

# Youth Perspectives on Critical Data Literacies

Samantha Hautea\*

Sayamindu Dasgupta\*†

Benjamin Mako Hill\*

\*University of Washington  
Seattle, WA, 98195  
{smhautea,sdg1,makohill}@uw.edu

†MIT Media Lab  
Cambridge, MA, 02142  
sayamindu@media.mit.edu

## ABSTRACT

As contemporary youth learn, play, and socialize online, their activities are often being recorded and analyzed. What should young people know about these data collection and analysis efforts? Although critiques of these new forms of data collection and analysis have grown increasingly loud, the voices of users, and particularly youth, have largely been absent. This paper explores the critical perspectives of youth who are programming with public data about their own learning and social interaction in the Scratch online community. Using a bottom-up approach based on ethnographic observation of discussions among these young users, we identify a series of themes in how these youth critique, question, and debate the implications of data analytics. We connect these themes—framed in terms of critical data literacies—to expert critiques and discuss the implications of these findings for education and design.

## Author Keywords

critical data literacy; data literacy; youth; learning; big data; data science

## ACM Classification Keywords

K.3.2 Computers and Education: Computer and Information Science Education; H.5.m Information Interfaces and Presentation (e.g., HCI): Miscellaneous

## INTRODUCTION

As technology’s ability to save and store information has increased, the creation and analysis of large-scale datasets has grown. In 2013, a study estimated that more than 90 percent of the data in the world had been produced over the last ten years, and that the world was generating 2.5 quintillion bytes of data daily. Moreover, this number is rapidly increasing [16].

The collection, management, and analysis of these rapidly expanding datasets are often described as issues of “big data” [28, 21] and “data science” [53, 49]. Although the sources of these new large datasets vary, many of the most popular accounts of the growing importance of data have focused on “digital footprints” [35, 26], “digital traces” [32], or “digital exhaust” [38]: data created as a side effect of human interaction with computing systems [19]. Quantitative analysis of these

social and behavioral data is used to uncover hidden patterns and correlations, which can then be used to inform design, decision making, and policy setting [1, 47].

Some of the excitement about data is directed at the potential to collect and analyze information about the activities of youth. As youth flock to online systems to play, learn, and socialize, their activities are measured, recorded, quantified, tracked, and analyzed in unprecedented ways.<sup>1</sup> In the field of digital learning systems, there is broad interest in new possibilities for digital data about youth, including using data for “personalizing instruction, improving student learning, improving student engagement, and providing rapid feedback to learners” [17, 2].

Recently, several scholars have begun to advance a critical discourse around data and its limitations. In particular, these critiques have emphasized the challenges associated with drawing valid conclusions from online behavioral data and the need for researchers and educators to reflect on the ethics of how data is acquired, interpreted, and utilized [5, 33, 34, 41]. These critiques suggest that the contemporary public discourse about data has an element of mythology embedded in it, causing “widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy” [5].

Within these critical conversations, the voices and accounts of end users who are the subject of data collection are rare, with several notable exceptions (e.g. [9, 7, 44]). As individuals at the leading edge of new data collection efforts, understanding the implications of new forms of digital data collection and analysis is particularly important for youth. What should young people know about the data being collected about them and about the attempts to analyze and understand these data in ways that can shape their experience? What are the skills they require to approach the subject of data critically?

In this paper, we present a bottom-up approach that attempts to highlight youths’ voices in this broader conversation about data and data literacy. In the following background section, we introduce several important concepts from prior research as a way of defining our terms. Next, we briefly introduce our empirical setting: a novel system built on top of Scratch called *Scratch Community Blocks*. Subsequently, we describe our own data and methodology. We present a unique dataset

<sup>1</sup>Although it is outside the scope of this paper, the legal and regulatory environment relating to collecting data on the online activities of children under 13 is complex and shifting. See boyd, Gasser, and Palfrey [6] for an overview of some of these issues.

of user discussions, ethnographic field notes, interviews, survey responses, and more than 1,600 creative artifacts that we collected over an approximately four-month period observing over 700 young programmers working with behavioral and social data in the Scratch online community. Using a grounded theory-based approach, we systematically coded and analyzed these qualitative data.

We reserve the bulk of this paper for discussing our findings. These findings are organized into five themes that reflect critical features of contemporary online data collection and analysis identified by the young people in our sample. We present these as youth perspectives on critical data literacies: (1) data collection and retention has privacy implications, (2) data analysis requires skepticism and interpretation, (3) data can come with assumptions and hidden decisions, (4) data-driven algorithms cause exclusion, and (5) measuring and reporting on data can affect the system that created the data. We use quotes and creative artifacts from youth to explicate each of these themes in detail. Furthermore, we connect these concerns to broader questions about data raised by researchers, critics, and educators. Based on these findings, we offer several suggestions for designers and educators on possible strategies for engaging youth in critically thinking about the role of data in their lives and in their world.

## BACKGROUND

Although our research design takes a bottom-up approach that is primarily grounded in the data we have collected, our work is deeply informed by previous research. First, the way in which we attempt to theorize how youth understand data is informed by Papert’s theory of *constructionism*. Second, the way we understand and frame our findings is informed both by recent attempts to understand “big data” as well as related literature on *data literacies*. Finally, our desire to support youth in critically engaging in the process of meaning creation and learning is informed by Freire’s concept of *critical pedagogy*.

*Constructionism* is a theory proposed by Seymour Papert and others [43, 30] that builds on Jean Piaget’s theory of constructivism [46], where the learner is seen as an active constructor of knowledge rather than being a passive recipient of information. Papert argued that this process happens in a more “felicitous” way when the learner is actively engaged in constructing an entity (physical, or virtual). This entity could then be used as an “object to think with.” An example of such an object is a set of differential gears that helped a young Papert grasp ideas from mathematics [42]. Our approach is informed by constructionism in that we have sought to understand youths’ conceptions of data and its limitations by looking at the ways in which youth engage in the construction of data artifacts in a social context.

The topic of *data literacy* has been defined in a number of ways by educators and education researchers. We follow the Oceans of Data Institute in using the term to describe the ability to “access, interpret, critically assess, manage, handle, and ethically use data” [40]. We believe that the term “critically” is central. Our definition of *critical* is drawn from Freire, who advocated learning processes in which learners would be enabled to critically perceive “the way they exist in the

world with which and in which they find themselves” [23]. Following Freire, our work seeks to contribute to the emerging literature on *critical data literacies* [20, 45]. This conversation has focused on calls for the cultivation of an understanding of the implications of large-scale data collection and analysis, which necessitates critical thinking about data, including issues of privacy, surveillance, and the power structures that enable data-driven processes to affect people’s lives.

Although literacy is often conceived of as being top-down—that is, defined by educators and policymakers—our approach is fundamentally bottom-up. We do not claim that the critical data literacies described by the youth in our paper are perspectives that all youth will, or should, share. Rather, we present these youths’ perspectives on critical data literacies as a way of including their voices in the emerging scholarly dialog around what data literacies might involve and what strategies may be of value to designers and educators in this space.

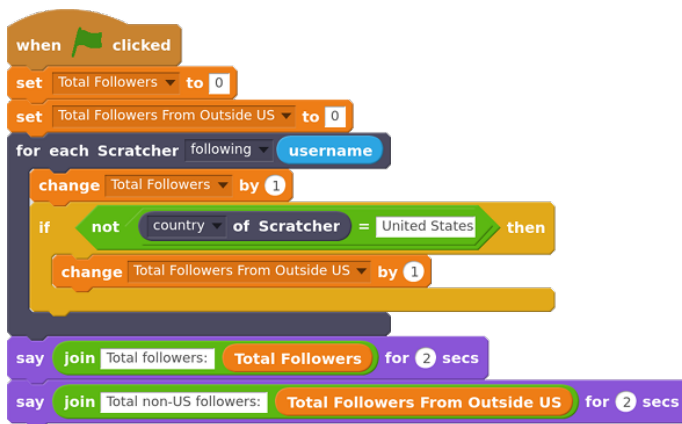
## EMPIRICAL SETTING

Our empirical setting is the Scratch online community. Scratch is a block-based visual programming language designed for youth aged 8–16 years old [50]. Interactive programs are constructed by dragging and dropping visual blocks together where blocks are roughly analogous to tokens in text-based programming languages. Scratch is embedded within an online social community with over 15 million registered users who can share projects, comment on each others’ work, follow each others’ activity, bookmark projects, and so on [37].

In early 2016, an experimental feature called *Scratch Community Blocks* was introduced to a subset of active Scratch users [15]. *Scratch Community Blocks* adds a new set of blocks to Scratch that allow users to access public metadata on users and projects from the Scratch website. The goal of the system is to engage Scratch users in programmatically accessing and analyzing these data. In particular, *Scratch Community Blocks* gives users access to the data about users’ followers, followees, shared projects, and favorited (bookmarked) projects. For each of these projects, it provides access to data including the number of views received, the number of “love-its” (similar to “likes” in other social media platforms), and code metadata, such as whether the project uses blocks that play music.

A sample project illustrating how *Scratch Community Blocks* can be used is shown in Figure 1. The code shown in Figure 1a calculates the total number of followers of the user who is viewing the project (using the “username” block embedded inside the “for each” block) as well as the total number of followers who are from outside the United States. Once the calculation of these numbers is complete, a graphical character on the screen “says” the result using a speech bubble (Figure 1b).

*Scratch Community Blocks* was beta tested on a fully functional copy of the Scratch website. Users with access to the beta site could create and publish projects that were visible to other invitees. Although the system was deployed on a separate website, the data fetched by the system was from the main Scratch website. Unless disabled by a project’s creators, users could leave comments on projects in a threaded discus-



(a) Code to access and process social data



(b) Results of running the code

Figure 1: Example of *Scratch Community Blocks* code. The code calculates the total number of followers of the viewer of the project and how many of them are from outside the United States.

sion area associated with each project. Additionally, there was a special threaded sub-forum in the main Scratch discussion forums dedicated to the community blocks and only accessible to invitees.

The system was beta tested over several months. Beta testers were drawn from a pool of Scratch users who had been active for at least six months and who had shared at least four projects. The self-reported age distribution of these users was close to that in the Scratch community in general—with the median age being 12 years, and with more than half being between 11 and 15. The gender ratio was approximately 48% male, 47% female, and 5% other. Selected users were given access to a separate instance of the Scratch website set up exclusively for trying out *Scratch Community Blocks*. Prior to the users being invited, the beta site did not have any content except for a few sample projects. A total of 2,500 Scratch users were invited in three stages. Over 700 of these users shared over 1,600 projects and 1,000 comments on the site. As the beta site became active, its landing page was populated with projects created by users. As with the main Scratch website, the source code for every project shared on the site could be viewed and remixed by any user.

## DATA AND METHODOLOGY

Our data was drawn from several sources, and our methodology consisted of a grounded theory-based ethnographic approach. Our primary source of data was ethnographic observation of users in the *Scratch Community Blocks* website during a four-month period of beta testing. All three authors visited both the *Scratch Community Blocks* website and forum frequently to browse projects, read comments, and to follow discussions in the special forum. Though primarily observers, we also participated by experimenting with the blocks ourselves, publishing several of our projects, and engaging in several discussions. Two of the authors played an active role in the design of the blocks, and one led the implementation. The first author exclusively focused on ethnographic observation. As a group, we attempted to be comprehensive in our

coverage of activity on the site and to view projects as they were published and discussions as they were ongoing.

The time that our team spent “in the field” was between February and July 2016 and varied day-to-day with activity on the site. The site saw more posts and projects during the days immediately following each wave of invitations and then slowly returned to a low level of activity. During the most active periods, our team members visited the site and forums several times a day. Over the course of our observations, we took copious field notes, which both quoted from discussions and included our own interpretations, feelings, and reactions.

We systematically collected the full text of the *Scratch Community Blocks* forums, which included 936 distinct messages left on 42 threads. We also drew from the more than 1,000 comments left on over 1,600 projects shared by users on the site (over 200 projects received comments). Additionally, we conducted five interviews with users of the system in which we discussed the blocks and the projects they had created. Three of these interviews were conducted with active users of the beta site over voice or video chat. The remaining two interviews were conducted face-to-face after the workshop described below. These interviews lasted a total of approximately three hours.

We also distributed a survey to users of the system in June 2016. All 2,500 users who were invited to the site were asked to participate in the survey, and over 400 responded. In this paper, we draw from eight open-ended questions that probed respondents’ reactions to and feelings about the blocks. We have included these questions in an online supplement. Finally, the team organized a three-hour face-to-face workshop where a total of 12 Scratch users were introduced to the blocks and then encouraged to explore, create, and discuss the blocks with the team and each other. Detailed notes were taken by the workshop facilitators.

We followed Charmaz’s [11] approach to grounded theory to code and analyze these various sources of data. One feature of

Charmaz’s approach is more explicit direction for researchers to incorporate prior ideas and existing theory, compared to Corbin and Strauss’ [12] original formulation. Following Charmaz, we complemented our inductive coding (the bulk of our codes) with “sensitizing” codes drawn from theory during initial coding as well as later-stage iterative recoding of data using codes informed by theory that we felt was relevant to emergent themes.

All of our coding and analysis was completed using the collaborative QDA software package Dedoose using a single codebook. Following Charmaz, we iteratively completed the following analytic cycle: coding data, discussing codes as a team, grouping similar codes together, identifying themes of interest, writing and revising memos, and recoding data. We coded and analyzed data continuously. As a result, data collected in the first month of the study (field notes, forums conversations, records of projects, and comments) were coded more openly to generate a large set of initial codes. This was intended to provide a basis for grouping and identifying themes that could be further explored. Data collected later in the study (such as the survey responses) were coded less openly.

Codes included phrases such as “user stalking,” “conflicts with other Scratchers,” “cyber-bullying,” and “too personal.” These codes, and many others, were collapsed into themes, such as “concerns about user behavior” or “concerns about privacy.” At the end of this process, we found that our themes and memos were largely focused on five key features of data collection and analysis identified by youth. We present these features as perspectives on critical data literacies.

## FINDINGS

In this section, each theme is framed with a brief introduction of the critical data literacy in question. We summarize youth’s perspectives on the issue they emerged in our analysis and provide evidence to support these summaries using quotes and examples from our fieldwork. We conclude each section by relating youth’s perspectives to existing scholarly criticism.

We have tried to provide context for quotes without revealing the identity of any users. We indicate the self-reported age of the Scratch user wherever possible but have altered all usernames. Because our survey was anonymous and all questions were optional, we have no identifiable or demographic information on survey respondents.

### Data collection and retention has privacy implications

A primary source of excitement about new forms of data collection is the growing ability to continually record data about individuals’ interactions and actions. However, because online data collection is passive, the subjects of observation are often not aware of it. Scholars have described this as “epistemic asymmetry” [9] because users are not aware that data collection, retention, and analysis is happening. Even though *Scratch Community Blocks* provides access to data that is public, the surfacing of this data collection caused many users to raise concerns. These concerns constitute our first critical data literacy: that data collection and retention has privacy implications.

Many of the projects on the beta site utilized blocks that automatically detected the username of the viewer. One project greeted users by asking them if they remember their first project before reciting that project’s title and a few statistics about it. This comparatively simple demonstration of the persistent nature of retained data was evocative for a number of users. In one comment on the project, a 10-year-old user said that she found it “creepy and very cool.” Another comment used the terms “scary” but “cool” to voice a similar sentiment.

The ability to retrieve information that had been published but forgotten about clearly surprised and intrigued Scratch users. Although they may have been aware that their activities were being recorded in some abstract sense, and although they explicitly chose to make their activities public by publishing their projects, *Scratch Community Blocks* revealed how easily this information could be retrieved computationally. Through comments on projects, survey responses, forum posts, and interviews, we observed how Scratch users struggled with the implications of what it means to have this data out there.

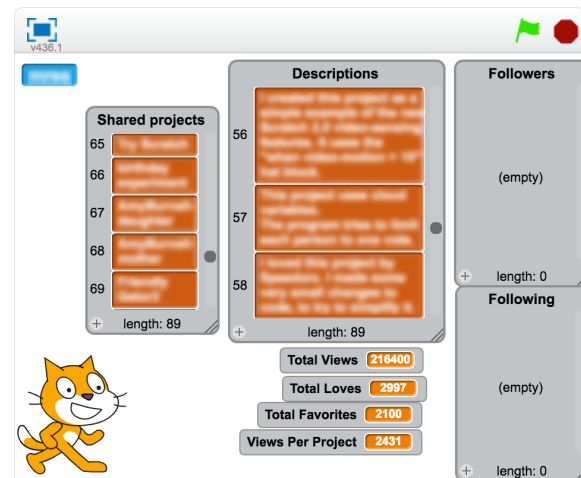


Figure 2: This “information retrieval” project calculates the cumulative number of views, loveits, and favorites received by a user over their entire trajectory. Textual data in the image has been blurred to preserve anonymity.

Some Scratch users quickly realized that computational access to their data enabled new possibilities. For example, although one can see the number of projects a user has created and their titles on the main Scratch website, calculating the cumulative views, comments, and loveits across all of a user’s projects without *Scratch Community Blocks* would involve clicking on each project individually and manually calculating the sums. With the new system, users could do so with a few blocks of code. An example created by a user is shown in Figure 2.

However, some Scratch users expressed concern about the retrieval of historical information and the potential for the blocks to enable “stalking.” Although the blocks only revealed information that was public, several Scratch users felt that the ability to programmatically retrieve information from anyone was invasive.

On the discussion forums, a 14-year-old user expressed concern that data about anyone on Scratch can be retrieved through the system:

If I had to say one thing, it may be a bit personal, considering you can type in ANYONE's username, but it's still great! That's only a minor flaw that can easily be fixed.

Another user (15 years old) echoed a similar sentiment in a forum post. She pointed out that programmatic access to public data makes it easier to actively monitor someone:

I will, however, add, that it does feel like a bit of a way to stalk Scratchers from a project. Though I suppose it doesn't make much of a difference since a stalker could always just stalk them on the site, but you get what I mean. Someone could set themselves up with some of these blocks and get themselves a constant relay of the person's activity.

Although she acknowledged that people are already looking at other user profiles and could "stalk" with or without the blocks, she also felt that programmatic access to data represented an important qualitative difference. In a survey response, a third user suggested that there should be a system that would ask for permission from a user before their data was accessed:

The blocks are a little personal; for example someone can find out anyone's follower account, views, loves, favorites, etc., so there should be something that asks for permission from the user.

Not everyone was concerned. When presented with the concerns that other users raised about privacy, one workshop participant (11 years old) provided a passionate defense of the use of the community blocks to retrieve information

I think that easier is never a problem. In fact, in our entire world. . . , everywhere in history, if you look around, technology upgrades either make our world easier, funner, or safer. Right? I mean like, vaccines, safer. Roller-coasters, funner. Computers, easier. And these Community Blocks definitely fall under easier and funner. It makes Scratch projects more diverse, which is great because projects are already diverse, and it makes it easier because you can access the information anyway. There's not any sort of privacy issues—but it makes it so much more easier and elegant to access.

The lack of consensus among the young users of *Scratch Community Blocks* on the privacy implications of new forms of data collection and new programmatic forms of data analysis mirrors disagreements among adults. boyd, et al. [7, 4], while describing *privacy* as, "both a social norm and a process," where privacy is "negotiated." Nissenbaum [39] argues for the development of "contextual" norms for data use and reuse. In other words, data collected for a stated use should not be used for another purpose unless permission is given.

What we observed through the conversation and feedback from Scratch users was essentially the process of privacy being "negotiated" (in the sense used by boyd et al.), as the socio-

technical system forming the contextual backdrop for data collection and reuse added a new affordance. The ability to recognize and navigate the often messy relationship between the details of data collection, access, and analysis with privacy reflects our first critical data literacy.

### Data analysis requires skepticism and interpretation

Though data can be a powerful tool for gaining insights, there are inherent challenges to data analysis as a means to understand the world. Data can be error prone and requires interpretation. Statistics educators developing curricula for elementary school students have described the "uncertainty and intrigue of real mathematics" [51] and have highlighted the need to have students discuss, debate, and argue as they conduct data analysis. Our observations indicate that youth using *Scratch Community Blocks* encountered these challenges in various ways. The examples that we provide in this section reflect the second critical data literacy—that data analysis requires skepticism and interpretation.

In our survey, we asked Scratch users what they liked best about the community blocks. Several users commented that they felt the blocks allowed for greater precision and confidence in the accuracy of information. For example, one respondent commented that the blocks made it possible to eliminate human error when retrieving the number of page views:

I found one project on the Scratch website (main) that added up your total views, but it required you to look through your projects and type in the number of views. This could easily result in human error, in a way that the community blocks won't [be] doing the same thing.

Another respondent echoed a similar opinion:

I think it's a pretty neat addition to the existing blocks. I am the kind of person who loves to know the exact numbers when it comes to stats. It honestly is an amazing idea; learning about how many people were on Scratch and etc. blew my mind.

Other Scratch users pointed out that the blocks occasionally appeared to report inaccurate information. For example, one interviewee described an inconsistency with data reported by the system:

At one point the follower blocks, it said I have slightly more followers than I do. And, that was kind of confusing when I was trying to make the project. [. . .] I pulled up a second [browser] tab and compared the [data from *Scratch Community Blocks* and the data in my profile].

Upon investigation, we established that the source of this discrepancy was a bug in *Scratch Community Blocks* where the block iterating over followers included users who had disabled their accounts. For users without visibility into the implementation of the system, the numbers returned simply appeared wrong.

These "incorrect" numbers were discussed by users because they did not match the numbers on the public user profiles. Through these discussions, we saw Scratch users engaging





Figure 3: The formula for ‘Total Scratcher Awesomeness’—an example of a project that uses a model to score and rate users.

in what has been described in a recently released workshop report on data literacy as data skepticism, or “questioning when something doesn’t make sense, particularly regarding the data sets and what they do/do not represent” [40]. Of course, *Scratch Community Blocks* provided access to some data that could not be validated. In many cases, such as the total number of Scratch projects, users had no way to know if data were wrong.

Moreover, many users aimed to measure less objectively quantifiable variables, such as “awesomeness” or “the best project.” To do so, Scratch users came up with models that combined a number of more accessible variables to represent what they thought was a reasonable proxy for the concept they were interested in. In this process, they drew on their knowledge of the community and their own values to make judgments about how to use the data they have access to.

For example, one user modeled “Total Scratcher Awesomeness” of a Scratch user by adding followers with cumulative loves, favorites, and views and subtracting the number of projects created. The Scratch code to perform this calculation is shown in Figure 3.

O’Neil [41] described models such as these as “opinions embedded in mathematics” and suggested that they are often very dangerous—particularly when they are opaque and unchangeable. Scratch users were expressing their opinion on what they think is an “awesome” Scratch user through the mathematics in the code shown in Figure 3. In this case, however, the model itself was both transparent and remixable, allowing other users to see what was incorporated and to offer their own suggestions. Indeed, we saw the author of this project incorporate feedback that suggested adding number of followers and total views to show how these models can potentially evolve to represent collective opinion. One such interaction is reproduced below:

**Commenter:** You should add how many people follow you and the total views! Cool idea!

**Creator:** good idea! I think i will add that...

Though we observed a number of “scoring” projects, and though these projects received a considerable amount of attention as measured by views and comments, we observed comparatively little concern or discussion about opinions being expressed through the models. On some of these projects, community members left comments to ask the project creators their reasoning behind the formulas they created. Creator responses were frequently variations on “I made it up.” What users chose to include and exclude in their models reveals some insight into their perceptions and beliefs. Several projects placed prominence on statistics that came out of interactions (such as likes and favorites) over those that came from user activity (such as projects created). We feel that exploring

ways to foster these discussions and how youth develop these models is a promising avenue for future work.

O’Neil argues that the process of interpreting data requires a thorough knowledge of context. The importance of context was clear among the youth we observed, not only in model building, but in other discussions as well. The culture of the Scratch community can skew certain variables. For example, a common cultural practice in Scratch involves users reporting their scores in game projects through the comments section. This means that games in Scratch often have more comments. This fact was not lost on the young data analysts we observed. In one project that attempted to identify the viewer’s most viewed and and most loved project, a commenter offered a suggestion to, among other things, display the viewer’s project with the most comments. The 11-year-old creator of the project responded positively but added that he felt that the number of comments might be more representative of genre than popularity.

**Commenter:** Woah, neat! A few ideas I have: Most commented on (as that may not be the same as most loved), newest project, and random project.

**Creator:** Nice ideas! Although, the most commented might end up being a game where there is a high-score system, as everyone comments their scores.

boyd and Crawford argue that without context, data can “lose meaning and value” [5]. Definitions of “data science” (e.g., [29]) also highlight the need for domain knowledge or subject expertise to understand the “why” of data as opposed to the “what” [48]. In the example above, we see Scratch users not only engaging in analyzing data but also doing so in a manner that indicates that they are aware of the mechanism through which this data is generated. This skepticism and knowledge that data must be put into context and interpreted reflects our second data literacy.

#### Data comes with assumptions and hidden decisions

Although data can have obvious errors, there are also more subtle issues that can cause the end result of data analysis to be biased or incorrect. Some scholars have framed this by suggesting that “raw data is an oxymoron” [25] or that raw data needs to be “cooked with care” [3]. The point of these critiques is that data is always shaped by decisions and assumptions made during the process of data collection, processing, and analysis. For example, if a surveyor decides to conduct a national poll over landline telephones, it will be biased toward a certain demographic. Any interpretation of the results of that poll should keep this potential source of bias in mind. Understanding the assumptions and decisions that accompany data is a critical step for consumers and purveyors of data science tools. This constitutes our third critical data literacy: that data comes with assumptions and hidden decisions.

In conversations involving projects created with *Scratch Community Blocks*, we saw that becoming fluent in this literacy is not always a straightforward process. For example, one project showed the cumulative views received by the viewer's projects as a percentage of the total number of users in Scratch. This led to some confusion as some users assumed that the project calculated what percentage of users in Scratch had viewed the project. The resultant conversation happened in the comment thread on the project page:

**Commenter A:** that's so cool! almost 00.5% of all the users on scratch have viewed my projects and that's a lot :B but crossstitch's<sup>2</sup> results are indeed slightly dubious. . . over 100% of people have viewed his projects which is awesome but impossible - love the project!! ^o^

**Commenter B:** @CommenterA I think it's because its based on views, not each specific player.

**Commenter A:** @CommenterB that's awesome :D people who haven't registered on scratch have viewed a significant amount of his projects yes

**Project Creator:** @CommenterA Yeah what @CommenterB said is correct.

Here we see an example of collective sense-making. Commenter A began by questioning the validity of a result and pointing out a discrepancy. In the following discussion, users tried to puzzle through the ways in which this discrepancy could have occurred. The creator of the project in question was aware of this particular quirk of the project, and when sharing the project, he had added the following note to the project description:

This project does not and cannot calculate unique viewers. So some views can be the exact same person who viewed the project earlier.

The discussion above reflects a broader issue in interpreting data made available by centralized socio-technical systems such as social networks. While data may be publicly viewable, the decisions, processes, and algorithms behind the construction of this data are not. Although the project creator expressed some confidence in the fact that the project view count returned by *Scratch Community Blocks* double-counted at least some users, the details of how this number is created is not public or known. The full answer is very complicated and the author of the system that computes the view count struggled to explain it to our team. In absence of documentation from the team implementing and running the Scratch website, understanding the mechanism of how this particular variable is constructed is difficult.

This issue is not specific to Scratch. YouTube's support page on view counts state:

Video views are algorithmically validated to maintain fair and positive experiences for content creators, advertisers, and users. To verify that views are real and accurate,

YouTube may temporarily slow down, freeze, or adjust the view count, as well as discard low-quality playbacks.<sup>3</sup>

It is not clear what is meant by fair, positive, or a low-quality playback. While there may be reasons for Scratch and YouTube's creators to not divulge the way views are counted, the result is that users make assumptions and guesses, often without realizing they are doing so. Numbers are treated as neutral and unproblematic facts. In the conversation above, Scratch users realized that even a simple number may have a complex story behind it and began reverse-engineering the process that created it.

In grappling with the often opaque nature of data, many Scratch users created projects that sought to add context and meaning to data. When asked what they liked best about the new blocks, one survey respondent said,

I really liked the possibility of creating a more meaningful high score list. Using an existing cloud high score list and being able to include a user's country and about me section would really be cool.

In this case, the survey respondent recognized that the existing high score systems in Scratch games provided an incomplete picture. By constructing a high score system that provided additional context, they were able to make sense of the data in a way they considered meaningful. In these conversations about context, assumptions, and hidden decisions, these users helped build toward an important critical data literacy.

#### Data-driven algorithms can cause exclusion

Data and data-driven algorithms can be used for discrimination and for exclusion [13]. The potential for data-driven decisions to result in exclusion was recognized in a 2014 report to the office of the US president, which stated:

A significant finding of this report is that recent changes in data collection and analysis have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace. [47]

This represents our fourth critical data literacy: that data-driven algorithms can lead to exclusion.

The possibility of discrimination based on data was recognized by Scratch users, and it was a frequent theme in their concerns and critiques. In a forum post, one 13-year-old pointed out that since it was possible to detect how many followers the viewer of a project had, users could create code to prevent users with a low follower count from accessing a project:

I love these new Scratch Blocks! However I did notice that they could be used to exclude new scratchers or scratchers with not a lot of followers by using a code: like this:

```
when flag clicked
  if then user's followers <300
    stop all.
```

I do not think this a big problem as it would be easy to

<sup>2</sup>crossstitch is the creator of some of the most popular projects on Scratch.

<sup>3</sup><https://support.google.com/youtube/answer/2991785> (accessed in September 2016)

remove this code, but I did just want to bring this to your attention in case this not what you would want the blocks to be used for.

Although often more subtle and complex than described here, algorithmic discrimination has received significant attention among scholars [24] and in the popular press [31]. *Scratch Community Blocks* offered an opportunity for Scratch users to grapple with this debate. The user quoted above was concerned about algorithms designed to exclude users who are new to the community or less popular than their peers. He also suggested that it would be possible to “remove this code”—hinting at the possibility for oversight or monitoring mechanisms that could address such discrimination.

In the conversations around this type of code, not all comments were expressions of concern. A number of users saw the new affordance as adding new possibilities for incentives in games built with Scratch:

You know how in certain apps, they have things like “like us on [Facebook], and get 3 coins” I could do something like “follow WillTheSuperman, and get 25 more power-ups” or something like that.

This quote, which describes a potential incentive system for gaining followers, shows how Scratch users built on their experience on platforms outside Scratch that use data to implement similar types of incentives. Scratch users recognized that they could use data to shape their peers’ behavior. Not all users agreed that such incentive mechanisms would be a positive addition to what is possible in Scratch. One 16-year-old wrote:

People can use these new blocks so that people are forced to follow them. I know, this could be a way for them to get followers, but I think it could go out of hand. Say, they were making a story, they post one chapter. They are hooked, so they go onto the next chapter. Then, on the second chapter, it says: “You have to follow me to read this chapter!” It’s forcing upon them to follow this user in order to read the next chapter. I mean, this is a effective way to get followers, but it just doesn’t seem right.

Using a concrete example of a hypothetical project that forces Scratch users to follow the creator to continue accessing the project, this Scratch user expressed her concern that an algorithm with access to behavioral data could intentionally shape user behavior.

While several users recognized problems, others articulated potential solutions like the first user quoted in this section. One survey respondent suggested that this would be a method of editing out the discriminatory algorithm:

I thought that this community blocks idea might lead to some exclusion. Some people might start using the code to only let popular people play their games. This won’t be too much of a problem though, because people can still edit the code and make it playable.

That said, this particular workaround is possible only because Scratch code is always open, visible, and editable by any user.

Even then, it is only possible when a Scratch user has the necessary skills to make the changes to the discriminating code.

Outside Scratch, most algorithms are inaccessible. They are typically considered trade secrets or are too complicated for non-specialists to understand, let alone modify [41]. The need for technical sophistication to address this issue was not lost upon Scratch users—a 16-year-old community member wrote in a forum post:

Just as a few others mentioned already though, we should make sure that it is not used to discriminate New Scratchers who not only likely not have many followers, projects, and comments, but also likely have a more limited knowledge of the technology, so they might not know how to change the project inside. I know that I wouldn’t have.

These comments reflect sophisticated views about how access to data could affect the experience of individual Scratch users. Scratch users were able to imagine ways to use data in ways they considered both desirable and undesirable. They could conceptualize the potential effects of data discrimination on both an individual and at the community level.

In the report to the US president cited earlier, the authors raise concerns about “digital redlining,” data-driven practices that may cause exclusion in the same way that banks refused to give loans to customers from certain neighborhoods during the first half of the 20th century. In the comments above, we see similar concerns within the context of Scratch, acknowledging that while data is a powerful means to understand the world, it can also establish new forms of discrimination and exclusion. In their discussion of digital exclusion and potential responses to it, Scratch users provided insight into a fourth critical data literacy.

#### **Measuring and reporting data can affect the system that created the data**

In trying to understand and evaluate a system and its effects, there is often a tendency to value measures that are easily quantifiable. Because of their increased value, these measures are often used to evaluate performance. As a result of this process, the system that produces the underlying data can become distorted. This phenomenon of quantitative indicators ending up distorting the process or the system was noticed and documented by Campbell [10]. A classic example of this can be seen in formal educational contexts where quantifiable test scores get more attention than may be warranted. Scholars and educators such as Holt [27] have argued that education as a field has become overly focused on test scores with teachers teaching specifically for tests and ignoring less quantifiable components of education such as fostering expression and enabling creative solutions to unanticipated problems. This represents the fifth and final youth-identified critical data literacy—that measuring and reporting data can affect the systems creating the data.

Some of this information made accessible by *Scratch Community Blocks* was not easily accessible otherwise. This was particularly the case with code metadata in Scratch projects.



Using *Scratch Community Blocks*, a number of projects represented and visualized code metadata from users' trajectories in different ways. These visualizations caused some users who viewed these projects to look back on their participation and to reflect on the type of blocks they tended to use more, and the types they used less. Fostering self-reflection through data was an explicit design goal of *Scratch Community Blocks* [15].

However, this affordance also introduced the risk of an unanticipated consequence. The original design goals of Scratch [50, 52] treated creativity, collaboration, and respect for one's peers as paramount. In such a context, there is a risk that users of *Scratch Community Blocks* could shift behaviors in Scratch in deleterious ways by highlighting and incentivizing behaviors—such as attracting large number of followers—that are easy to measure and report but are, at best, orthogonal to Scratch's goals.

One survey respondent, when prompted for any additional comments on the community blocks, stated their belief that the projects created with the blocks had the potential to affect what people perceived as important:

I am just a bit worried that people will make these projects and take it the wrong way, saying that followers are the most important thing on Scratch.

Echoing a similar concern, a 12-year-old user wrote in the forums that this can lead to disagreements and conflict within the community:

These new features are really helpful! However, I think that it would have a possibility to leading to flame wars or fights about followers.

Similarly, another 12 year old user imagined that users with a smaller number of followers could be made fun of and harassed:

But a few things, you can easily make fun of someone for example, "You only have 2 followers! Ha! Well I have 10!"

These concerns were not totally unfounded. We found that users were treating number outputs by projects (e.g., the Scratcher awesomeness calculator described earlier, which took into account number of followers) as scores and reporting them (sometimes competitively) as project comments. For example, in one comment, a user wrote "9189. ITS OVER NINE THOOOUUUUSAAAAAND!!!!!!" We also found a few instances of users creating projects that would generate mocking comments for users with a smaller number of projects and followers.

boyd and Crawford, quoting Du Gay and Pryke [18], note that, like accounting tools, data analysts "do not simply aid the measurement of economic activity, they shape the reality they measure," so Big Data stakes out new terrains of objects, methods of knowing, and definitions of social life" [5]. Similarly, the introduction of new affordances around data into Scratch can have far-reaching consequences.

The problems youth observed and described are not entirely new to Scratch. Indeed, incentivizing following behavior ex-

isted in Scratch before the introduction of *Scratch Community Blocks*. That said, Scratch users recognized that as collecting and using data became easier, it could influence behavior and, ultimately, the nature of the data itself. The ability of youth to recognize the way these blocks could affect the community in which they participate is a manifestation of a final critical data literacy.

## DISCUSSION

Our findings show that the way that youth critique, question, and debate the implications of data analytics reflects some of the existing critical discourse pertaining to data. That said, not all themes explored by scholars were raised by youth. Two notable critiques that were absent from the conversations are the point that larger data sets may not necessarily mean better data and the point that new forms of data collection and analysis can create new types of digital divides by enabling only a handful of organizations and individuals with significant resources to perform data collection and analysis [5]. We feel that these topics did not come up because they pertain more to "big data" than the types of social media analytics supported by *Scratch Community Blocks*. Moreover, youth using *Scratch Community Blocks* had little say in what data was being collected or presented to them. Relatedly, we do not know whether the critical perspectives that we described in the context of Scratch translate into insights about data practices in other social networks (e.g., Facebook or Snapchat). Understanding and addressing these gaps remain a topic for future work.

Another limitation of our approach was that although *Scratch Community Blocks* was based on Scratch, our setting was a context where we explicitly encouraged the use of data. We know from previous work by Dasgupta et al. [14] that a majority of Scratch users use only a subset of the approximately 130 code blocks that are available to them. We do not know what patterns of engagement would have emerged if these were introduced in the larger Scratch community. Strategies to engage larger groups in discussions of these questions remain an open challenge.

Even within this limited setting, discussions were driven by the agency of the Scratch users concerned. Users encountered and explored the five critical data literacies described in this paper on their own. These users were operating within the culture of Scratch, a familiar context, as they engaged in the discussions and sense-making that we have highlighted in this paper. Bruner describes the importance of such contexts in meaning making as follows:

Meaning making involves situating encounters with the world in their appropriate cultural contexts in order to know "what they are about." Although meanings are "in the mind," they have their origins and their significance in the culture in which they are created.[8]

In a similar vein, Papert and his collaborators have described the idea of Mathland, a cognitive space where mathematical ideas are encountered naturally by the inhabitants [42]. Learning mathematics in Mathland, constructionists argue, should

be similar to the experience of a child learning French in France.

In light of these theories, we can see that as the young people using *Scratch Community Blocks* grappled with the implications of social data analytics, they were supported by the fact that their encounters were situated in an appropriate and familiar cultural context. As educators consider a critical pedagogy of data, our work shows that situating the learning experience in a familiar cultural context can play an important role. Our work points to the possibility that enabling learners to analyze social and behavioral data can spark conversations that explore critical data literacies.

In May 2016, *Science* magazine published an article [36] describing the long-term vision for “big ideas” at the US National Science Foundation. One of the specific research areas constituted the development of “learning opportunities and educational pathways” driven by an “understanding of the knowledge and skill demands needed by a 21st century data-capable workforce.” Earlier during the year in January 2016, the US Federal Trade Commission released a report with the title “Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues” [22]. These acknowledgements, along with the concerns raised by academics and advocates, indicate that an essential piece of a “data-capable workforce” are critical data literacies of the kinds we have described here. These reports suggest that what may be at stake is not just economic productivity and competitiveness but also some of our core democratic and civic values.

In this paper, we show that it is possible for young people to discover, critique, and question the implications of digital data collection in their lives, frequently echoing the concerns raised by academic scholars. Defining critical data literacies is still a topic of ongoing conversation. We hope that the young people’s perspectives presented here can contribute to this discussion. We show that not only are youth receptive to discussions about critical data literacies, they also have something to teach us as well.

#### ACKNOWLEDGMENTS

Parts of this paper are included in Dasgupta’s PhD dissertation, and some of the text and images in the paper appear in his thesis. We would like to thank members of the Scratch community for their time and for inspiring us with their creative work. We would also like to acknowledge Mitchel Resnick, Natalie Rusk, Hal Abelson, and our anonymous reviewers for their support and thoughtful feedback. Financial support for this work came from the National Science Foundation (grants DRL-1417663 and DRL-1417952).

#### REFERENCES

1. Hal Abelson, Ken Ledeen, and Harry Lewis. 2008. *Blown to Bits: Your Life, Liberty, and Happiness After the Digital Explosion* (1 edition ed.). Addison-Wesley Professional, Upper Saddle River, NJ.
2. Marie Bienkowski, Mingyu Feng, and Barbara Means. 2012. *Enhancing teaching and learning through educational data mining and learning analytics: An issue*

*brief*. Technical Report. US Department of Education, Office of Educational Technology.

3. Geoffrey C. Bowker. 2005. *Memory Practices in the Sciences*. MIT Press.
4. danah boyd. 2014. *It’s Complicated: The Social Lives of Networked Teens*. Yale University Press.
5. danah boyd and Kate Crawford. 2012. Critical Questions for Big Data. *Information, Communication & Society* 15, 5 (2012), 662–679. DOI : <http://dx.doi.org/10.1080/1369118X.2012.678878>
6. danah boyd, Urs Gasser, and John Palfrey. 2010. How the COPPA, as Implemented, Is Misinterpreted by the Public: A Research Perspective. (April 2010). [http://cyber.harvard.edu/sites/cyber.harvard.edu/files/COPPA\\_Hearing\\_Statement\\_boyd\\_Gasser\\_Palfrey\\_4-29-10.pdf](http://cyber.harvard.edu/sites/cyber.harvard.edu/files/COPPA_Hearing_Statement_boyd_Gasser_Palfrey_4-29-10.pdf)
7. danah boyd, Alice Marwick, and others. 2011. Social Privacy in Networked Publics: Teens Attitudes, Practices, and Strategies. (2011).
8. Jerome S. Bruner. 1996. *The Culture of Education*. Harvard University Press.
9. Finn Brunton and Helen Nissenbaum. 2011. Vernacular resistance to data collection and analysis: A political theory of obfuscation. *First Monday* 16, 5 (2011).
10. Donald T. Campbell. 1979. Assessing the impact of planned social change. *Evaluation and Program Planning* 2, 1 (Jan. 1979), 67–90. DOI : [http://dx.doi.org/10.1016/0149-7189\(79\)90048-X](http://dx.doi.org/10.1016/0149-7189(79)90048-X)
11. Kathy Charmaz. 2006. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. Sage Publications, London.
12. Juliet M. Corbin and Anselm C. Strauss. 2007. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (3rd ed.). Sage Publications, Inc.
13. Kate Crawford, Mary Gray, and Kate Miltner. 2014. Big Data| Critiquing Big Data: Politics, Ethics, Epistemology | Special Section Introduction. *International Journal of Communication* 8, 0 (2014). <http://ijoc.org/index.php/ijoc/article/view/2167>
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 Conference on Computer Supported Cooperative Work & Social Computing (CSCW ’16)*. ACM.
15. Sayamindu Dasgupta and Benjamin Mako Hill. 2017. Scratch Community Blocks: Supporting Children as Data Scientists. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ’17)*. ACM, New York, NY, USA. DOI : <http://dx.doi.org/10.1145/3025453.3025847> (forthcoming).

16. Åse Dragland. 2013. *Big Data—for better or worse*. Technical Report. Sintef.  
<http://www.sintef.no/en/latest-news/big-data--for-better-or-worse/>
17. DRL. 2013. *Data-Intensive Research to Improve Teaching and Learning - An Ideas Lab to Foster Transformative Approaches to Teaching and Learning*. Program Solicitation 13-565. National Science Foundation. <https://www.nsf.gov/pubs/2013/nsf13565/nsf13565.htm>
18. Paul du Gay and Michael Pryke. 2002. Cultural Economy: An Introduction. In *Cultural Economy: Cultural Analysis and Commercial Life*. Cultural economy: Cultural analysis and commercial life. SAGE Publications Ltd, London, UK, 1–20. <http://sk.sagepub.com/books/cultural-economy/n12.xml>
19. Elizabeth Dwoskin. 2014. Big Data’s High-Priests of Algorithms; ‘Data Scientists’ Meld Statistics and Software for Find Lucrative High-Tech Jobs. *Wall Street Journal (Online)* (Aug. 2014).  
<http://search.proquest.com/newsstand/docview/1552020409/abstract/D70B27FC5DA74D5APQ/1>
20. Catherine D’Ignazio and Rahul Bhargava. 2015. Approaches to Building Big Data Literacy. In *Proceedings of the Bloomberg Data for Good Exchange Conference 2015*. New York, N.Y. [https://dam-prod.media.mit.edu/x/2016/10/20/Edu\\_D\\_Ignazio\\_52.pdf](https://dam-prod.media.mit.edu/x/2016/10/20/Edu_D_Ignazio_52.pdf)
21. Liran Einav and Jonathan D. Levin. 2013. *The Data Revolution and Economic Analysis*. Working Paper 19035. National Bureau of Economic Research. <http://www.nber.org/papers/w19035>
22. Federal Trade Commission. 2016. *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues*. Technical Report.  
<https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>
23. Paulo Freire. 2000. *Pedagogy of the oppressed*. Bloomsbury Publishing.
24. Tarleton Gillespie. 2014. The Relevance of Algorithms. In *Media technologies: Essays on communication, materiality, and society*, Pablo J. Boczkowski and Kirsten A. Foot Tarleton Gillespie (Ed.). MIT Press, Cambridge, MA, 167.
25. Lisa Gitelman. 2013. *Raw Data Is an Oxymoron*. MIT Press.
26. Scott A. Golder and Michael W. Macy. 2014. Digital Footprints: Opportunities and Challenges for Online Social Research. *Annual Review of Sociology* 40, 1 (July 2014), 129–152. DOI : <http://dx.doi.org/10.1146/annurev-soc-071913-043145>
27. John Caldwell Holt. 1982. *How Children Fail*. Da Capo Press.
28. NSF-NIH Interagency Initiative and others. 2012. *Core techniques and technologies for advancing big data science and engineering (BIGDATA)*.
29. Iain Johnstone and Fred Roberts. 2014. *Data Science at NSF*. Technical Report. National Science Foundation. [http://www.nsf.gov/attachments/129788/public/Final\\_StatSNSFJan14.pdf](http://www.nsf.gov/attachments/129788/public/Final_StatSNSFJan14.pdf)
30. Yasmin B. Kafai. 2006. Constructionism. In *The Cambridge Handbook of the Learning Sciences* (1st ed.), Keith R. Sawyer (Ed.). Cambridge University Press, 35–46.
31. Lauren Kirchner. 2015. When Discrimination Is Baked Into Algorithms. *The Atlantic* (Sept. 2015).  
<http://www.theatlantic.com/business/archive/2015/09/discrimination-algorithms-disparate-impact/403969/>
32. Bruno Latour. 2007. Beware, your imagination leaves digital traces. *Times Higher Literary Supplement* (April 2007).  
<http://bruno-latour.fr/sites/default/files/P-129-THES-GB.pdf>
33. David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. The Parable of Google Flu: Traps in Big Data Analysis. *Science* 343, 6176 (March 2014), 1203–1205. DOI : <http://dx.doi.org/10.1126/science.1248506>
34. Caitlin Lustig, Katie Pine, Bonnie Nardi, Lilly Irani, Min Kyung Lee, Dawn Nafus, and Christian Sandvig. 2016. Algorithmic Authority: The Ethics, Politics, and Economics of Algorithms That Interpret, Decide, and Manage. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA ’16)*. ACM, New York, NY, USA, 1057–1062. DOI : <http://dx.doi.org/10.1145/2851581.2886426>
35. Mary Madden, Susannah Fox, Aaron Smith, and Jessica Vitak. 2007. *Digital Footprints: Online identity management and search in the age of transparency*. Pew Internet & American Life Project, Washington, DC. [http://www.pewinternet.org/files/old-media/Files/Reports/2007/PIP\\_Digital\\_Footprints.pdf.pdf](http://www.pewinternet.org/files/old-media/Files/Reports/2007/PIP_Digital_Footprints.pdf.pdf)
36. Jeffrey Mervis. 2016. NSF director unveils big ideas, with an eye on the next president and Congress. *Science* (May 2016). DOI : <http://dx.doi.org/10.1126/science.aaf5713>
37. Andrés Monroy-Hernández. 2007. ScratchR: sharing user-generated programmable media. In *Proceedings of the 6th international conference on Interaction design and children (IDC ’07)*. ACM, New York, NY, USA, 167–168. DOI : <http://dx.doi.org/10.1145/1297277.1297315>
38. Dale Neef. 2014. *Digital Exhaust: What Everyone Should Know About Big Data, Digitization and Digitally Driven Innovation* (1 edition ed.). Pearson FT Press, Indianapolis, IN.

39. Helen Nissenbaum. 2009. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.
40. Oceans of Data Institute staff. 2016. *Building Global Interest in Data Literacy: A Dialogue*. Technical Report. Oceans of Data Institute, Education Development Center, Inc. 24 pages. <http://oceansofdata.org/our-work/building-global-interest-data-literacy-dialogue-workshop-report>
41. Cathy O’Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown/Archetype.
42. Seymour Papert. 1980. *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, Inc., New York, NY, USA.
43. Seymour Papert and Idit Harel. 1991. Situating constructionism. In *Constructionism*. Vol. 36. Ablex Publishing, New York, NY, US, 1–11.
44. Chanda Phelan, Cliff Lampe, and Paul Resnick. 2016. It’s Creepy, But It Doesn’t Bother Me. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI ’16)*. ACM, New York, NY, USA, 5240–5251. DOI : <http://dx.doi.org/10.1145/2858036.2858381>
45. 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 (2013), 103–120. DOI : <http://dx.doi.org/10.1007/s10758-013-9202-4>
46. Jean Piaget. 1970. *Genetic epistemology*. Trans. E. Duckworth. Columbia University Press, New York, NY, US.
47. John Podesta. 2014. *Big Data: Seizing Opportunities, Preserving Values*. Technical Report. Executive Office of the President, United States of America.
48. Jake Porway. 2013. You can’t just hack your way to social change. *Harvard Business Review* 7 (2013).
49. Foster Provost and Tom Fawcett. 2013. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* 1, 1 (Feb. 2013), 51–59. DOI : <http://dx.doi.org/10.1089/big.2013.1508>
50. 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 (2009), 60–67. DOI : <http://dx.doi.org/10.1145/1592761.1592779>
51. Susan Jo Russell and Rebecca B. Corwin. 1989. *Statistics: The Shape of the Data. Used Numbers: Real Data in the Classroom. Grades 4-6*. Dale Seymour Publications, Palo Alto, CA. <http://eric.ed.gov/?id=ED328451>
52. Douglas Thomas and John Seely Brown. 2011. *A New Culture of Learning: Cultivating the Imagination for a World of Constant Change* (1 edition ed.). CreateSpace Independent Publishing Platform, Lexington, Ky.
53. Matthew A. Waller and Stanley E. Fawcett. 2013. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics* 34, 2 (June 2013), 77–84. DOI : <http://dx.doi.org/10.1111/jbl.12010>