# ‘Don’t make assumptions about me!’: Understanding Children’s Perception of Datafication Online

GE WANG, Department of Computer Science. University of Oxford, UK
JUN ZHAO, Department of Computer Science. University of Oxford, UK
MAX VAN KLEEK, Department of Computer Science. University of Oxford, UK
NIGEL SHADBOLT, Department of Computer Science. University of Oxford, UK

*Datafication*, which is the process in which children’s actions online are pervasively recorded, tracked, aggregated, analysed, and exploited by online services in multiple ways that include behavioural engineering, and monetisation, is becoming increasing common in the online world today. However, we know little about how children feel about such practices and how they perceive datification. Through online interviews with 48 children aged 7-13 from UK schools, we examined how children perceive datafication practices, especially how such practices could make inference on them. We identified three key knowledge gaps in children’s perceptions, including their lack of recognition of who were involved in the data processing and how, data being transmitted across platforms, and their data ownership. Through situating our findings under a critical algorithmic literacy framework, our findings provided some immediate indications regarding how we could better support children in the datafied society through more transparency and autonomy-supportive designs, as well as the need for a fundamental shift of the current data governance structure.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; • **Security and privacy** → **Social aspects of security and privacy**.

Additional Key Words and Phrases: Datafication, Data Inference, Online Platforms, Children

**ACM Reference Format:**
Ge Wang, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2022. ‘Don’t make assumptions about me!’: Understanding Children’s Perception of Datafication Online. In *CSCW ’22: ACM Conference on Computer-Supported Cooperative Work and Social Computing, Woodstock, NY*. ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/1122445.1122456

## 1 INTRODUCTION

Today, children are spending more time online than ever before. Online technologies are increasingly important for children, providing access to vital education, socialisation, participation, well-being, and entertainment resources [48]. In the UK, for instance, 96% of children aged 5-15 are online, and more than half of ten-year-olds go online using their own devices [64]. Thus, the Internet has now become seen as an essential enabler for children to learn, have fun and grow, especially during a pandemic, in which much of ordinary life has shifted online [64].

This rapid adoption and increasing reliance on the online world has raised corresponding concerns about the long-term effects of *datafication*, in which children’s actions are pervasively recorded, tracked, aggregated, analysed, and exploited by online services in multiple ways that include behavioural engineering, and monetisation [54, 57, 86]. At the core of this datafication is online services’ ability to make *data inference* on users, that is to analyse data, supported by algorithms,


*CSCW ’22, January, Woodstock, NY*

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

with the aim to evaluate certain personal aspects relating to a natural person [48], in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements [4]. Such datafication is practically impossible to avoid, or undo through deletion [48]. Children's life are now being routinely quantified, measured and used to profile and predict practices that could in return, have short and long-term implications for them [18, 51, 86]. Such activities take place invisibly behind the scenes of apps and services, and are less well understood or discussed as risks than other kinds of more easily characterised harms, such as the collection or disclosure of particular kinds of sensitive data. Given that most adults have little understanding of how their own data are being collected, processed, and used to shape their digital environments [13], it is not particularly surprising that children too, lack a robust understanding or adequate mental models of how their data are processed or used [43, 84]. Specifically, children were shown to have less awareness about data collection relating to institutional, or commercial use and exploitation [74]; namely, how their data were used by companies to serve them adverts or target and nudge them into taking particular actions or directions [66].

We feel that it is important to understand children's perceptions and how they interpret those online datafication practices, not only because it might identify common barriers and knowledge gaps in children, but also lead to better ways to empower and protect them online. In this paper, we used the YouTube platform as an example datafication platform [55, 60] experienced by children on a everyday basis, throughout the world [27, 63]. We report our results based on one-to-one interviews with 48 children, aged 7-13, from UK schools undertaken between November and December 2021. This paper is the first to contribute an understanding of the perceptions of children on the datafication and the more specific data inference practices that dominates their information consumption online. We identified three key knowledge gaps in children's current awareness and perceptions of datafication practices online, including their lack of recognition of (i) who are involved in the data processing and how, (ii) data being transmitted across platforms, and (iii) their data ownership. Through situating our findings under a critical algorithmic literacy framework, our findings provided some immediate indications regarding how we could better support children in the datafied society through more transparency and autonomy-supportive designs, as well as identified the need for a fundamental shift of the current data governance structure.

## 2 RELATED WORK

### 2.1 Datafication on Online Platforms

To establish the scope of our investigation, we first aim to define what we mean by *datafication* online, this refers to the process that children's actions are pervasively recorded, tracked, aggregated, analysed, and exploited by online services in multiple ways that include behavioural engineering, and monetisation [54, 57, 86]. To be more specific, we would like to draw on Livingstone's digital data types framework [48], in which digital data is categorised into three types: *data given*, which is the data contributed by individuals during their participation online; *data traces*, which is the data left by participation online and captured via data-tracking technologies such as cookies, web beacons or device/browser fingerprinting, location data and other metadata; and *inferred data*, the data derived from analysing data given and data traces, often by algorithms, possibly combined with other data sources. The *Data Inference* process for inferred data is core to datafication online, and refers to the process of analysing data, supported by algorithms, with the aim to evaluate certain personal aspects relating to a natural person [48], in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements [4].

By making inferences about individual's lives, datafication has been seen as more than a violation of privacy, or issue of data protection [13]; instead, it is more appropriately seen as a potential threat to human autonomy [78] brought by increasingly sophisticated dataveillance techniques [72]. On the other hand, datafication practices are becoming increasingly common in the online world today, and in fact, can be found on almost any online platform [14, 39, 68]. Through a study done on three different companies - Bluekai, Google and Yahoo, Rao et al. [68] found a wide range of data including demographic data, data on interests and attitudes and more have been used to make inferences about individuals or groups, which includes predictions about their future actions and inactions, general characteristics and specific preferences. Facebook has been found to make inferences on its users to form an 'interested reading' of their digital trace data [70], so as to create interest classifications that produce sales for advertisers and maintain user engagement on the news feed [76]. Research on Instagram also showed that there have been profiling practices on its users, and users were nudged towards certain content such as idealised images which could have negative impacts on the body satisfaction of young girls [65]. Similarly, there has been evidence on YouTube conducting inference on users to maximise their engagement on the platform, which could be particularly problematic for the minors [65].

Meanwhile, various regulations have attempted to protect children from such practices, such as COPPA [1] protects against the online collection of personal information of children under 13 years of age, and GDPR [16] sets restrictions on profiling on children. However, the above mentioned platforms managed to find exemptions for themselves by claiming there being a minimum age requirement of 13 year old for children to use their service. That said, there has been clear evidence that children, especially those under 13s are still on these platforms, and in fact, have quite heavy usage on such platforms [62, 75]. A recent report on 2,002 US children showed that 45% of kids under 13 were already on Facebook, and 40% already used Instagram [75]. YouTube in particular, although has a 'YouTube Kids' version that was claimed to be for under 13s, the most recent Ofcom report still showed that YouTube remained to be the most popular video-sharing-platform among 8-12 year-olds, and more than 85% of preschoolers were found to most commonly used YouTube to watch content [62].

### 2.2 Children's Perceptions of Datafication Online

There has been growing concern expressed relating to the datafication of children, especially as children may lack the awareness, knowledge, or mental faculties to be able to understand or be aware of such practices. Kumar et al. [42] found that children between the ages of 8 and 11 started to understand that data collection on online platforms could create some risks for them, but tended to associate such risks mainly with 'stranger-danger'. Zhao et al. [85] found through focus group studies with UK children aged 6-10, that while they could identify and articulate certain privacy risks well, such as information oversharing or revealing real identities online; however, they had less awareness of other risks, such as online tracking or personalised promotions. A recent report in 2020 [74] based on focus group interviews with 169 UK children aged 11–16 showed that children primarily conceptualized data in relation to interpersonal contexts, but had misapprehensions about how personal data was collected, inferred and used by organizations, whether public institutions, such as their schools, or private/commercial businesses. These findings were compatible with findings from a study in 2017 [67], in which teenagers aged 14 to 18 were found to have more concerns over interpersonal contexts, but often failed to understand or perceive potential threats to their privacy from ways first and third parties might make use of their data, and how personal data could be used in predictive ways to shape their future experiences and behaviors. Children between 12-17 demonstrated some awareness of 'data traces' they left online [83] and of device tracking [53], but found it hard to make a personal connection or apply such knowledge to themselves [7].

Research has also found that even at this age, children had little knowledge of data flows and cross-platform data sharing, and they found it hard to view data flows as a dynamic process instead of a static one [11]. It was also found that children struggled to grasp the relation between privacy and data, and they would only focus on data they know they give, much more than data that is taken or inferred [48].

While children would likely have difficulties in fully understanding the complexities of datafication or its means, some recent research has shown that children were well-equipped and capable of grasping essential concepts related to datafication, such as that their personal data (such as activity history) could be processed and used to sell products to users such as themselves, if they were given sufficient scaffolding and nudges from parents and educators , and their understanding grows with experience [48]. That said, a project working with over 200 children aged 8-12 with the goal of developing critical data literacy, showed that repeated exposure to a phenomenon, in this case online datafication practices, did not produce knowledge in and of itself [66]. Instead, children needed support and guidance from adults and educators to develop critical data literacy, as such literacy did not tend to simply develop from time spent online [36].

While the privacy implications of the datafication of children, such as sensitive data capture, or data disclosures to third parties have been heavily discussed across the literature of online safety, research relating to how children perceive the algorithmic processes or inferences remains scarce. In fact, only recently have efforts at digital literacy begun to explore AI/algorithmic literacy [49]. However, such efforts were still mainly focused on broader concepts, such as social robotics [9, 41], AI [58] and machine intelligence [22]. Children's perceptions of agents powered by AI algorithms have been shown to be strongly influenced by their previous interactions with them [80]. Previous work also showed that when given sufficient support, children were capable of understanding the role of data in determining machine behaviours [59]. Perhaps the most relevant research would be the line of work around 'critical algorithmic literacy', which focused on helping children to understand the implications of data processing, thus to promote their ability to engage in the critique of algorithmic practices [8, 21, 37].

## 3 STUDY DESIGN

Given our focus on investigating children's perceptions of datafication practices online. We chose the YouTube platform to be used as an example, and conducted one-to-one semi-structured interviews with children to elicit their responses to a collection of tasks that attempted to recreate their everyday experiences on YouTube, followed by a collection of scenarios that reflected different types of datafication practices in relation to how YouTube could process and make use of their data. The reason we selected YouTube as the exemplar datafication platform is because it contains a variety of data processing practices, and is familiar by most children for it being one of the most consumed entertainment platforms among children around the world [27, 63].

### 3.1 Part 1: YouTube Tasks

In part 1 of the interview process, we chose to walk through a series of tasks with participant children on their own devices (Figure 1) for the purpose of recreating and reminding them of their everyday experience on YouTube. In this process, children were encouraged to actively interact with YouTube while being instructed to solve a series of tasks.

Our problem-solving tasks were carefully designed based on 'critical interaction points' on YouTube, and were divided into three sessions: In *Task 1*, children were asked to show the researchers *how they would normally find their favorite videos on YouTube*. Entering the YouTube is a key interaction point and task 1 aimed to observe how children would go to their favorite videos (such as through a subscription list or just through searching function); in *Task 2*, children were asked to

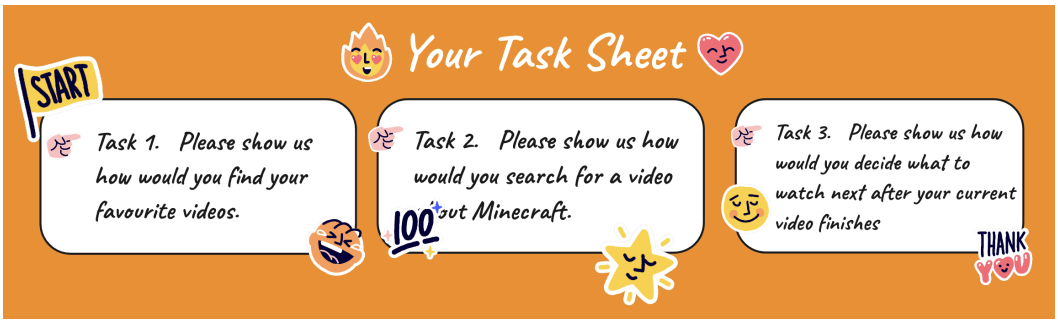



Fig. 1. Children were instructed to complete three tasks on YouTube.

show the researchers *how they would find a video on a specific topic*. We consider the search function as a 'critical interaction point' as it's the most important functionality on the YouTube platform in order to find a video to watch. Here task 2 aimed to observe how children made use of the search function and subsequently, how children would choose and decide which video to click on from the search list; in *Task 3*, children were asked to wait for a current video to finish, and then show the researchers *how they would decide what to watch next*. A few 'critical interaction points' were involved in this task, including the 'next-up' videos that show up after a current video is finished, the 'videos on the right' which includes a list of videos recommended to children based on different categories of information (e.g. the genre of the current video watched, other videos from the same YouTuber, videos based on the watching history), as well as the personalised advertisements that show up at the beginning of the next video. Task 3 aimed to observe how children would normally react to such choices and examine their perceptions on such recommendation practices.

As children were going through the three tasks, they were also invited to explain their understanding about and experiences of certain technological terminologies (*homepage, autocomplete, autoplay*, and *personalised adverts*), if they appeared during their tasks, to evoke further discussions around their perceptions on the relevant data practices. For example, when adverts appeared during their completion of the task, we asked about their perceptions on personalised adverts and how they dealt with them. During the process, we tried to use the tasks to encourage children to recall their experiences as much as possible. Instead of focusing on discussing what were returned, we particularly focused on asking children questions about *how would you* or *what do you do then*.

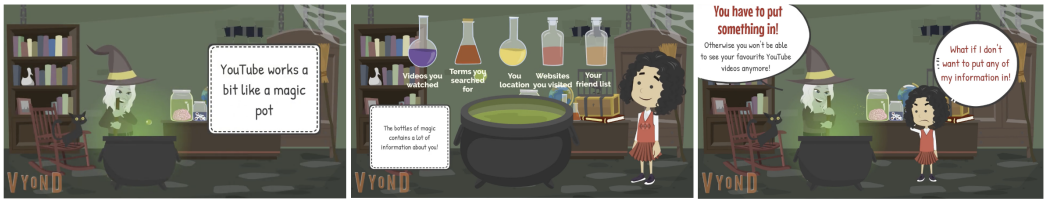
### 3.2 Part 2: Video Scenarios

In the second part of the interview, we chose to present the children with two videos about a fictional character named Lola, a 10-year-old girl who likes to watch videos on YouTube, and how she learns about data collection and processing on YouTube. In the videos, we used metaphors and compared the collection and processing of children's data against throwing bottles of elixir into a magic pot. Previous research showed that children as young as 5 can start to comprehend metaphors and could provide verbal explanations for metaphorical expressions [71], and that metaphors and stories are effective way of building children's understanding of abstract concepts [10, 15].

Each video lasts about 1.5 minutes and was shared with participant children through a link to each video. After watching each video by themselves, children were invited to comment on specific plots presented in the video. A screen capture as well as the question were presented to the children through screen share by the researchers so that children can recall the content, and the questions were expressed in a language appropriate for the participant's age and development. We

have not intended to design or introduce the videos for education or learning purposes because the study was relatively short, lasting about 1 hour. Instead, the videos were mainly there for invoking discussions.

**Video 1: General Perception of Datafication Online**




**Video 2: Perception of Data Inference Online**

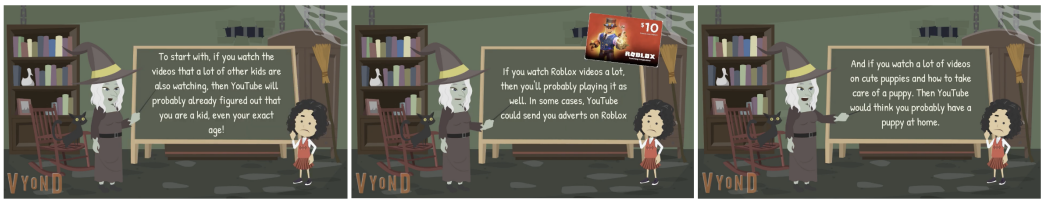



Fig. 2. Screenshots of video 1 and 2. Video 1 is about children's general perceptions of datafication online. Video 2 is about children's perception of data inference online.

Video 1 - *General Perception of Datafication Online.* This video pictures YouTube as 'a magic pot' that requires access to a range of data in order to perform its magic, including video we watched, terms we searched for, websites we visited, our friend lists and our location. After having watched the video, children were prompted to articulate on how they perceive the general datafication practices on YouTube. Specifically, whether they were surprised by or happy about the data being collected by YouTube, and what do they think will happen to their data.

Video 2 - *Perception of Data Inference Online.* This video provides more details about *what* magic can be performed by YouTube and *how*. In this way, we talked about the data inferences performed by YouTube, using data collected by themselves as well as those from other platforms (such as web search history on other sites). Then, we invited children to think about and articulate on the data inference practices: specifically, whether they were surprised by or happy about how YouTube use different types of data about them to recommend new videos, learn more about their personal life, or send them adverts more personalised to their interests. In this way, we focused on the data inference part of the datafication process - online platforms could learn more subtle things about users (apart from interests) based on their data. This refers to the likelihood that YouTube could infer about a child on a more personal level, such as inferring their age, their socio-economic status, and more.

## 4 STUDY METHOD

Participants were recruited from local primary and secondary schools, and a public forum for recruiting family participants. Recruitment started in November 2021 after obtaining ethics approval, and 48 children were interviewed between November and December 2021. Participants were given the choice to take the study either online or in-person due to Covid restrictions. 46 children were interviewed online and 2 were interviewed in-person. At each of these sessions, parents

helped children to set up their devices and then left the children alone until the interview was completed. The majority of the participants were recruited through school newsletters. Each study was facilitated by at least two researchers.

### 4.1 Participant Information

We had 48 participant children, including 23 boys and 25 girls. 24 were among the age range of 7 to 10, another 24 were among the age range of 11-13, with an average age of 10 (range = 7-13, s.d. = 1.76). Details about participants can be found in Table 1.

| Age | #Boys | #Girls | #Total |
|---|---|---|---|
| 7-yo | 2 | 2 | 4 |
| 8-yo | 3 | 1 | 4 |
| 9-yo | 5 | 5 | 10 |
| 10-yo | 2 | 4 | 6 |
| 11-yo | 4 | 6 | 10 |
| 12-yo | 5 | 6 | 11 |
| 13-yo | 1 | 2 | 3 |

Table 1. Summary of participants' ages and genders

### 4.2 Study Process

Each interview contained four parts, including an introduction session, Part 1 - a walk-through of tasks on YouTube, Part 2 - a walk-through of videos on data practice scenarios, and finally an open-ended session about children's thoughts and needs as well as any issues not so far discussed. The whole study was planned to last around to approximate 1 hour.

Children were asked to bring their most used device when accessing YouTube to maximally recreate their everyday experience. In Part 1 of the study, children were given some time to read through the task sheet given. One of the researcher read along the tasks with them, and helped them throughout the process. As they were completing the tasks, researchers observed their activities and asked questions to invoke discussions. Children were told there are no "right answers" to whichever task they were on or question asked. In Part 2 of the study, children were shown with two videos on datafication practices scenarios. One of the researcher watched the videos along with them, and offered clarifications if children had trouble. After they finished watching each video, screenshots of the video were shown along with questions, to remind children of the content of the video. While children were first asked to respond to the predefined questions, like *Are you happy to share some information than others*, our researchers followed up any responses that required further clarification and we encouraged children to share their personal experience or values related to the scenarios, by asking questions like *has it surprised you* and *what would you do then*.

Screen and audio recordings were taken during the studies. Screen recordings on children completing tasks on YouTube were played back during the transcription and data analysis phases to highlight any notable patterns.

## 5 DATA ANALYSIS METHOD

We transcribed the interviews and analysed the data using a grounded, thematic approach [12] to develop codes and themes related to data processing practices reported and discussed by the children.

Results from part 1 of the study contained both children's experiential descriptions about their own experiences of using YouTube, invoked by the tasks, and their perceptions of key tasks. So, we identified the experiential reflections mentioned by the children, and then carefully coded the specific synchronic elements mentioned (such as their emotions, sequence of actions, and information used to inform their action) by the children during their completion of each specific task. In this way, we could gain a more in-depth understanding about how participant children currently managed critical interactions points on YouTube and what elements of knowledge were used by them. This gave us a set of codes about children's usage patterns, their existing knowledge points, and their general expression of experiences about YouTube.

With respect to the data from the video-guided interviews, we tried to calibrate how children perceived YouTube's data processing practices, from the general datafication to the more specific data inference practices, by carefully examining their use of language, for example, how they described data, those who collected their data, what and how the data was processed etc.

The thematic coding process started by dividing the transcriptions into two equal-sized sets. The first two authors independently analysed the first set of transcriptions to derive an initial set of codes. They then met to consolidate and reconciled codes into a common codebook, with a Cohen's kappa of 0.78. The first author then completed coding the remaining half of the transcriptions using this common codebook.

## 6 RESULTS

We present our results by first outlining children's overall experience and usage of YouTube. We then present children's general perceptions of datafication online, followed by an in-depth analysis regarding children's perceptions of the more specific data inference practices. Finally, we present our analysis on children's coping strategies and the design changes they want. While our participant children demonstrated different perceptions and varied level of understanding, we found no strong differences between children of different age or gender. We present individual children's quotes with their participant id. For information on their age and gender, please see Table 2 in Appendix.

### 6.1 Children's Overall Experience and Usage on YouTube

Most children (37/48) in our study owned their own devices (phone, tablet, computer), while the rest used their parents' devices or shared with their siblings. A fair amount of children (19/48) had and would sign in to their own YouTube accounts [1], 7 children used their parents' YouTube accounts, while the remaining 22 children did not have their own account and would not sign in to an account at all on YouTube. In terms of usage, around a third of children in our study spent over 5 hours per week on YouTube, 11 others spent 3 to 5 hours per week, and the remaining 21 children spent 1 to 3 hours per week on YouTube. Children in our study were generally attracted to game videos, animations and educational videos on YouTube.

In general, most children (41/48) reported that they started their use of YouTube using the search button. They generally found 'autocomplete' quite useful for their search, and a few children (17/48) reported that they thought their search intentions could be influenced by the autocomplete list. However, the majority of them (41/48) reported that they were able to decide which search terms to use. The autocomplete could be helpful to remind them of something they searched and they generally thought themselves unlikely to be 'persuaded' by these terms.

Once the play of a video was completed, the majority of the children (35/48) reported that they knew what autoplay is, they commonly experienced this and they demonstrated different ways to respond to autoplay. The majority of the children (29/48) said they normally wouldn't pay any

[1]The minimum age to have an account is 13 years-old, as claimed by YouTube [82].

attention to the autoplay function, thus let the video to automatically go to the next one ('I think I won't do anything' - P20). Some children (23/48) said that they would only follow the videos in autoplay if they were from the same series that they were following ('Well if I'm watching a series then yes, cause the next video would be from that series' - P29); while others (25/48) said that they were more likely to choose what to watch next from the list on the right, instead of permitting autoplay ('I would just pick one from this list' - P4). However, only a few children (11/48) knew how to turn off autoplay.

A fair amount of children (19/48) had their own YouTube accounts and all of them reported they knew or have used subscriptions. Almost all children (46/48) reported that they knew what advertisements are and they generally disliked advertisements on YouTube, mainly because the advertisements being 'not interesting', 'not useful', and 'unskippable'.

In general, most children (38/48) reported having a positive experience on YouTube, for YouTube 'offering fun videos to watch', and 'I can learn stuffs on YouTube'. The remaining 10 children said they had a neutral experience on YouTube, mainly because of 'boring advertisements', and some reported seeing 'weird stuffs' - inappropriate contents on YouTube before. None reported their experience on YouTube as negative.

### 6.2 Children's General Perceptions of Datafication Online

To start with, we present our findings on children's general perceptions of datafication practices on YouTube. Specifically, how children perceive what data is collected and what happens to these data.

*6.2.1 Data Collected: Essential? Personal? And Owned by Whom?* All children knew that YouTube would collect data from them, however, they had different perceptions in terms of what data was being collected, and such perceptions were often related to whether they think such data was essential or personal.

To start with, all children we interviewed knew that data such as the videos they watched before, and terms they searched would be collected by YouTube. And again, all of them were perfectly fine with YouTube doing so, because they thought these data were *essential* for YouTube to generate videos for them:

> I'm not surprised by that (YouTube collecting videos watched and terms searched) at all, and I'm fine with it. Of course they would need that to provide videos for us. I don't see how else they could do that. (P10)

On the other hand, children were more reluctant for YouTube to collect data they think that were 'personal to them', and many did not think YouTube was already collecting such data as they could not see why their personal information such as their location information, their age and gender, would be needed for generating videos:

> Definitely not the location. I feel like the location is personal to you, and honestly I don't know why they would need that. (P3)

There was a strong theme of discussion around what data children considered as their 'personal data'. Interestingly, while almost all children strongly believed that data such as their location information and their age and gender was their own personal data, a majority of children regarded their behavioral data (e.g. videos watched, terms searched, channels subscribed) as 'not personal to me' and therefore could be accessed and quantified by platforms:

> Personal data is something that is not already on their platform, like my location. But I don't think the videos watched and terms are that personal to me, as they are on their platform already. (P16)

Furthermore, we found an interesting common thinking among children regarding who owns their behavioral data online (e.g. videos watched, terms searched, channels subscribed). A surprising

amount of children (32/48) thought that it was YouTube (or the platform) that own these data, or have the rights to these data:

> YouTube should have the ownership. Because YouTube are providing these videos? They should have the right to collect these things I guess, I mean, these are like their thing. (P37)

> I don't know. It feels like I kind of technically own them. But it's kind of like YouTube's right to know and control them? Because we are on their platform? (P21)

In response to our question about "whether you are happy to share these information" or "Are you surprised", the majority of the children were happy to share data they regarded as essential for YouTube to function, but less happy about those they regarded as 'personal' or 'non-essential for YouTube':

> I'm happy to share videos I watched, terms I searched before, channels I subscribed. But definitely not my location, that is personal to me. (P14)

> The others are all fine cause I see why YouTube would need that. Location? No. I can't really think of how that would help them. (P31)

*6.2.2 Data as Part of a Process: (Only) Used to Provide Better Services for Me.* The majority of the children perceived that the data were used by YouTube to provide services for them, within YouTube:

> I knew they(recommended videos) are based on what I watched before and I think it's pretty smart of them to do so. And I always get videos I like. (P34)

Meanwhile, children also demonstrated this perception about the exchange of data for services on other online platforms. For example, some children mentioned that they thought Amazon would collect data such as their purchase history and search terms to recommend products to them; and some children also mentioned their experience on apps stores and games stores, that they thought platforms would collect what apps or games they previously downloaded to personalise the things recommended to them:

> It's just like Amazon right? They will know what stuffs I searched or bought, and then send me similar products. (P31)

> This is a suspicion, but I think Steam is doing that as well? I remember seeing somewhere that they would use stuffs like what I played before to recommend more games for me? (P48)

On the other hand though, while children had good understanding on how YouTube, and other online platforms would provide service for them using their data, most children thought that the *only* reason why online platforms would collect and process their data was *just for* offering them better services. Only very few children (7/48) mentioned data could be used by the platforms for monetisation. However, their understanding of data monetisation were still limited to - platforms will use their data to provide better service for them/users, so that websites will have more users/engagement, thus making more money:

> If they offer better service, they would attract more users, or longer watch time, then they will make money. (P6)

> The more people they have on their website, they could show that to their sponsors, and the companies would pay YouTube money to put up more advertisements. (P41)

And when children thought the platforms were just trying to offer them better services, they were generally fine with and perceived such monetisation as fair, because they were receiving better services:

> It's pretty fair. They are just trying to give you a better experience, and they are doing a pretty good job. I think they should earn some money. (P29)

> I don't really mind if they use my data to make money. It's a win-win situation actually. At least it's useful to me. (P11)

*6.2.3 Data Silos: My Data will Only Stay on Where I Go.* A key theme emerged was children's perception about where these data about them went: the majority of them believed that the data were collected by a certain platform and would be only used by this platform locally.

> They collect your data on YouTube, and used that on YouTube. I think it's pretty obvious? I don't think the others (other platforms) could use YouTube's data. (P29)

Some children showed some awareness about data being shared across platforms but only very few of them (5/48) were able to articular how such data sharing might be carried out:

> Maybe if you have YouTube tab open, and you change that to a different website. They might be able to track you to the next website. (P46)

Interestingly though, a fair proportion of children (22/48) thought data sharing/tracking was enabled through Google:

> Because YouTube is owned by Google, and YouTube and Google could see whatever you do on each. But if you have a different search engine then YouTube won't be able to track you. (P25)

Even for the few children who had some notion about the general data sharing practices, they struggled to picture how the data flow took place. And they commonly thought that their data would only be transferred among the websites they actually visited:

> I think it (my data) would only go into all the websites that I usually use. (P16)

> It's like footprints. I will only have footprints in places I go to, but not beyond them, like other places. (P28)

It was perhaps thus not surprising to see that a fair amount of children (19/48) thought that they could just delete the data they left on the Internet by simply uninstalling the apps they usually use:

> I guess I could take back all my information if I just delete the apps? (P11)

### 6.3 Children's Perceptions of Data Inference Online

Following children's perceptions of the general datafication practices online, we further investigated their understanding and perceptions of the data inference practices online.

*6.3.1 What is Data Inference.* Continuing the notion about data as being an element and part of a process, most children described data inference as the process in which YouTube is trying to makes guesses and assumptions about people:

> YouTube does not actually know any facts about you. So it has to guess about your age, what kind of school you go to, any news in your home. (P45)

Some children thought that data inference also included categorising people into certain groups based on what is inferred about them:

> It assumes about you, and put you together with the majority. (P48)

> They would categorise you in different groups, so that they could do things in each different groups. (P19)

While most children described data inference as 'making guesses' and 'broad assumptions' about them, there were a few children believed that data inference would enable YouTube (and other platforms) to know exact factual details about them:

> They will try to learn everything about you. And with all the information they have, they could even somehow know my name, and everything about me, what jobs my parents do etc. I don't want this but I think it's unlikely for them to get things wrong, with all that power they have. (P14)

*6.3.2 How is Data Inference Conducted.* Children showed different understanding in terms of how a data inference process was conducted. We summarised their understanding into four major themes: *operators* - who/what was conducting the data inference, *inputs* - what was the data inference based on, *algorithmic processing* - what was the (computational) process happen in between, and *outputs* - what were the outputs of data inference.

Starting with the *operators*, the majority of the children reported that they believed that it was not 'human' that 'manually' processed their data. They would describe the operators as some kind of 'machine' using terms such as 'a computer', 'a robot', 'an AI' or '010101'. On the other hand, only a minority of children (11/48) had some awareness that algorithms were used to process their data:

> I doubt it if just some person just sitting there doing it. So it's probably some kind of algorithm. (P7)

For these children, they showed varied perceptions about what an algorithm was. Some of those believed an algorithm was written by human, and followed a set of rules:

> A person wrote the algorithm, like they told the computer what to do. An algorithm just do what it's coded to do. (P24)

Interestingly, when trying to explain how algorithms follow a set of rules, many of these children did not recognised the 'data-driven' aspect of algorithm, but instead described algorithms as following very exact rules. For example, when asked about why they think algorithms could generate gender-stereotypical contents for people, they tend to think that algorithms were set to only give certain content to certain gender:

> I think it (YouTube algorithm) might be doing that intentionally, like it's set to only give certain contents to certain groups of people. (P3)

That said, many children described there was data *input* for this machine/algorithm, and the algorithms were closely based on what data they knew: they generally thought the data collected on YouTube (e.g. videos watched, terms searched) could be used to make inference about them:

> I mean if you only watch yr7 videos on YouTube, they would probably guess that you are a kid. (P27)

> Depending on what kind of videos you watched? Like if you watch videos in another language, it's pretty obvious to them that you speak that language. (P29)

On the other hand, many of these children reported that they thought only their data resulting from their interactions on a particular platform would be used to make inference about them. To be more precise, almost all children only mentioned the possibility of YouTube using data generated from their interactions on YouTube (e.g. videos watched); none, meanwhile, mentioned the possibility of YouTube using other data sources, such as what they did on other websites. This agrees with previous findings from studies with both adults and children, that individuals perceive data in silos on each platform separately and have little awareness of cross-platform tracking and information sharing:

> I don't really think that's (data inference based on cross-platform data) possible? Well I guess it's possible if they want to. Emm, I never really thought about this... (P34)

Finally, as for the *outcome* of data inference, many children mentioned the concept of a 'profile'. And they described such profiles as a list of things, or records about them:

> I guess they would have a profile about you, like what type of person they think you are. It's different from what they think you might be interested in. I don't know, it's hard to describe. (P18)

*6.3.3 How Children Perceive Data Inference Practices - When Reflecting on Hypothetical Scenarios.* During our interview, we used two hypothetical scenarios to establish a deeper conversation with

children about their perceptions around how platforms could make inference about them. The first scenario illustrated how YouTube would send personalised ads on things of different prices to different children; and the second one showed how YouTube would send advertisements on Barbie Dolls and makeups to girls, and send advertisements on science books and robotic toys to boys. When reflecting on these scenarios, children showed strong oppositions against platforms conducting data inference on them, for a variety of reasons.

To start with, children had strong opinions and generally believed that platforms shouldn't make assumptions of people, arguing that such practices could be used against them or people like them:

> Don't make assumptions about me! They're just kind of saying that some people should spend more money than others. (P19)

> This is just stereotypical. They are trying categorise people by their gender and make guesses on them. But those guesses could be wrong. (P48)

Children further described such data inference practices spying on them without them knowing:

> It's rude. It's just rude. There are lines. In way, what YouTube does is technically spying on your personal life. (P19)

> It's like you go into someone's house without them knowing, take stuff away from them. YouTube is just doing that. (P18)

They then continued to argue there exists an unequal relationship between them and YouTube. When being asked about what they thought about the trade-off between YouTube trying to learn so much about them, while at the same time, YouTube was offering them videos to watch, they thought what YouTube took from them was more than what they gained on the platform:

> It's just unfair. The videos they give us might not be useful to us, but the things they learn from us is useful to them. (P40)

Some children continued to argue that YouTube's only job should be providing videos, and it didn't need to learn so much about them:

> I'm just here for the videos. Their only job is to generate videos. It's completely unnecessary for them to know anything about me. Well maybe just what videos I watch, but that's it. (P23)

*6.3.4 How Children Perceive Data Inference Practices - When Reflecting on Their Own Experience.* However, when being taken outside the scenarios and asked to reflect on their own experience. It was interesting to see that, the exact same children who demonstrated different feelings and values, as well as critical thinking before (under the hypothetical scenarios), were unable to relate such thinking to their own everyday experience on YouTube. For example, many of them did not think or weren't aware that data inference practices could happen in their own lives:

> I don't think they are actually trying to do that though, cause I don't really see any stereotypical content. (P28)

> I actually never thought about that. Now that I think about it, they could be doing that? It's pretty scary. (P46)

And when being further asked about why YouTube, or more generally, platforms would try to infer about them and learn things about them, the majority of the children thought this was dominantly used for "making the website more popular" or "offering me better videos to watch", even though they demonstrated different feelings when they were shown how different 'values' could be generated by the inference process:

> They try to learn things about users because they want to give them better services. And if you offer better services, your website would become more popular. (P2)

> It's because so that they will know what's new in your life and offer you better videos to watch. (P35)

And when reflecting on their own experiences, many children reported that they didn't care as much because they didn't feel such data inference, if any, would make any difference or affect them in their real lives:

> To be honest, I don't really care. It's not gonna influence me in any way even if they learn things about me. (P40)

On the other hand, for the children who had some real life experience in terms of how data inference could have impacts on them, they were more alert of such practices:

> I definitely don't want them to know about me. I remember my parents mentioned that there was this time they tried to refurnish the house, maybe they have searched for something on that? And then they started to get calls from agencies, like they know they own a house and asking whether they want to rent out the house. I can't remember exactly but it's really creepy. (P43)

### 6.4 Coping Strategies and Desire for Legibility and Control

When asked about how they managed their data, the children revealed several coping strategies, for dealing with the undesirable aspects of these datafication practices, which revealed a number of common concerns and perspectives among children.

*6.4.1 Don't Know How to Cope.* During our interviews, children mentioned several coping strategies that pertained to their data privacy, specifically around how to stop data collection from websites and how to better protect their personal information online. These strategies included turning off location tracking, and adjusting privacy settings.

However, there was little mentioned (or that children could do) to help them address how their data were used by the algorithms. They expressed confusion and frustration at their inability to stop, or defend themselves more directly against certain data processing practices, describing them as 'more subtle movements':

> So like at school, we've already definitely learnt not to just randomly give out your age, your personal details and stuff. But with these subtle movements from YouTube and probably other companies, I don't know what to do. (P34)

Some children mentioned they tried to read up more carefully on the consent agreements and terms of service, but found it frustrating that data processing practices weren't made clear in either, and that such practices were impossible to manage or control from their end:

> They don't really talk about these things you know, well, at least they weren't very explicit. (P9)

> It's just impossible for me to deal with it. What can I do? Maybe I can hide some data from them, like my location. But I can't control what happens after they got the data. (P14)

*6.4.2 Desire for Legibility and Control.* On the other hand, when discussing future mechanisms, almost all children (46/48) mentioned that they wanted to be able to stop their data from being processed and used in ways that were unknown to them, mainly through two types of mechanisms - transparency and control.

To start with transparency, almost all children (41/48), regardless of their age, talked about how they would like to have information regarding the data processing practices: not only about what data of them is being collected, but more importantly, how their data will be processed. For example, some children described a transparency mechanism in which they would be able to see how a certain video or advertisement was selected for them:

> Cause it just feels more comfortable knowing what they're really doing. Like how they generate this video, is it based on something I purchase on another website? Maybe they're doing something extra they don't tell you. (P32)

Some other children mentioned they wanted to be able to see if YouTube had a profile about them, and what YouTube put down about them in the profile:

> Yes I do want to see what they think about me. Probably be happier if they get things wrong though. I'd rather they didn't know everything about me. (P31)

Meanwhile, children also urged for more control mechanisms: more specifically around how they could more directly control what platforms could infer about them, and how platforms would conduct this inference process:

> I want to have like a settings or something like that, that I can adjust what they could learn about me. Maybe adjust what they could use to learn as well. (P5)

Some children just said they wanted all platforms to stop data inference from the beginning, and instead only focus on 'their job':

> To be honest, I just want them to do their job and nothing else. Just stop learning things about people cause that's not their job. (P28)

*6.4.3 Pessimism About Platform Change.* However, nearly all of the children (45/48) we interviewed expressed pessimism that YouTube, and other online service platforms in general, would change anything about their current practices. Several children discussed that they thought nothing would change, as such datafication and inference online had become such established and ingrained, and it would be hard to change unless there were regulations on it:

> If there's a law on it then yes. But if no they'll just carry on doing it. (P26)

And a surprising amount of children (37/48) thought no changes would be made, for the reason that companies would only care more about money:

> Well, I guess it impossible. Because they just want to make more money, to make their business bigger, and to get wealthier. They just care more about money. (P7)

## 7 DISCUSSION

### 7.1 Key Findings and Contributions

Children are often regarded as not as capable or competent as adults for coping with the complexities of online life, including aspects of privacy, safety, and datafication [47], viewed as 'diminished versions of adults' [66]. However, our findings showed that not only do children care significantly about various aspects of datafication, but they already possess rudimentary conceptual understandings of it, and a significant willingness to take action to shape it to their desires, possibly even more than adults. We found that children were well aware of the data collection practices on YouTube, especially on data such as the videos they watched before, and terms they searched. They largely knew that these collected data will be processed, and generally understood that such processing is not conducted by human, but through some kind of machine. They had some awareness in terms of online datafication practices would "make guesses on them", and the results of such guesses could be used on users such as themselves, although not necessarily understanding the full picture. On the other hand, our results reinforced existing findings that children do not always comprehend online datafication practices to a full extent [48, 66, 74], and through a knowledge construction lens, we identified three key knowledge gaps in children's perception of data inference practices, including their lack of recognition of *who are involved in the data processing and how*, *data being transmitted across platforms* and *their data ownership*. These findings provide critical contributions regarding our understanding of how 7-13 years old children perceive datafication in their everyday platform; instead of looking broadly at children's understanding of 'what an algorithm is', our research provided deeper insights regarding children's ability to interpret the implications of data processing and their barriers. These findings also provided important future design directions for supporting the development of children's algorithmic literacy and self-autonomy.

### 7.2 Implications for Critical Algorithmic Literacy Development

Although the focus of our research was to explore how children experience and perceive a data-driven algorithmic platform, our findings also provided crucial insights for the recent 'critical algorithmic literacy' (CAL) development, confirming the urgency of extending children's cognitive computational thinking. The CAL framework also provides a useful guidance for us to interpret the observed gaps of knowledge for devising key future design considerations.

There have been a variety of developments looking into how to support children in achieving 'digital literacy' [17, 26, 35, 77]. However, such frameworks were often oriented around data privacy or online safety, with the algorithmic processing of data of online systems scarcely mentioned. There has been some work from researchers around AI literacy [23, 24, 32, 38, 49]. However, such work generally pertains to AI systems (machine-based systems that could make predictions, recommendations, or decisions that influence real or virtual environments [81]), rather than the more specific data-driven algorithms. On the other hand, 'critical algorithmic literacy' [8, 21] puts particular emphasis on *understanding the implications of data processing*, by directing children's attention towards data and the algorithmic processes applied to them. The goal of CAL is not merely assisting the development of knowledge about algorithms but also an ability to engage in critique of algorithmic systems reflexively.

CAL is closely aligned with recent extension of computational thinking. Kafai et al [37] proposed that computational thinking should include three key frames: the cognitive, the situated, and the critical thinking. *Cognitive thinking* focuses on the understanding of key computational concepts, practices, and perspectives and the associated skill building and competencies; *situated thinking* encourages learning to take place in contexts that the learner cares about so that they include their personal expression and social engagement in their pathway of learning; and finally *critical computational thinking* emphasises the importance of supporting the questioning of larger structures and processes behind the computational phenomenon. Our key findings provided the crucial empirical evidence for the need of situated and critical thinking in children's algorithmic literacy.

To start with, our findings resonates the emphasis that CAL should go beyond the basic *cognitive* understandings and *situated* thinking will complement children's understanding of the social aspects of algorithmic system. For example, although most of our study participants were able to recognise different types of data being collected by YouTube and how they were processed to provide better video recommendations, they struggled to situate these understandings in the complex and diverse social contexts under which data may be collected, processed, analysed or exploited. Supporting children to connect with different algorithmic situations is the key objective of *situated thinking*. Indeed, previous studies have shown how children's understanding of privacy and algorithmic implications could be boosted by participating in carefully designed computational tasks involving the sharing and processing of social media data [20, 31]. We can envisage that similar experiments could be designed to enhance children's situated thinking regarding the sharing of data across platforms and what their data ownership means under different contexts.

Other findings from our study demonstrated the importance of introducing *critical* thinking in algorithmic literacy. For example, many of our participants struggled to perceive their data were shared across different platforms and how these data from outside YouTube could affect their experiences on the platform. They were unable to perceive themselves and their data as part of one connected data-driven digital society, in which an extensive amount of personal data about each individual can be processed in various unanticipated ways. This ability of situating datafication in a broader digital society is rarely discussed in existing research of algorithmic literacy [8, 49, 66]. However, *critical thinking* promotes an understanding of the existing structure of power in order to increase children's awareness of ideologies, privilege, and opportunities, and several research

have shown how the approach can push children to "conceptualise, create, and disseminate digital projects that break silences, expose important truths, and challenge unjust systems." [37]

Situating our findings under CAL also provided tremendous inputs for our design of future digital experiences for children. At the same of time of exploring a constructive application of behaviourally or psychologically grounded approaches to enhance children's self-autonomy and resilience, we must recognise the importance of supporting children's CAL. Existing approaches taken by the CAL community have shown some fruitful results. A good critical computational thinking ability would pave the crucial foundation for children to exercise informed choices in a transparent algorithmic system and chances of exercising their data autonomy.

### 7.3 Implications for Designing Future Digital Experiences for Children

Our understanding of children's current experience of datafication prompted an urgent need for rethinking what future data-driven digital experiences should be like for children, so that we can reduce the negative effects they have on children [40, 61]. Our findings have provided some immediate indications regarding how we could better support children in the datafied society and a need for a fundamental shift of the current data governance structure.

To start with, children's inability to recognise data transitions across platforms and their data ownership indicated how the lack transparency of the current datafication approach is damaging children's development of identity [54, 85] and their ability to effectively link such practices with, let alone effectively recognise and comprehend, data processing and inferences. As a result, we had the majority of the children believing that datafiication was largely for generating better services and that datafication was a very localised phenomenon. Transparency mechanisms have been extensively studied to be brought in as 'hints' for children to remind them of and to help them better posit their decisions during their interaction with the online platforms [69, 79]. However, previous attempts on transparency mechanisms tend to focus on raising children's awareness on certain practices on the online platforms (e.g. their data is being collected). These existing transparency mechanisms were mainly oriented around empowering children with more *cognitive understanding* of key computational concepts through child-friendly ways, such as Lego's Caption Safety [45] and Google's Be Internet Legends [28]. On the other hand, as our findings as well as the CAL framework indicated, we must extend these existing transparency mechanisms to support children developing *situated recognition* in broader and more diverse contexts. A recent design example from the UK ICO has shown how the transparency of implicit data privacy risks could be conveyed to children in a much more effective way if these risks were situated in a variety of scenarios from children's lives that they are more familiar with [34]. Future transparency mechanisms should focus on seeking to promote children's critical thinking, that is to increase their awareness of potential outcomes, ideologies and values associated with datafication practices, instead of the factual hints on the surface.

In fact, the lack of transparency mechanisms could result in children not being able to make informed decisions online. Self-determination theory explains that children's ability to self-regulate requires intrinsic motivation to enact specific behaviors and internalization of norms [30]. Understanding social ideologies and values related to their decisions enables children to exercise executive function, control impulses, and make more informed decisions online. This has been reflected during our studies, that children particularly struggled at some of the critical interaction points on YouTube includes recognising autocomplete when they are conducting searches, personalised advertisements embedded in the videos, autoplay for the next-up videos, or a list of recommended videos personalised for them. Children's discussions showed that they were not always capable of recognising the datafication behind such interaction points: a lot of them would randomly select terms from autocomplete list without fully recognising how each term is generated for them, permit

autoplay without thinking too much about how and why autoplay videos were generated for them, or click on advertisements links without realising that those advertisements were specifically targeting at them.

On the other hand, such behaviours of children resonate the importance of designing for children's digital experience at these critical points [50], which could go beyond transparency, and more generally about enabling children to think a bit more and make more informed decisions online. Several studies have explored how to scaffold children at various critical interaction points. A recent study on YouTube, for example, introduced redesigned mechanisms to offer users with different level of control at points such as search, autoplay, recommendations, thus resulted in users feeling greater sense of agency [50]. Related research with children have shown that changing the designs of how a video finishes would greatly alter how children decide what to do next [33]. Meanwhile, there have been a variety of research on how psychologically-based design mechanisms could be brought in at these critical interactions. Studies have attempted to adopt the use of dual system theory [19] to support users' ability to regulate their digital experience [52, 73], by carefully framing messages to encourage more automated and reflective exercise of self-autonomy. Our study showed that children had a strong desire of gaining more control over how their data was being exploited and processed. Supporting children to exercise better control and autonomy in datafication has not been extensively addressed in existing literature. Our research has confirmed a need for future design in this direction and the importance of exploring how children's existing algorithmic literacy may affect the building of their agency. This resonates with existing literature which has emphasised that the development of children's media and digital literacy can be key to developing their resilience with harmful content and experiences online [61].

### 7.4 Future Data Governance: Towards A More Ethical Data Structure

The current datafied childhood demands a fundamental shift of the current data governance structure. The key contributing factor to the current ubiquitous datafication in our children's lives, as well as our society, is the centralised data monopoly and concentration of power by a number of platform companies. In recent years, policy development for children online has started to put more emphasis on the 'algorithmic-nature' and the 'AI-enabled' online practices for children [3, 6, 46]. While the impact and the potential risks and concerns of datafication have been heavily discussed in recent policy development [2, 16, 29], discussions around how to fundamentally address the problem of online data governance were largely missing. Data governance here refers to the fundamental problem of who should have the right to access and control users' data. Our findings showed that children had strong demands to control the processing of their own data through more direct means. However, under the current online data ecosystems, such demand is almost not achievable. Unlike data privacy management, in which children could still be part of process to determine whether they want or don't want certain data of them to be collected by platforms, datafication and the algorithmic processing of their data is another issue, during which users were not even part of this process, the current data ecosystem does not allow them to have any form of control on how the processing would be conducted once their data is collected and gone to the platforms. We therefore argue there need to be fundamental changes to the current data ecosystem that dominate our information consumption online [25]. We argue that future online service providers should be more ethical when they were accessing and processing people's data, which might require a fundamental revolution to the current data ecosystem that is very much service providers oriented.

There have been several recent movements towards this, and researchers have been actively engaged in the development of new data governance structure. A wave of new decentralised paradigms for data sharing and ownership [5] are being explored to expand individual data subjects' ability to control access to their data, by enabling collective access requests through representative

intermediaries such as NGOs and trade unions, therefore increasing the agency of individual data subjects. However, creating a new data governance structure is not without challenge [5]. Under this new and decentralised paradigm one must have careful thinking about *who* will be responsible for the data sharing, data control and data curation, how to attribute responsibilities if something went wrong, which can range from the compromised quality of the data provided by a data subject, misuse of shared data by the data processor (e.g. data re-sharing or re-identification), as well as how to help users adapt to this new data governance structure. Our research have confirmed a need from children to expand their ability to access and control their data, and a fundamental change in the data governance structure could bring new opportunities and challenges. We encourage future research making further explorations regarding how to facilitate the creation and uptake of this new data governance paradigm, so that users can regain their data rights and we can build towards a more ethical data architecture for the future society.

## 8 LIMITATIONS AND FUTURE WORK

There are several important limitations of this work; the first pertains to sample size of study population. First, the parents who signed up for their children may already be more interested and cautious about datafication practices than the average population. In fact, we have noticed that some children mentioned their parents have been having conversations with them about their data online, and some have reported that their parents would configure online privacy settings for them.

Second, while we did not collect information about participants' family income, the families' areas of residences were centred around the largest city in the UK and area near a university city. This may have resulted in our findings reflecting a greater digital literacy than the general population, not only due to familial influences, but also due to local schools. Our findings showed that age is not a critical factor in children's ability in understanding and describing data practices. Instead, although we did not formally collect this information from the participants, we noticed from our conversations with children that family and school education greatly influenced children's digital literacy. For the children who mentioned having conversations with their parents, or have learnt about data-related issues at school, they were generally more capable of describing datafication practices and showed deeper understanding.

Third, we based our analysis on children's self-report data, and children may have moderated their responses according how the questions were asked or due to the experimental setting. For example, we can see how the choice of metaphors or the use of languages may affect children's connections to the phenomenons to certain extent, and the choice of our stances could affect how participants look upon datafication. Nevertheless, the focus of our paper is not assessing children's experience with the platforms or their digital well-being. Our focus is to assess children's perceptions of datafication: what they can recognise and what they struggle to recognise. Meanwhile, we attempted to mitigate this through several ways, one was by making sure that children knew there are no "right answers" to whichever task they were on or question asked, and by asking questions in a language appropriate for the participant's age and development. We also made sure that children were not 'nudged' into any answers or understanding by choosing our wording carefully, and attempted to design our questions to avoid pre-established assumptions.

The reflections on the hypothetical scenarios we devised were necessarily reactions to simplified fictions of what would happen in the real-world, and thus were limited in terms of ecological validity. For instance, we chose our example in the scenario to exhibit representational algorithmic (gender and socio-economic) bias, as this is still very common in online recommendation systems. However, this may have focused participants on this particular aspect of that scenario.

Finally, datafication is a broad topic related to a variety of different issues - ranging from monetisation, behavioral manipulation, belief shaping, and more. In this study, we were only able to

address a few of the topics relating to more immediately relevant aspects of the YouTube experience. Although we chose YouTube because of the important role it plays in kids' online experiences, our choice of YouTube is likely to have shaped our results relating to datafication significantly to those around those issues most relevant to this app; studies using different apps (e.g. games or educational software) may yield additional perspectives and issues complementary to the ones we identified.

Future work aims to explore how may we design approaches that could enable children to expand their knowledge about datafication practices, and develop ways to support children's critical algorithmic literacy development. We intend to run co-design workshops [44, 56] with children as follow-up studies, which may include transparency and autonomy-supportive design mechanisms, and explore approaches to help children better navigate online datafication practices. Meanwhile, we aim to develop child-friendly prototypes based on some new data governance structures, and explore ways for such data structure to become more useful for children.

## 9 CONCLUSION

As children are growing up in an age of datafication, children's data are now being routinely used to profile, analyse and make predictions of them. Their actions online are not only recorded, tracked, aggregated, but also analysed and monetised. Such practices are hard to understand even by adult users, let alone children. This paper is the first to contribute an understanding of the perceptions of children on the datafication and the more specific data inference practices that dominates their information consumption online. We identified three key knowledge gaps in children's perceptions, including their lack of recognition of who are involved in the data processing and how, data being transmitted across platforms, and their data ownership. Our findings provided crucial insights for the recent critical algorithmic literacy development, confirming the urgency of extending children's cognitive computational thinking. Through situating our findings under critical algorithmic literacy, our findings have provided important immediate indications regarding how we could better support children in the datafied society through more transparency and autonomy-supportive designs, as well as identified the need for a fundamental shift of the current data governance structure. We hope that our findings will support both designers for children, as well as those of educational material seeking to address gaps in their understanding of datafication practices online, and finally, lay down the foundation for a more ethical data governance structure in the future.

## REFERENCES

[1] 1998. Children's Online Privacy Protection Rule ("COPPA"). https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule

[2] 2020. Age appropriate design: a code of practice for online services. https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services/

[3] 2020. Guidance for Regulation of Artificial Intelligence Applications. https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

[4] 2020. What is automated individual decision-making and profiling? https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/

[5] 2021. Ethical Web And Data Architectures. https://www.oxfordmartin.ox.ac.uk/ethical-web-and-data-architectures/

[6] 2021. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206

[7] Amelia Acker and Leanne Bowler. 2017. What is your Data Silhouette? Raising teen awareness of their data traces in social media. In *Proceedings of the 8th International Conference on Social Media & Society*. 1–5.

[8] Ezequiel Aleman, Larysa Nadolny, Alejandro Ferreira, Bruno Gabetti, Guillermo Ortíz, and Martín Zanoniani. 2021. Screening Bot: a Playground for Critical Algorithmic Literacy Engagement with Youth. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*. 198–202.

[9] Alex Barco, Chiara de Jong, Jochen Peter, Rinaldo Kühne, and Caroline L. van Straten. 2020. Robot Morphology and Children's Perception of Social Robots: An Exploratory Study. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 125–127. https://doi.org/10.1145/3371382.3378348

[10] Richard M Billow. 1981. Observing spontaneous metaphor in children. *Journal of Experimental Child Psychology* 31, 3 (1981), 430–445.

[11] Leanne Bowler, Amelia Acker, Wei Jeng, and Yu Chi. 2017. "It lives all around us": Aspects of data literacy in teen's lives. *Proceedings of the association for information science and technology* 54, 1 (2017), 27–35.

[12] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[13] Moritz Büchi, Eduard Fosch-Villaronga, Christoph Lutz, Aurelia Tamò-Larrieux, and Shruthi Velidi. 2021. Making sense of algorithmic profiling: user perceptions on Facebook. *Information, Communication & Society* (2021), 1–17.

[14] Moritz Büchi, Eduard Fosch-Villaronga, Christoph Lutz, Aurelia Tamò-Larrieux, Shruthi Velidi, and Salome Viljoen. 2020. The chilling effects of algorithmic profiling: Mapping the issues. *Computer law & security review* 36 (2020), 105367.

[15] Lynne Cameron. 1996. Discourse context and the development of metaphor in children. *Current Issues in Language & Society* 3, 1 (1996), 49–64.

[16] European Commision. 2021. General Data Protection Regulation (GDPR). https://gdpr-info.eu/

[17] European Commission. 2016. The European Digital Competence Framework for Citizens.

[18] Nick Couldry and Ulises A Mejias. 2019. *The costs of connection.* Stanford University Press.

[19] Fiery Cushman. 2013. Action, outcome, and value: A dual-system framework for morality. *Personality and social psychology review* 17, 3 (2013), 273–292.

[20] 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*. 3620–3631.

[21] Sayamindu Dasgupta and Benjamin Mako Hill. 2020. Designing for Critical Algorithmic Literacies. *arXiv preprint arXiv:2008.01719* (2020).

[22] Stefania Druga and Amy J Ko. 2021. How Do Children's Perceptions of Machine Intelligence Change When Training and Coding Smart Programs?. In *Interaction Design and Children* (Athens, Greece) *(IDC '21)*. Association for Computing Machinery, New York, NY, USA, 49–61. https://doi.org/10.1145/3459990.3460712

[23] Stefania Druga, Sarah T Vu, Eesh Likhith, and Tammy Qiu. 2019. Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019*. 104–111.

[24] Stefania Druga, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2018. How smart are the smart toys? Children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. 231–240.

[25] Catherine D'Ignazio and Rahul Bhargava. 2015. Approaches to building big data literacy. In *Proceedings of the Bloomberg data for good exchange conference*.

[26] UK Council for Internet Safety. 2020. Education for a Connected World – 2020 edition. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/896323/UKCIS_Education_for_a_Connected_World_.pdf

[27] 5 Rights Foundation. 2021. Playful by Design: Free play in a digital world. https://digitalfuturescommission.org.uk/wp-content/uploads/2021/11/A-Vision-of-Free-Play-in-a-Digital-World.pdf

[28] Google. 2021. Be Internet Legends. https://beinternetlegends.withgoogle.com/en_ie

[29] HM Government. 2020. Online Harms White Paper.

[30] Wendy S Grolnick, Edward L Deci, and Richard M Ryan. 1997. Internalization within the family: The self-determination theory perspective. *Parenting and children's internalization of values: A handbook of contemporary theory* (1997), 135–161.

[31] Samantha Hautea, Sayamindu Dasgupta, and Benjamin Mako Hill. 2017. Youth perspectives on critical data literacies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 919–930.

[32] Clint Andrew Heinze, Janet Haase, and Helen Higgins. 2010. An action research report from a multi-year approach to teaching artificial intelligence at the k-6 level. In *First AAAI Symposium on Educational Advances in Artificial Intelligence*.

[33] Alexis Hiniker, Sharon S Heung, Sungsoo Ray Hong, and Julie A Kientz. 2018. Coco's Videos: An Empirical Investigation of Video-Player Design Features and Children's Media Use. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 254.

[34] ICO. 2021. Designing data transparency for children. https://ico.org.uk/media/2620177/designing-data-transparency-for-children.pdf

[35] DQ Institute. 2021. Digital Intellegence Framework. https://www.dqinstitute.org/global-standards/

[36] Engin Isin and Evelyn Ruppert. 2020. *Being digital citizens*. Rowman & Littlefield Publishers.

[37] Yasmin Kafai, Chris Proctor, and Debora Lui. 2020. From theory bias to theory dialogue: embracing cognitive, situated, and critical framings of computational thinking in K-12 CS education. *ACM Inroads* 11, 1 (2020), 44–53.

[38] Peter H Kahn, Batya Friedman, Deanne R Perez-Granados, and Nathan G Freier. 2006. Robotic pets in the lives of preschool children. *Interaction Studies* 7, 3 (2006), 405–436.

[39] Gabriella Kazai, Iskander Yusof, and Daoud Clarke. 2016. Personalised news and blog recommendations based on user location, Facebook and Twitter user profiling. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 1129–1132.

[40] Baroness Kidron, Alexandra Evans, and Jenny Afia. 2018. Disrupted Chldhood. (2018).

[41] Jacqueline M. Kory-Westlund and Cynthia Breazeal. 2019. Assessing Children's Perceptions and Acceptance of a Social Robot. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19)*. Association for Computing Machinery, New York, NY, USA, 38–50. https://doi.org/10.1145/3311927.3323143

[42] Priya Kumar, Shalmali Milind Naik, Utkarsha Ramesh Devkar, Marshini Chetty, Tamara L. Clegg, and Jessica Vitak. 2017. 'No Telling Passcodes Out Because They're Private': Understanding Children's Mental Models of Privacy and Security Online. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 64 (dec 2017), 21 pages. https://doi.org/10.1145/3134699

[43] Priya Kumar, Shalmali Milind naik, Utkarsha Ramesh Devkar, Marshini Chetty, Tamara L. Clegg, and Jessica Vitak. 2018. No Telling Passcodes Out Because They're Private?: Understanding Children's Mental Models of Privacy and Security Online. In *Proceedings of ACM Human-Computer Interaction (CSCW '18 Online First)*. ACM.

[44] Priya Kumar, Jessica Vitak, Marshini Chetty, Tamara L Clegg, Jonathan Yang, Brenna McNally, and Elizabeth Bonsignore. 2018. Co-designing online privacy-related games and stories with children. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. ACM, 67–79.

[45] Lego. 2021. Captain Safety. https://www.lego.com/en-gb/life/digital-safety

[46] David Leslie. 2019. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. https://doi.org/10.5281/zenodo.3240529

[47] Sonia Livingstone. 2018. Children: a special case for privacy? *Intermedia* 46, 2 (2018), 18–23.

[48] Sonia Livingstone, Mariya Stoilova, and Rishita Nandagiri. 2019. Children's data and privacy online. *Technology* 58, 2 (2019), 157–65.

[49] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.

[50] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[51] Deborah Lupton and Ben Williamson. 2017. The datafied child: The dataveillance of children and implications for their rights. *New Media & Society* 19, 5 (2017), 780–794.

[52] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools. In *proceedings of the 2019 CHI conference on human factors in computing systems*. 1–18.

[53] Shawna Malvini Redden and Amy K Way. 2017. 'Adults don't understand': exploring how teens use dialectical frameworks to navigate webs of tensions in online life. *Journal of Applied Communication Research* 45, 1 (2017), 21–41.

[54] Giovanna Mascheroni. 2020. Datafied childhoods: Contextualising datafication in everyday life. *Current Sociology* 68, 6 (2020), 798–813.

[55] Ariadna Matamoros-Fernández. 2017. Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society* 20, 6 (2017), 930–946.

[56] Brenna McNally, Priya Kumar, Chelsea Hordatt, Matthew Louis Mauriello, Shalmali Naik, Leyla Norooz, Alazandra Shorter, Evan Golub, and Allison Druin. 2018. Co-designing Mobile Online Safety Applications with Children. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 523.

[57] Ulises A Mejias and Nick Couldry. 2019. Datafication. *Internet Policy Review* 8, 4 (2019).

[58] Gaspar Isaac Melsión, Ilaria Torre, Eva Vidal, and Iolanda Leite. 2021. Using Explainability to Help Children UnderstandGender Bias in AI. In *Interaction Design and Children* (Athens, Greece) *(IDC '21)*. Association for Computing Machinery, New York, NY, USA, 87–99. https://doi.org/10.1145/3459990.3460719

[59] David Mioduser and Sharona T Levy. 2010. Making sense by building sense: Kindergarten children's construction and understanding of adaptive robot behaviors. *International Journal of Computers for Mathematical Learning* 15, 2 (2010), 99–127.

[60] Lilian Mitrou, Miltiadis Kandias, Vasilis Stavrou, and Dimitris Gritzalis. 2014. Social media profiling: A Panopticon or Omniopticon tool?. In *Proc. of the 6th Conference of the Surveillance Studies Network*. Barcelona, Spain.

[61] House of Lords House of Commons; Joint Committee on the Draft Online Safety Bill. 2021. Draft Online Safety Bill. https://www.gov.uk/government/publications/draft-online-safety-bill

[62] Ofcom. 2021. Children and parents: media use and attitudes report 2020/21. https://www.ofcom.org.uk/__data/assets/pdf_file/0025/217825/children-and-parents-media-use-and-attitudes-report-2020-21.pdf
[63] Ofcom.org. [n. d.]. children and parents: media use and attitudes report 2019. https://www.ofcom.org.uk/__data/assets/pdf_file/0023/190616/children-media-use-attitudes-2019-report.pdf
[64] Ofcom.org. [n. d.]. Ofcom Children's Media Lives: Life in Lockdown. https://www.ofcom.org.uk/__data/assets/pdf_file/0024/200976/cml-life-in-lockdown-report.pdf
[65] Joint Committee on the Draft Online Safety Bill. 2021. Draft Online Safety Bill. https://committees.parliament.uk/publications/8206/documents/84092/default/
[66] Luci Pangrazio and Lourdes Cardozo-Gaibisso. 2021. "Your Data Can Go to Anyone": The Challenges of Developing Critical Data Literacies in Children. In *Critical digital literacies: Boundary-crossing practices*. Brill, 35–51.
[67] Luci Pangrazio and Neil Selwyn. 2017. 'My Data, My Bad...' Young People's Personal Data Understandings and (Counter) Practices. In *Proceedings of the 8th International Conference on Social Media & Society*. 1–5.
[68] Ashwini Rao, Florian Schaub, and Norman Sadeh. 2015. What do they know about me? Contents and concerns of online behavioral profiles. *arXiv preprint arXiv:1506.01675* (2015).
[69] Mitchel Resnick, Robbie Berg, and Michael Eisenberg. 2000. Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *The Journal of the Learning Sciences* 9, 1 (2000), 7–30.
[70] Bernhard Rieder. 2017. Scrutinizing an algorithmic technique: The Bayes classifier as interested reading of reality. *Information, Communication & Society* 20, 1 (2017), 100–117.
[71] Meredith L Rowe, Şeyda Özçalışkan, and Susan Goldin-Meadow. 2008. Learning words by hand: Gesture's role in predicting vocabulary development. *First language* 28, 2 (2008), 182–199.
[72] Marijn Sax. 2016. Big data: Finders keepers, losers weepers? *Ethics and Information Technology* 18, 1 (2016), 25–31.
[73] Kavya Sharma, Xinhui Zhan, Fiona Fui-Hoon Nah, Keng Siau, and Maggie X Cheng. 2021. Impact of digital nudging on information security behavior: an experimental study on framing and priming in cybersecurity. *Organizational Cybersecurity Journal: Practice, Process and People* (2021).
[74] Mariya Stoilova, Sonia Livingstone, and Rishita Nandagiri. 2020. Digital by default: Children's capacity to understand and manage online data and privacy. *Media and Communication* (2020).
[75] THORN. 2021. Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting, and Blocking. https://info.thorn.org/hubfs/Research/Responding%20to%20Online%20Threats_2021-Full-Report.pdf
[76] Kjerstin Thorson, Kelley Cotter, Mel Medeiros, and Chankyung Pak. 2021. Algorithmic inference, political interest, and exposure to news and politics on Facebook. *Information, Communication & Society* 24, 2 (2021), 183–200.
[77] UNESCO. 2019. Digital Kids: Asia-Pacific competence framework. https://www.gcedclearinghouse.org/sites/default/files/resources/190165eng.pdf
[78] Sandra Wachter. 2020. Affinity Profiling and Discrimination by Association in Online Behavioral Advertising. *Berkeley Tech. LJ* 35 (2020), 367.
[79] Jacqueline M Kory Westlund and Cynthia Breazeal. 2016. Transparency, teleoperation, and children's understanding of social robots. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 625–626.
[80] Randi Williams, Hae Won Park, and Cynthia Breazeal. 2019. A is for artificial intelligence: the impact of artificial intelligence activities on young children's perceptions of robots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.
[81] Karen Yeung. 2020. Recommendation of the council on artificial intelligence (oecd). *International Legal Materials* 59, 1 (2020), 27–34.
[82] YouTube. 2021. Terms of Service. https://kids.youtube.com/t/terms
[83] Brahim Zarouali, Koen Ponnet, Michel Walrave, and Karolien Poels. 2017. "Do you like cookies?" Adolescents' skeptical processing of retargeted Facebook-ads and the moderating role of privacy concern and a textual debriefing. *Computers in Human Behavior* 69 (2017), 157–165.
[84] Jun Zhao, Ge Wang, Carys Dally, Petr Slovak, Julian Edbrooke-Childs, Max Van Kleek, and Nigel Shadbolt. 2019. I make up a silly name' Understanding Children's Perception of Privacy Risks Online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[85] Jun Zhao, Ge Wang, Carys Dally, Petr Slovak, Julian Edbrooke-Childs, Max Van Kleek, and Nigel Shadbolt. 2019. 'I Make up a Silly Name': Understanding Children's Perception of Privacy Risks Online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300336
[86] Shoshana Zuboff. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019*. Profile books.

## A PARTICIPANT IDS

| Participant ID | Age | Gender | Participant ID | Age | Gender | Participant ID | Age | Gender |
|---|---|---|---|---|---|---|---|---|
| P1 | 13yr | Girl | P17 | 11yr | Boy | P33 | 7yr | Boy |
| P2 | 13yr | Boy | P18 | 9yr | Boy | P34 | 7yr | Girl |
| P3 | 12yr | Boy | P19 | 11yr | Girl | P35 | 10yr | Girl |
| P4 | 8yr | Boy | P20 | 7yr | Boy | P36 | 11yr | Girl |
| P5 | 9yr | Boy | P21 | 12yr | Boy | P37 | 12yr | Girl |
| P6 | 9yr | Girl | P22 | 9yr | Boy | P38 | 12yr | Girl |
| P7 | 9yr | Boy | P23 | 12yr | Boy | P39 | 12yr | Boy |
| P8 | 9yr | Boy | P24 | 12yr | Girl | P40 | 13yr | Girl |
| P9 | 10yr | Boy | P25 | 11yr | Girl | P41 | 11yr | Boy |
| P10 | 12yr | Girl | P26 | 10yr | Girl | P42 | 9yr | Girl |
| P11 | 9yr | Girl | P27 | 11yr | Girl | P43 | 11yr | Boy |
| P12 | 10yr | Girl | P28 | 10yr | Boy | P44 | 8yr | Boy |
| P13 | 8yr | Boy | P29 | 12yr | Girl | P45 | 9yr | Girl |
| P14 | 11yr | Boy | P30 | 11yr | Girl | P46 | 7yr | Girl |
| P15 | 8yr | Girl | P31 | 10yr | Girl | P47 | 12yr | Girl |
| P16 | 9yr | Girl | P32 | 11yr | Girl | P48 | 12yr | Boy |

Table 2. Participant IDs with ages and genders