend of massive online youth communities is notoriously difficult to come by because most companies consider this information proprietary and do not share it with outside members.

The reason we focus on this 2225 sub-sample is that the remaining 2779 users did not engage key activities used for the study. In addition, though these remaining users logged on to Scratch and likely browsed the site during the time of the study, we do not have information about what they did on Scratch. Most likely they viewed webpages without leaving any networking residues: they did not click "like" or favorite projects and they did not leave comments. This data was unavailable because the Scratch Team at MIT did not record this data in log files. This division of users who created and shared projects (and of those some who left comments or "love-its" or favorites) and those who did not create and share projects was a surprise to us. Based on our analyses, sharing a project on the Scratch site defines the baseline of all other active participation beyond viewing. Thus our next step in understanding broad trends of programming and participation on the Scratch website focused on participation profiles of these project creators, identifying how users engaged in downloading, commenting, remixing, "loving," or friending in the online Scratch community, treating it as a type of DIY social networking forum (Grimes & Fields, 2015).

Data Analysis: Latent Class and Transition Analysis

At any given period on Scratch, players engage in multiple modes of participation. As described earlier, we categorize these forms of participation into *DIY participatory activities* (i.e.,., downloading and remixing projects), *socially supportive actions* (i.e.,, loving and favoriting project), and *socially engaging interactions* (i.e., commenting and friending),. We apply latent class analysis (LCA) to identify whether they are distinct types of players who share common modes of participation. LCA's advantages relative to other statistic techniques (e.g., mean splits or cluster analyses) are its conservative ability to identify similar groups of individuals that are uniquely different from other groups (i.e., classes), provide probabilities for classifying individuals into each class, and examine the influence of covariates (e.g., gender) on membership. For instance, LCA can identify whether there are groups (or classes) of players who only focus on DIY participatory activities and do not in engage in more complex social activities and others who do the exact opposite; it then estimates the likelihood that each player is placed into these classes. After these classes have been identified, latent transition analysis (LTA) examines whether individual players transition to different classes of participation across time or stay where they are comfortable (Collins & Lanza, 2010).

This process of analysis begins with LCA. The goal of which is to identify the optimal number of latent classes through an iterative post hoc process (Hagenaars & A. McCutcheon, 2002; Muthen, 2002). For example, given six indicators of participation, LCA would first examine whether a model with two classes (e.g., social vs. non-social players) would provide a better fit than a one-class model (e.g., non-social). If so, LCA continues to test models with additional latent classes until model fit indices and substantive interpretation is satisfactory. The interpretation process examines the participation patterns (based on the extent of use for each indicator in a given class) and the number of individuals in each class to determine whether the specific number of latent classes and membership are meaningful. This LCA process of identifying the optimal number of latent classes is repeated across all time points (e.g., January, February, and March) to determine the number and consistency of classes. Latent transition analysis examines whether and how individuals within these classes change membership across time; in other words it examines the likelihood that novice users remain novices or move onto different forms of participation. Through the same process of analyses, LTA also examines the influence of other variables (e.g., gender, membership time) in the classification process. This would examine whether gender plays a role in participation patterns and whether newbies and oldies (veteran players) utilize Scratch in the same fashion.

In terms of statistical criteria, multiple indicators of model fit are often used as there is no definitive model fit index for these analyses. For this study, the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the sample-size adjusted BIC (aBIC), the Lo-Mendel Rubin Likelihood Ratio Test (LMR-LRT), and entropy values are provided. Model selection is often based on the lowest values on the AIC, BIC, and aBIC, or a scree-like test, in which selection was based on where the indices begin to level off. The LMR-LRT compares models with different numbers of classes, wherein a non-significant value indicates whether a simpler model with one fewer classes provides a better fit for the data. The entropy value is a standardized measure of classification accuracy based on the model's posterior probabilities; this value range from 0 to 1, with higher values reflecting better classification. The (average) posterior probabilities reflect the most likely (or probable) class membership across all users. When the average probabilities for the most likely class are high (above 80%), coupled with low probabilities (below 20%) for the other classes (i.e., misclassification), these numbers suggest good fit. Given the potential ambiguity in model fit indices, the substantive aspect of LCA allows the researcher flexibility in identifying the optimal number of latent classes to balance statistical and theoretical interpretation of each model. This avoids the potential of identifying classes with only a few users or a class that is generally similar to another except for minor statistical differences in specific observed activity.

4.3. Analysis: Gender and Scratch Membership

To further test whether length of membership or gender were proportionately represented in each of the latent classes, multiple chi-square tests for independence analyses were performed for each of the three months. These analyses utilize results from LCA, where each player is classified into a specific latent class (based on how they participate). These tests of independence will then examine whether classes of participation play are linked to gender and length of Scratch membership (the total lifetime of the user's account as of January 2012). Length of membership was distributed across four categories of members: users with brand new accounts created in January 2012 (newbies), users with accounts up to three months old (young), accounts up to 12 months old (one-year), and accounts over one-year old (oldies) (see Table 1A and 1B). Notably, age distribution was roughly equal between the overall sample and the subsample of 2225 project-sharing participants. There was a slightly larger percentage of newbies and slightly smaller percentages of one year and oldie participants, but these differences are small. A significant chi-square test would show that there is a relationship between gender (or membership) and latent classes profiles. Follow-up standardized residual scores test whether the actual count of individuals in a given cell is greater than ($z > |2|$ or $|3|$) or less than expected ($z < |2$ or $3|$) at $p = .05$ or $p = .01$. For example, a significant standardized residual would indicate that the number of females in a given membership class is significantly greater ($z > 2$) or less ($z < 2$) than expected.

Table 1A. Distribution of Scratch membership in entire sample (n=5004).

Scratch Membership	Frequency	Percent of Sample
Newbie (new account)	1436	28.7
Young (0-3 months)	1364	27.3
One Year (4-12 months)	973	19.4
Oldie (12+ months)	1165	23.3
Missing data	66	1.3%

Table 1B. Distribution of Scratch membership amongst project-sharing participants (n=2225).

Scratch Membership	Frequency	Percent of Sample
Newbie (new account)	756	33.9%
Young (0-3 months)	628	28.2%
One Year (4-12 months)	411	18.5%
Oldie (12+ months)	404	18.2%
Missing data	26	1.2%

Findings

Profiles: Project Creators Versus Browsers

Our examination of a random sample of 5,000 users revealed that participation on the Scratch website begins with project-creation. To our surprise, creating and sharing projects was a baseline for all other kinds of online participation, demonstrating the centrality of programming and project creation on the Scratch website and providing a potentially new model for social networking forums from the bottom up. Prior statistics on Scratch participation highlighted only the frequency trends of all users over the entire age of the Scratch site (see: scratch.mit.edu/statistics). From these statistics and from case study research, others have reported that Scratchers tend to prefer either project creation or commenting, usually divided by gender (with male users engaging in more project creation and female users posting more comments; see Brennan, 2011). However, our analyses of Scratchers suggest a different pathway, namely that project-creation is the basic form of participation on the Scratch website (beyond simply browsing which we could not study). Further, nearly all commenters on the Scratch site are also project-sharers. For instance, in the month of January, there were no users who posted comments who did not create at least one project, whereas there were many users who created projects but did not post comments. The simple finding that users who did not create projects largely did not participate in any other traceable way on the Scratch site suggests a new model of social networking forum that focuses on user-created content sharing rather than the more commonly thought of activities conducted on social network sites (boyd & Ellison, 2007) that feature reports of personal daily activity (e.g., Facebook, Vine, Twitter).

Patterns: Transitions in Participation over Time

We conducted latent class analyses conducted to identify the types of participation patterns for each time point (January, February, and March). Each month suggested a different number of classes. Table 2 presents the multiple goodness-of-fit indices for each of the three waves of analyses. For January, a 5-class model provides the most optimal fit based on decreasing model fit indices (BIC, aBIC) and a non-significant LMR-LRT at the 6 class model; in the 5 class models, players had high average probabilities of being classified into a specific class (with the most likely class membership probability between 75.6% and 94.2%) compared to being classified into another class (with a misclassification probability between 0.1% and 24.4%). Moving on to February, the LMR-LRT, aBIC, and substantive interpretation of three different models suggest a 4-class model would provide the most meaningful model. In addition, the average posterior probabilities range from 75.0% to 96.8% for the highest probability class, and between 0.1% and 20.7% for misclassification. For the analyses of March data, the BIC and aBIC hit their lowest point at the 3-class model, and the class sizes and substantive interpretation of the other models also point to a 3-class model. Similar to the other models, the average posterior probabilities for the most likely class membership ranged from 85.5% to 98.0%, and misclassification numbers were between 0.1% to 5.7%.

Table 2

Model-fit indices for participation profiles in January, February, and March 2012.

Note: **Bold** type indicates the best fitting model based on the given fit index.

DICH, JANUARY (N = 2225)							
	likelihood	free par	BIC	aBIC	LMR-LRT p-value	Entropy	AIC
1	-7550.943	6	15148.132	15129.069	N/A	N/A	15113.887
2	-6150.005	13	12400.208	12358.905	0.0000	**0.852**	12326.011
3	-6043.122	20	12240.395	12176.852	0.0009	0.699	12126.245
4	-6006.607	27	12221.316	12135.533	0.0072	0.688	12067.213
5	-5976.236	34	**12214.528**	**12106.504**	**0.0000**	0.839	12020.472
6	-5967.036	41	12250.081	12119.817	0.1129	0.790	**12016.073**

7	**-5960.683**	48	12291.327	12138.824	0.0066	0.821	12017.367

DICH, FEBRUARY (N = 2225)							
	likelihood	free par	BIC	aBIC	LMR-LRT p-value	Entropy	AIC
1	-6260.815	6	12567.875	12548.812	n/a	n/a	12533.630
2	-4158.034	13	8416.265	8374.962	0.0000	**0.943**	8342.067
3	-4040.316	20	**8234.781**	8171.238	0.0000	0.868	8120.631
4	-4017.073	27	8242.249	**8156.466**	**0.0183**	0.859	8088.147
5	-4007.576	34	8277.207	8169.184	0.0799	0.894	**8083.152**
6	-4002.081	41	8086.161	8320.169	0.1361	0.910	8086.161
7	**-3997.482**	48	8364.925	8212.421	0.3965	0.856	8090.964

DICH, MARCH (N = 2225)							
	likelihood	free par	BIC	aBIC	LMR-LRT p-value	Entropy	AIC
1	-5400.273	6	10846.791	10827.728	N/A	N/A	10812.546
2	-3354.877	13	6809.952	6768.649	.0000	**.963**	6735.754
3	-3250.281	20	**6654.712**	**6591.168**	.0000	.904	6540.561
4	-3238.291	27	6684.684	6598.901	.0011	.937	6530.581
5	-3228.587	34	6719.229	6611.206	**.0312**	.947	6525.174
6	-3219.988	41	6755.984	6625.720	.1582	.939	**6521.976**
7	**-3213.203**	48	6796.367	6643.864	.3879	.938	6522.407

Based on the latent class analysis results, we next conducted latent transitions analyses to examine whether and how users changed membership from one month to the next. In the analyses, the model and thresholds for each month were constrained to consist of the same number and pattern of classes discussed above. In addition, the influence of gender and length of membership as covariates in the classification process were also assessed. The transition process and the influence of gender and length of membership on the classification of players into each class will be discussed in the interpretation sections.

Interpretation of Latent Classes. Our latent class analyses revealed five classes of project-sharers on the Scratch site, which we describe in Table 3 as Low Networkers, Downloaders, Commenters, Networkers, and High Networkers. As described earlier in the model results section, one class disappeared every month, a phenomenon we explain later in interpreting our latent transition analysis. Below we describe each class as well as the changes we saw in each month based on the number and types of profiles in that month.

Table 3
Description of participation profiles
in the Scratch online community from January—March 2012.

Name	**Abbr.**	**Description**	**Months Present**

Browsers	55.5% (n=2779)	**Browsers**	B	Browses the website leaving no discernable trace to others (or in the available data)	Jan, Feb, Mar
Latent Classes – Project Sharers	44.5% (n=2225)	**Low Networkers**	LN	Creates & shares projects but does nothing else visible on the site.	Jan, Feb, Mar
		Downloaders	D	All of the above + downloads projects	Jan
		Commenters	C	All of the above + comments on projects	Jan, Feb, Mar
		Networkers	N	All of the above + some likelihood of "love-its" or "favorites" and some friending	Jan, Feb
		High Networkers	HN	All of the above + usage of "love-its," favorites, and friending as well as a higher likelihood of remixing	Jan, Feb, Mar

January. Among the five distinct latent classes identified in January (see Figure 4), the majority (43.9%) of users were classified as Low Networkers, who are unlikely to do anything except post a project during the month. Moving to a more engaged class, Downloaders (17.2% of the sample) have a 100% chance of downloading projects from the Scratch site in addition to posting a project, but exhibit almost none of the other activities. Commenters (16.1%) exhibit a strong likelihood of downloading projects, commenting on projects, and friending, with low likelihoods of favoriting or loving projects. Networkers (17.2%) are very likely to participate in downloading, commenting, favoriting, loving, and less likely friending. Finally, the High Networkers (8.4%) stand out as the Scratchers most likely to be involved in nearly every social activity on the Scratch site: they have a 55% chance of posting a remix, a 100% chance of downloading a project, and very high (above 85%) chances of commenting on or favoriting a project, and a 100% chance of loving a project and making a friend request. They stand out beyond the Networkers particularly in the areas of favoriting, remixing and friending, being twice as likely as Networkers to engage in remixing and friending.

Thus from this month, each profile appears to provides both quantitative and qualitative higher levels of participation. Low Networkers and Downloaders engage in DIY participatory activities, namely sharing projects and (for Downloaders) additionally downloading projects, Commenters take part in socially engaging actions as well through commenting. Networkers further include socially supportive actions, namely favoriting and loving projects, two activities we originally thought would be much more common across users. High Networkers stand out as more likely than Networkers to engage in all of the social and identity building activities available on the

Scratch site. They are much more likely to engage in favoriting (which has an identity building role in addition to the socially supportive role it plays) and also have the strongest likelihood of participating in all of the above activities as well as friending and remixing.

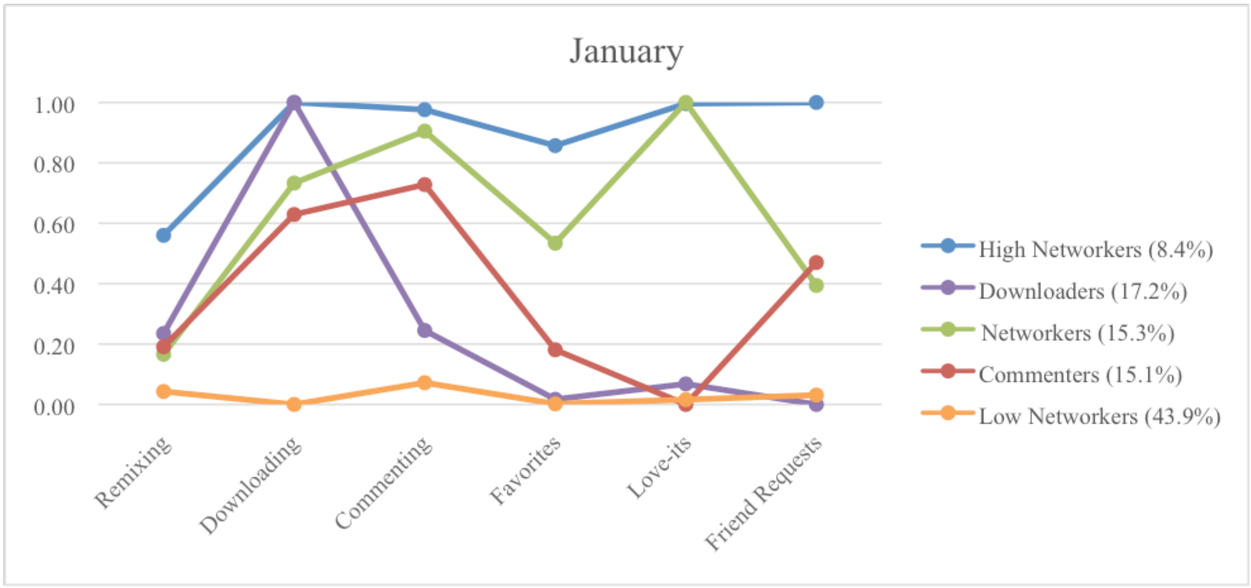

Fig. 4 Latent Class Patterns for January.

February. Among the five classes discovered in January, four of them emerged in February (see Figure 5). The Low Networker (70.1%), Commenter (12.5%), Networker (8.8%) and High Networker (8.6%) classes showed similar latent class profiles as those found in the prior month.

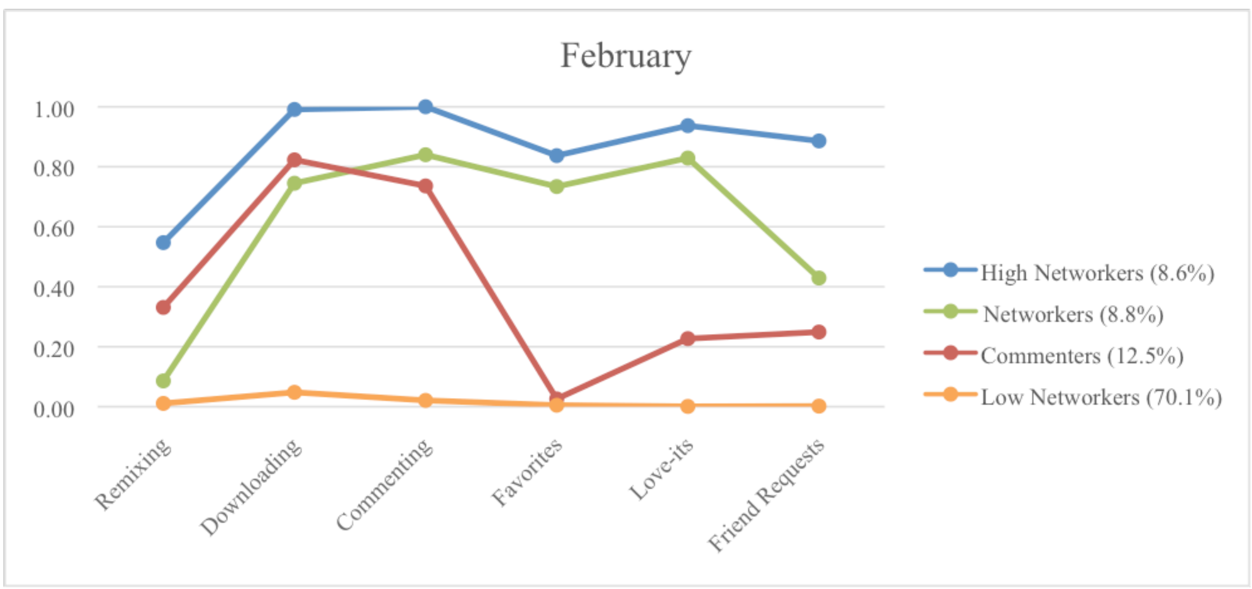

Fig. 5 Latent Class Patterns for February.

March. Among the classes discussed in the previous months, latent classes of Low Networkers (77.1%), Commenters (11.4%), and High Networkers (11.6%) also appeared in March. Although the patterns of participation (or latent class profiles) were very similar to the previous months, the likelihood of friend requests was much lower compared to the previous months for all three latent classes. Figure 6 shows the profile patterns in March.

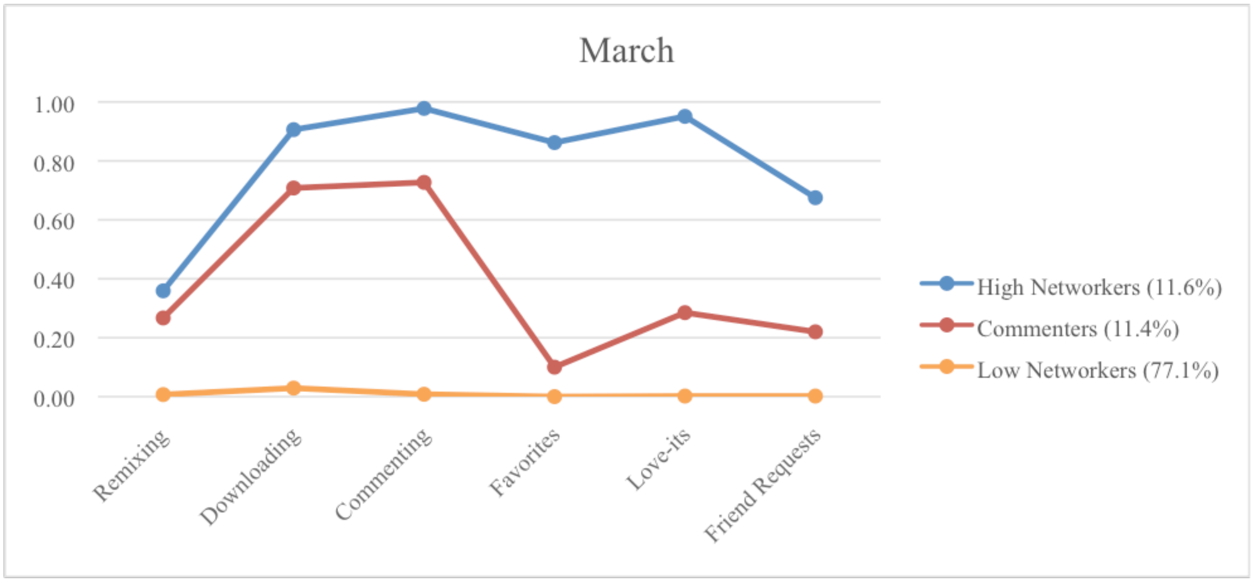

Fig. 6 Latent Class Patterns for March.

Against our expectations, there was no class of individuals who were likely to participate in socially engaging or socially supportive actions (e.g. commenting, favoriting, loving, or friending others) without also having a strong likelihood of downloading projects, an activity which suggests that Scratchers are not just sharing self-created projects but investigating and looking into them. In other words, besides posting a project, downloading a project is a second gatekeeper to social activity on the Scratch site, then commenting, and finally other types of social networking (i.e., favorites, loves, friending, and remixing). Although leaving socially supportive networking residues such as favorites, love-its, and friend requests originally seemed to us to involve the lowest bar of participation (i.e., simply clicking a button), this actually appears to be a practice in which only those who are most involved in a full range of practices on the site participate, namely the Networkers and High Networkers.

Further, we can already see changes in participation through the ways each class grows, shrinks, or disappears. The Low Networkers class grew substantially from January to February (from 43.7% to 70.1%) and grew slightly larger in March (to 77.1%). In contrast, the High Networkers class stayed constant in numbers between January (8.4%) and February (8.6%) and increased a little in March (11.6%). All of the other classes (i.e., Downloaders, Commenters, and Networkers) slowly shrink or disappear entirely. On one level this may suggest that those users who do not start engaging in socially engaging or socially supportive practices (i.e., loving, favoriting, and friending) may not stay as engaged on Scratch, highlighting the importance of social activities for website activity, even on a site focused on project creation. To further understand these phenomena we look toward gender and Scratch membership to see if those hold hints about changing patterns of participation.

5.2.2. ***Interpretation of the Latent Transitions.*** We found that participation online shifted dramatically over the three-month time period of the study. Tables 4 and 5 show the probabilities of players classified in one class transitioning to another class the following month; Figure 7 illustrates these same patterns of change. In general, Scratchers who were not engaged in any activity (Low Networkers) were likely to stay in that class across the months. This transition was less dramatic from January to February as a quarter (27%) of Low Networkers transitioned into Commenters, Networkers, and even involved High Networkers. From February to March, only 13.5% of Low Networkers evolved to more advance players (as either Networkers, or High Networkers). A similar shift in participation also appeared for High Networkers. That is, there was a strong likelihood that High Networkers stayed as High Networkers from month-to-month.

The majority of Commenters and Networkers followed traditional website trends shifting to lower and lower engagement (as Low Networkers) or continuing similar participation practices across time with very fewer players moving upward in their participation. It is also interesting to note that Downloaders, a pattern of participation that ceased to exist after January, became less active as Low Networkers (89.3%) with a few members shifting to the practices of Commenters. Moving to the last month, although Commenters disappeared as a class, a large number of these users (34.4%) showed the most promise in their play and turned toward participation as Networkers (downloading, commenting, and otherwise networking) or High Networkers (engaging in all aspects of Scratch).

Table 4

Likelihoods of members of one profile transitioning to another profile.

January \ February	**High Networkers** (8.67%)	**Commenters** (12.6%)	**Networkers** (8.79%)	**Low Networkers** (69.9%)
High Networkers (8.34%)	65.4%	2.5%	11.7%	20.4%
Networkers (15.6%)	8.0%	11.5%	16.1%	64.4%
Commenters (15.3%)	6.9%	16.5%	4.5%	72.1%
Downloaders (16.9%)	0.0%	9.2%	1.4%	89.3%
Low Networkers (43.7%)	1.9%	14.8%	10.3%	73.0%

Table 5

Likelihoods of members of one profile transitioning to another profile.

F \ March	**High Networkers** (11.9%)	**Networkers** (12.1%)	**Low Networkers** (75.9%)
High Networkers (8.67%)	74.7%	10.5%	14.8%

Networkers (8.79%)	21.6%	10.4%	68.0%
Commenters (12.6%)	5.4%	29.0%	65.6%
Low Networkers (69.9%)	4.1%	9.4%	86.5%

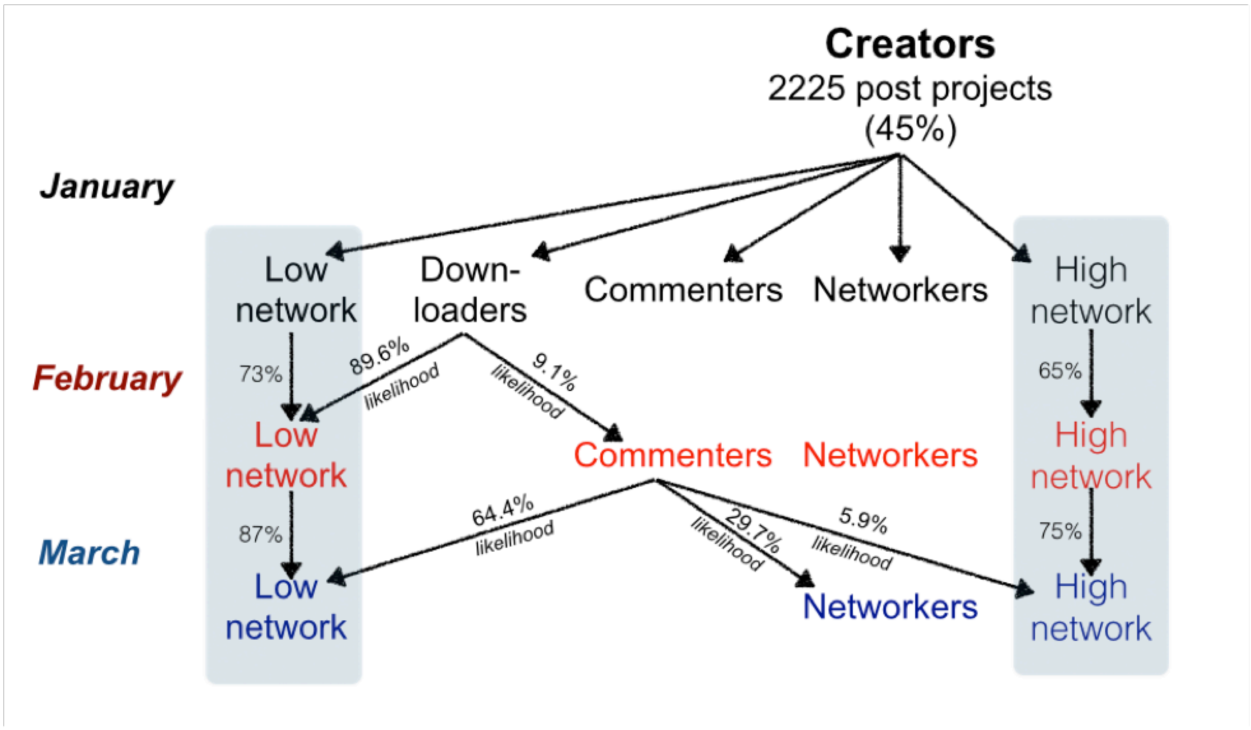

Fig. 7 Visualizing some of the transitions to "low network" participation and the tendency of high networkers to stay as high networkers.

Gender and Membership in Scratch Community

We now turn to two additional features of members in Scratch online community, gender and length of membership. Two sets of further analyses were conducted to examine the influence of these variables on latent class membership. Under the umbrella of LTA, we first utilized multinomial logistic regression to test whether gender and length of membership influenced membership at each time point. Second, we examined the distribution of gender and membership groups across the latent classes (based on the highest probability classification) through the use of chi-square tests of independence.

Multinomial logistic regression. Using results from our latent transition analyses, multinomial logistic regression analyses were performed to predict latent class membership at each time point using gender and length of membership as predictors. Prior to these analyses, the Low Networker class, the largest class for most months, was selected as a reference group; thus, the outcome variable is a dichotomous variable with the higher value indicating membership into a specific class (e.g., High Networker class) relative to the Lower Networker class. Table 6 presents the influence of gender and length of membership as regression coefficients (and odds ratio) predicting class membership. For January classes, results show that increased length of membership on Scratch was a significant predictor of membership into the High Networker class (relative to the Low Networker class). However, this length of membership significantly decreased the likelihood of membership into the Commenters and Downloaders class in favor of the Low Networker class. Thus, at the initial month, there is a duality in terms of membership length: for some, the longer they stayed on Scratch, the more likely they will be highly involved (as High Networkers); for others, lengthier membership status encouraged inactivity in relation to being Commenters or Downloaders. For February, the length of membership functioned as a significant predictor of increased classification as Commenters and Networkers (relative to the Low Networkers), suggesting social involvement becoming key to retention. Similarly, in March, length of membership played a significant role in predicting increases in membership in both the active classes of Commenters and High Networkers. No other significant results were found for length of membership.

The story of gender as a predictor of latent classes suggests equity in membership. In general, gender played a marginal role in terms of how users participated in Scratch. In January, results showed that girls were significantly more likely than boys to be in the High Networker class (rather than the Low Network class). In February and March, girls were significantly less likely than boys to be in the Commenters class. Gender was not found to significantly predict membership for the other classes. This lack of significant differences suggest that patterns of participations are less dependent on gender and more dependent on enduring membership.

Table 6

Multinomial logistic regression [coefficient (odds ratio)] result predicting latent class membership based on gender and length of membership

	High Networkers	**Commenters**	**Networkers**	**Downloaders**
January				
Length of Membership	0.37*** (1.44)	-0.16* (.85)	0.02 (1.02)	-0.34*** (.71)
Gender (female)	0.51** (1.66)	0.15 (1.16)	0.22 (1.25)	-0.01 (.99)
February				
Length of Membership	0.09 (1.09)	0.27*** (1.31)	0.51*** (1.67)	
Gender (female)	0.01 (1.01)	-0.39* (.67)	-0.10 (.90)	
March				
Length of Membership	0.59*** (1.81)	0.60*** (1.82)		
Gender (female)	0.19 (1.21)	-0.38* (.69)		

Note: *p < .05, **p<.01, ***p<.001. For each comparison, Low Networkers serve as the reference group.

Tests for gender and membership. To examine the distribution of gender and length of membership across the latent classes for each month in greater detail, chi-square tests of independence were conducted. Prior to these analyses, all users were categorized into their highest probability class for each month.

Gender. In general, gender played a marginal role in class membership. Although the chi-square test of independence for January revealed a significant relationship between gender and latent class memberships [χ^2 (4) = 9.635, p = .047], an examination of the standardized residuals revealed only one marginally significant finding: a higher proportion of girls who were categorized as High Networkers in January than expected (z = 2.030). The results for February did not yield a significant relationship, χ^2 (2) = 5.613, p = .132. For March, there was a lower proportion of girls in the Commenters than expected (z =-2.132), χ^2 (2) = 10.040, p = .007.

These analyses suggest that while males dominate the population of Scratch at large, within participation profiles gender differences are minimal, a remarkable finding for a youth amateur design site focused on programming. Notably, our own prior ideas about the Scratch online community suggested that girls dominated comments by sheer numbers while boys dominated projects, a pattern easily visible in simple frequency data on comments, projects, and gender (see scratch.mit.edu/research). However, by looking at participation patterns our analysis opens up an alternative look at these trends. From this perspective, all active users of the site are project creators, and amongst those are three groups of individuals who engage in commenting and other social networking activities (e.g., Commenters, Networkers, and High Networkers). There are almost no gender differences amongst these classes of users, and certainly no gender differences that hold over time.

Length of membership. The length of time users had accounts on Scratch.mit.edu (i.e. their Scratch membership) was most certainly related to which participation classes they were in, especially for users who created new accounts in January 2012 (i.e., “newbies” who joined the month our data collection began) and for more senior Scratch users. Tables 7, 8 and 9 show the chi-square results and the proportional distribution of users by membership over each month. Overall, we see several interesting trends. First, while the percentages of Scratchers that were Low Networkers increased each month from January to March, this trend was much stronger for those whose accounts were new in January (newbies) or less than three months old. More senior Scratch users were far more likely to be in a more involved class of participation than the younger users. Second, in January there were far more newbies in the Downloaders class than expected. This may be an explanation for why this class disappeared between January and February. Overall, each month the newest users, those who joined Scratch at the start of the study, are less likely to be represented in the more involved participation classes (i.e., Commenters, Networkers, and High Networkers). This shows a fairly typical form of online engagement seen in many other sites where new users join, engage in the website, then shift to lower participation or disappear altogether from the site (see Kafai & Fields, 2013).

Table 7

Distribution of January latent class members by by the age of their Scratch accounts[1], χ^2 (12) = 118.82, p < .001

January					
	High Networkers (8.40%)	**Downloaders (16.9%)**	**Commenters (15.4%)**	**Networkers (15.6%)**	**Low Networkers (43.7%)**
Newbie (new account) N = 740	**3.4%--**	**26.6%++**	**13.4%+**	14.9%	41.8%
Young (0-3 months) N = 625	7.0%	20.2%	9.9%	14.4%	48.5%
One-Year (4-12 months) N = 409	**15.4%++**	**11.7%--**	9.3%	18.3%	45.2%
Oldie (12+ months) N = 403	**14.1%++**	**12.4%--**	9.2%	14.4%	49.9%

Table 8
Distribution of February latent class members
by the age of their Scratch accounts, χ^2 (9) = 108.94, p <.001.

February				
	High Networkers (8.67%)	**Commenters (12.6%)**	**Networkers (8.79%)**	**Low Networkers (69.9%)**
Newbie (new account) N = 740	**4.5%--**	**8.1%-**	**5.5%-**	**81.9%++**
Young (0-3 months) N = 625	8.0%	11.2%	7.5%	73.3%
One-Year (4-12 months) N = 409	**17.6%++**	13.0%	**11.2%+**	**58.2%--**
Oldie (12+ months) N = 403	11.7%	**15.4%+**	**11.2%+**	**61.8%-**

Table 9. Distribution of March latent class memberships
by the age of their Scratch accounts, χ^2 (6) = 178.22, p < .001.

March			
	High Networkers (11.9%)	**Networkers (12.1%)**	**Low Networkers (75.9%)**
Newbie (new account) N = 740	**5.3%--**	**6.1%--**	**88.6%++**

[1] Notation: +z > 2.0, ++ z > 3.0. - z < -2.0), -- z < -3.0; these notations indicated whether there are significantly more (z > 2.) or fewer (z < -2.0) players in the cell than excepted

Young (0-3 months) N = 625	10.4%	**8.6%**	81.0%
One-Year (4-12 months) N = 409	**19.6%++**	**18.3%++**	**62.1%--**
Oldie (12+ months) N = 403	**17.6%++**	**22.3%--**	**60.0%--**

Parallel to these trends amongst junior Scratchers, more senior users (all those whose Scratch accounts > 3 months old) were more likely to be involved in all aspects of the Scratch site overall. For instance, the top two categories of senior users were more likely to be in the High Networkers class each month (excepting senior users in the month of February, see Table 8) as well as the Networkers class in February and March. Similarly they were markedly less likely to be in the Downloaders class in January. After January they were also considerably less likely to be in the Low Networkers class. Thus we would expect that users who continue to post projects on Scratch after at least a few months would be more likely to participate in more aspects of the Scratch community over time.

Discussion

This chapter examined broad qualitative and quantitative trends of social practices that shape participation in a youth do-it-yourself (DIY) social networking forum focused on the production of programming projects. While revealing visible distinct types of users that define participation on a massive scale, our findings also call into question some earlier views about participation on the website. In the following sections we discuss our new insights on enduring participation in the Scratch community, consider implications for equity, discuss the relationship of programming and participation, outline considerations for designing for collaborative learning on a massive scale, and propose directions for future research.

Project-Focused Participation: DIY Social Networking Forums

Perhaps most surprisingly, our findings suggest that the key forms of participation on the Scratch site are sharing and downloading content, activities that reflect that Scratch is most predominantly a DIY community. Remarkably, nearly 45% of Scratch users posted projects, a tremendously high level of user contribution in a massive online community. We suggest that this denotes a very different form of basic participation than more well-known patterns in traditionally thought of social network sites visited by far larger numbers of users (e.g., Facebook, Vine, MySpace) where users commonly post happenings and events in their daily lives. Instead, we suggest that DIY social networking forums may have their own unique patterns of participation where sharing one's own content is the baseline of participation rather than more socially engaging or socially supportive actions. One reason for this may be that in DIY social networking forums, sharing self-created content involves not just adding content to a site, but is the most core form of identity display in those online communities. It is all too easy to differentiate users as "project-creators" or "socializers." Rather, in Scratch at least, all active users who left any traces of their participation were project-sharers, and that this project sharing is both a participatory activity (sharing a creation with other users) as well as an identity building activity (where projects reflect who one is on the site). Amongst those project-sharers, users engaged in different types of social activity that differentiated their types of participation with high networkers being the most stable class of users.

Interestingly, the seemingly easy socially supportive actions (simple networking residues like loving and favoriting) were only evident amongst the most involved users: Networkers and High Networkers. Of course, these users engaged in the entire spectrum of social practices we identified: from sharing projects to commenting, loving, favoriting, friending, and remixing, truly forming the 'core' group of Scratch users. Further several of the above actions may play another role beyond social interaction through direct display on users' profile pages. For instance, on the Scratch site, *sharing* and *favoriting* projects holds far more prominence on users' personal pages than the thumbnail picture and city/country information on a user's profile (see Figure 2). These activities have the added layer of identity building on the site in that they represent a user's abilities, interests, and preferences. This puts a different lens on sharing projects as a basic form of participation in DIY social networking forums. Not only is it a type of content creation but it is also the primary way of establishing a presence in the online community. Favoriting projects, an activity engaged in the most by the High Networkers, also holds identity building meaning on the site. Seeing these DIY social networking activities in light of establishing an identity may provide another layer of interpretation as to their importance.

While large numbers of participants in sites with millions of registered users result in overall high activity, it is in fact often the smallest group of users that drives the most activities and attains the most visibility (i.e., Kafai & Fields, 2013). In other words, while everyone has access to the site, not everyone is as highly engaged or contributes in the same manner in informal online communities. What does this mean? For one, it means that those users most likely to draw the attention of designers and researchers are a relatively small group. Researchers who focus on case studies or ethnographies as well as designers who respond to users' posts and concerns are dependent on users who engage in commenting or other forms of written communication. If other massive DIY social networking forums follow the trends of Scratch.mit.edu, then those who leave comments may actually be a small minority of the overall population. Those who stay socially engaged month-by-month are an even smaller minority. The celebration of rich opportunities for learning in studies of affinity spaces, gaming communities, and social networking sites may thus only apply to a small proportion of users on a site. This observation indicates that actual collaborations in massive online communities are limited to a far smaller number of users than the overall size of the community seems to suggest.

Participation vs. Programming

Overall, we found an encouraging lack of gender differences amongst classes of users in Scratch.mit.edu based on engagement in social practices online. Given that programming communities are heavily male-dominated (even Scratch is 67% male), the fact that girls are proportionately part of all participation classes is remarkable. However, interesting questions about gender and participation arise when we compare classes of participants to classes of programmers. In our related study on the same sample of users we analyzed classes of programmers, finding four stable classes based on the relative sophistication of programming commands used in Scratch projects (see Fields, Giang & Kafai, 2014). In this case, gender differences appeared in the highest and lowest classes of programmers: girls were much more likely to be in the largest, most novice programming class (for example, not moving beyond loops in their programming) and much less likely to be in the "Advanced" and "Experienced" classes of programmers that used many different types of more challenging commands at relatively high levels of frequency (e.g., Booleans, variables, conditionals)[2]. This finding raises interesting questions about the differential appearance of gender differences: while there is a gender difference in programming at a high programming profile there is essentially no difference with regard to gender in any of the participation profiles.

Further, when we compared participation classes to programming classes from the same sample we found no relationships *except* between the "High Networker" class of participation and the highest, "Experienced" class of programmers which strongly overlapped (Fields, Giang, & Kafai, 2014). In the tech community, there has been a strong push to involve women in the socialization of computer science, assuming that such socialization will result in more involved and higher levels of coding. Yet these results indicate that we need better understandings of how social engagement may or may not relate to depth of programming engagement. Beyond programming, this raises questions about the relationship between deep participation and deep expertise in any given domain of design in online communities (e.g., writing, drawing, video making, etc.). While Ito and colleagues' work (2010) suggests a trajectory of participation from hanging out to messing around to geeking out, we found that these social networking activities, even at high levels, may not directly result in moving into "geeking out," at least at the higher levels of more sophisticated programming. Although case studies of successful Scratch users (Brennan, 2013) share the stories of members who managed to transition into more extensive programming, more research is needed to understand to what extent these transitions happen on a larger scale, for which classes of users, and over what kinds of time frames.

Designing for Online Participation

The larger goal of this research is to illuminate participation practices in massive communities that support learning and design, to see who is participating and collaborating in those activities, and to evaluate how to sustain those types of activities. One area that our findings contribute to is the affordances of different social networking features in online communities. Our findings suggest that sharing self-created projects may be a strong entry point for participating in online amateur design sites. Designers of DIY social networking forums, an up and coming genre of website for children and youth (Grimes & Fields, 2015) should note the key role of sharing one's creations in participation. Many, many websites that promote or provide tools for making things do not actually support sharing, yet this designed-for ability may be a key feature of promoting social engagement in interest-driven communities that support user design.

At the same time, users may need assistance in developing "participatory competencies" (Kafai & Burke, 2014) with more conversational types of networking residues such as comments, "likes," favorites, and friending/following. In our analysis it was highly unlikely for new users to engage in these features (i.e., to be in the Commenter, Networking, or High Networking classes). Interestingly, those users who engaged in the full range of networking features were highly likely

---

[2] Just because a user only uses loops and does not use seemingly more complex commands does not necessarily mean that their programs are less sophisticated. However, our analysis, detailed in Fields, Giang, and Kafai (2014), supports a view of increasing eliteness in programming based on the latent classes of programmers we identified. See Fields, Giang, and Kafai (2014) for a fuller discussion of this topic.

to stay engaged in the long-term. However, this does not necessarily mean that engaging in "loving" and "favoriting" will result in more enduring participation (i.e. a causal interpretation)—it is simply a part of the activities of the most engaged users. Our current analysis does not allow for a clear-cut interpretation of this finding but will require more research. Case study and ethnographic analyses of the Scratch community provide similar insights into the more social features of the site and how important developing relationships or a sense of community is to participation. Users who engage in commenting and who receive constructive and positive comments, tend to credit their engagement to those socially engaging activities (see Brennan, Valverde, Prempeh, Roque, & Chung, 2011). Our research of another online community, the virtual world Whyville.net, also highlighted the importance of reciprocal social engagement (i.e. conversing and hanging out) in the most involved (top 7%) users (Giang, Kafai, Fields & Searle, 2012). Yet users often may not know how to begin to comment appropriately on projects or even how to reply to others' comments, how to find collaborators, how to get feedback on their projects, and how to become a part of the community. Our research in local Scratch workshops confirms that many youth may feel disconnected from or even intimidated by larger online communities (Kafai, Fields & Burke, 2010; Fields, Vasudevan, & Kafai, in press). Further, another challenge is online communities are users who engage in discouraging behaviors, leaving insulting comments, copying projects without giving credit, pressuring others to be similar rather than creative (Brennan, 2011).

One effort we have made to support users' participatory and programming competencies has been to hold special "Collab Camps" where users are invited to program a special themed project in a small group (2+ users) (see Roque & Resnick, this volume??). From 2012-2013 we ran a series of three Collab Camps that utilized a specific timeline where groups (or collabs) had to post a draft of their project by a specific time, receive constructive criticism from the Scratch Team and trained Scratchers (Collab Counselors), then post a final version 2-3 weeks later. This successfully supported project revisions and deepening of programming and media skills (Fields, Kafai, Strommer, Seiner & Wolf, 2014) and an increase in constructive criticism left by participants on each others' projects (Roque, Kafai & Fields, 2012). We also implemented Collab Camps locally with novice Scratch students. In our third Collab Camp we integrated design features from the online challenge in a local workshop, training high school students to provide each other with positive, constructive feedback and providing transparency into each others' projects. Students cited these efforts as enabling them to improve their projects and identify more strongly with computing (Fields, Vasudevan, & Kafai, in press). Interestingly, the local users we engaged found the local audience of their peers the most meaningful; they were generally not interested in participating further on the website, though they valued the feedback and audience of the broader community. These provide but a few examples of the potential for helping local and online users build participatory competencies in DIY social networking forums and for utilizing design strategies implemented online in face-to-face and hybrid settings.

Directions for Future Research

We are only at the initial phase of understanding learners in amateur online design communities (or DIY social networking forums), especially youth programming communities. The type of broad scale research we conducted is useful for noting widespread trends not easily visible from more qualitative analyses, enabling us to put findings from case studies and ethnographies into a larger perspective. At the same time, by itself it has clear limitations in the depth of what it can say about users within the community and within each identified class. Other studies of online communities, namely gaming and social network communities (Boelstorff, Nardi, Pearce & Taylor, 2013; Hine, 2000; Williams, Yee, & Caplan, 2008) reveal an unhealthy split in either quantitative or qualitative research approaches. For instance, survey methods and statistical data mining seem to drive many efforts in coming to grips what engages members in these massive online communities. On the other end, we have ethnographies of single massive communities (see Boelstorff, 2008; Taylor, 2006) that inform us with a fine-grained detail of cultural practices and activities. Of course, others have rejected this dichotomy and argued for a mixed methods approach (Williams, 2005), but it is difficult to bring together the diverse expertise and resources (much less permission for backend data from websites) needed to accomplish both thick and broad analyses. In our view it is not just about juxtaposing data sources and analytical methods but also about developing perspectives that integrate both approaches in a productive manner. As a case in point, we have suggested and employed connected ethnographies that make use of the data mining and reduction in large data sets to identify particular participants based on their contribution profiles and to cross reference and develop these through in-depth ethnographies (Kafai & Fields, 2013; see also, Reimann, 2009). Such analyses leverage the explanatory potential of each method and allow us to contextualize cases within larger community trends.

Finally, the participation and patterns from the Scratch website do not generalize easily to other communities. Rarely, the Scratch website was created and developed in a university environment, one with a particular ethos of openness expressed through an open source computing tool (Scratch) as well as broad openness on the website (all comments and shared projects are fully public). Indeed, the breadth of networking features on Scratch is relatively rare when compared to other DIY social networking forums for kids (Grimes & Fields, 2015). Thus, in addition to rich, mixed method research into individual sites we also need research that systematically compares the designs and participation of multiple types of sites. Studying and supporting collaborative forms of learning in massive online communities is not simply a matter of involving larger numbers of participants but also of considering the nature of activities; the various roles of participants, educators, and designers; and the creation, sharing, and socializing around artifacts.

References

Ares, N. (2008). Cultural practices in networked classroom learning environments. *International Journal of Computer Supported Collaborative Learning, 3*(3), 301-326.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom.* New Haven and London: Yale University Press.

Black, R.W. (2008). *Adolescents and Online Fan Fiction*. New York: Peter Lang.

Boellstorff, T. (2008). *Coming of age in second life: An anthropologist explores the virtually human.* Princeton, NJ: Princeton University Press.

Boellstorff, T., Nardi, B., Pearce, C. & Taylor, TL (2013) (Eds.). *A Handbook of Ethnographic Methods for Virtual Worlds*. Princeton, NJ: Princeton University Press.

boyd, d. (2013). *It's complicated*. New Haven, CT: Yale University Press.

boyd, d., & Ellison, N. B. (2007). Social network sites: Definition, history and scholarship. *Journal of Computer-Mediated Communication*, 13.

Brennan, K., Valverde, A., Prempeh, J., Roque, R., & Chung, M. (2011). More than code: The significance of social interactions in young people's development as interactive media creators. *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2011(1), pp. 2147-2156.

Brennan, K. A. (2013). *Best of both worlds: Issues of structure and agency in computational creation, in and out of school* (Doctoral dissertation, Massachusetts Institute of Technology).
Brennan, K. (2011). Mind the gap: Differences between the aspirational and the actual in an online community of learners. In *Proceedings of the International Conference of Computer Supported Collaborative Learning*.
Chen, M. (2012). *Leet noobs: The life and death of an expert player group in World of Warcraft*. New York: Peter Lang.
Collins, L. M. & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences.* John Wiley & Sons: Hoboken, NJ.
Eddy, Y.C.L., Chan, C., & van Aalst, J. (2006). Students assessing their own collaborative knowledge building. *International Journal of Computer Supported Collaborative Learning, 1*(1), 57-87.
Fields, D. A., Giang, M. T., Kafai, Y. B. (2014). Programming in the wild: Patterns of computational participation in the Scratch online social networking forum. In *Proceedings of the 9th Workshop in Primary and Secondary Computing Education* (WiPSCE '14). ACM, New York, NY, USA, 2-11. http://doi.acm.org/10.1145/2670757.2670768
Fields, D. A., Giang, M. & Kafai, Y. B. (2013). Understanding collaborative practices in the Scratch online community: Patterns of participation among youth designers. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds), *To see the world and a grain of sand: Learning across levels of space, time, and scale: CSCL 2013 Conference Proceedings, Volume 1, Full Papers & Symposia*. International Society of the Learning Sciences: Madison, WI, 200-207.
Fields, D. A. & Kafai, Y. B. (2010). Knowing and throwing mudballs, hearts, pies, and flowers: A connective ethnography of gaming practices. *Games and Culture,* (Special Issue), 5(1), 88-115.
Fields, D. A. & Kafai, Y. B. (2009). A connective ethnography of peer knowledge sharing and diffusion in a tween virtual world. *International Journal of Computer Supported Collaborative Learning,* 4(1), 47-68.
Fields, D. A., Kafai, Y. B., Strommer, A., Wolf, E. & Seiner, B. (2014). Interactive storytelling for promoting creative expression in media and coding in youth online collaboratives in Scratch. In *Proceedings of Constructionism*, Vienna, Austria. http://constructionism2014.ifs.tuwien.ac.at/papers/3.2_2-8540.pdf
Fields, D. A., Pantic, K. Kafai, Y. B. (2015). "I have a tutorial for this": The language of online peer support in the Scratch programming community. In *Proceedings of Interaction Design and Children*, (IDC '15). ACM, New York, NY, USA.
Fields, D. A., Vasudevan, V., Kafai, Y. B. (in press). The programmers' collective: connecting collaboration and computation in a high school Scratch mashup coding workshop. *Interactive Learning Environments*.
Gee, J. (2003). *What videogames have to teach us about learning and literacy*. New York: Palgrave.
Giang, M. T., Kafai, Y. B., Fields, D. A., & Searle, K. A. (2012). Social interactions in virtual worlds: Patterns and participation of tween relationship play. In J.Fromme &

A. Unger (Eds.), *Computer games and new media cultures: A handbook on the state and perspectives of digital games studies.* New York: Springer Verlag. 543-556.

Grimes. S. M & Fields, D. A. (2015). Children's media making, but not sharing: The potential and limitations of child-specific DIY media websites for a more inclusive media landscape. *Media International Australia,* 154, 112-122.

Grimes, S. & Fields, D. (2012). *Kids online: A new research agenda for understanding social networking forums.* New York. The Joan Ganz Cooney Center at Sesame Workshop. Available online at http://www.joanganzcooneycenter.org/reports-38.html.

Hagenaars J. & McCutcheon A. (eds). (2002). *Applied Latent Class Analysis*. Cambridge, UK: Cambridge University Press.

Hine, C. (2000). *Virtual ethnography*. London: Sage Publications.

Ito, M., Baumer, S., Bittanti, M., boyd, d., Cody, R., Herr, B., Horst, H. A., Lange, P. G., Mahendran, D., Martinez, K., Pascoe, C. J., Perkel, D., Robinson, L., Sims, C., & Tripp, L. (2009). *Hanging out, messing around, geeking out: Living and learning with new media.* Cambridge, MA: MIT Press.

Kafai, Y. B. & Burke, Q. (2014). Connected Code: Why Children Need to Learn Programming. Cambridge, MA: MIT Press.

Kafai. Y. B. & Fields, D. A. (2013). *Connected Play: Tweens in a Virtual World*. Cambridge, MA: MIT Press.

Kafai, Y.B., Fields, D.A. & Burke, W.Q. (2010). Entering the clubhouse: Case studies of young programmers joining Scratch community. *Journal of Organizational and End User Computing* 22(2), 21-35.

Kafai, Y. B., Heeter, C., Denner, J., & Sun, J. Y. (2008). Beyond Barbie [R] and Mortal Kombat: New Perspectives on Gender and Gaming. *MIT Press (BK)*.

Lammers, J. C., Curwood, J. S., & Magnifico, A. M. (2012). Toward an affinity space methodology: Considerations for literacy research. *English teaching: Practice and critique*, *11*(2), 44-58.

Luther, K., Caine, K. Ziegler, K. & Bruckman, A. (2010) "Why It Works (When It Works): Success Factors in Online Creative Collaboration." In *GROUP '10: Proceedings of the ACM Conference on Supporting Group Work*, New York: ACM Press, 1-10.

Magnifico, A.M. (2010) Writing for shom? Cognition, motivation, and a writer's audience. *Educational Psychologist* 45(3), 167-84.

Maloney, J., Peppler, K., Kafai,Y., Resnick, M., & Rusk, N. (2008, March). *Programming by choice: Urban youth learning programming with Scratch*. Paper presented at the SIGCSE 2008 Conference, Portland, OR.

Margolis, Jane, Estrella, R., Goode, J., Holme, J. J., and Nao, K. (2008). *Stuck in the Shallow End: Education, Race, and Computing*. The MIT Press.

Margolis, Jane, and Allan Fisher. 2001. *Unlocking the Clubhouse*. MIT Press.

Monroy-Hernandez, A. & Resnick, M. (2008). Empowering kids to create and share programmable media. *Interactions*, *15*(2), 50-53.

Muthen B. (2002). Statistical and substantive checking in growth mixture modeling. Retrieved January 2007 from http://www.gseis.ucla.edu/faculty/muthen/full_paper_list.htm.

Nickerson, J.V., & Monroy-Hernández, A. (2011). Appropriation and Creativity: User-initiated Contests in Scratch. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS '11).*

O'Donnell, A. M. (2006). The role of peers and group learning. In P. Alexander & P. Winne (Eds.), Handbook of educational psychology, 2nd Edition. Mahwah, NJ: Lawrence Erlbaum.

Reimann, P. (2009). Time is precious: Variable- and event-centered approaches to process analysis in CSCL research. *International Journal of Computer Supported Collaborative Learning, 4*(3), pp. 239-257.

Resnick, M., Maloney, J., Hernández, A. M., Rusk, N., Eastmond, E., Brennan, K., Millner, A. D., Rosenbaum, E., Silver, J., Silverman, B., & Kafai, Y. B. (2009). Scratch: Programming for everyone. *Communications of the ACM, 52(11),* 60–67.

Rick, J. & Guzdial, M. (2006). Situating CoWeb: A scholarship of application. *International Journal of Computer-Supported Collaborative Learning*, 1(1), 89-115.

Roque & Resnick, (this volume). Designing supports for creative collaboration

Roque, R., Kafai, Y. B., & Fields, D. A. (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, 220-223.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Eds.), Liberal education in a knowledge society (pp. 67–98). Chicago: Open Court.

Scardamalia, M., & Bereiter, C. (1991). Higher Levels of Agency for Children in Knowledge Building: A Challenge for the Design of New Knowledge Media. *Journal of the Learning Sciences, 1*(1), 37-68.

Steinkuehler, S., & Duncan, S. (2009). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology, 17(6),* 530–543.

Taylor, T. L. (2006). *Play between worlds*. Cambridge: MIT Press.

van Aalst, J. (2009) Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer Supported Collaborative Learning, 4*(3), pp. 259-287

van der Pol, J., Admiraal, W. & Simons, P. R. J. (2006). The affordance of anchored discussion for the collaborative processing of academic texts. *International Journal of Computer-Supported Collaborative Learning*, 1(3), 339-357.

Webb, N. & Palincsar, A. (1996). Collaborative learning. In D. Berliner (Ed.), *Handbook of Educational Psychology* (pp. 345-413), New York: Macmillan.

Williams, D. (2005). Bridging the methodological divide in game research. *Simulation & Gaming, 36*(4), 1-17.

Williams, D., Contractor, N., Poolec, M. S., Srivastad, J., & Cale, D. (2011). The Virtual Worlds Exploratorium: Using Large-Scale Data and Computational Techniques for Communication Research. *Communication Methods and Measures*, 5(2), 163-180.

Williams, D., Yee, N. & Caplan, S. (2008). Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer Mediated Communication*, *13*(4), 993-1028.

Yee, N. (2014). The Proteus Paradox: How Online Games and Virtual Worlds Change Us-And How They Don't. Yale University Press.

Zhang, J., M. Scardamalia, R. Reeve, and R. Messina, R. Designs for Collective Cognitive Responsibility in Knowledge Building Communities. *Journal of the Learning Sciences* 18, no. 1 (2009): 7–44.

8. ACKNOWLEDGMENTS

This material is based upon work supported by a collaborative grant from the National Science Foundation (NSF#1027736) to Yasmin Kafai with Mitchel Resnick, and Yochai Benkler. The views expressed are those of the authors and do not necessarily represent the views of the National Science Foundation, Utah State University, St. Mary's College, or the University of Pennsylvania. Particular thanks to Anant Seethalakshmi for help with gathering data and to the Scratch Team for providing repeated feedback on analysis. An earlier version of the latent class analysis appeared in the proceedings of the Annual Conference of Computer Supported Collaborative Learning (Fields, Giang & Kafai, 2013).