Work-in-Progress: UnBias Youth Juries

Liz Dowthwaite
University of Nottingham
Horizon Digital Economy Research
Nottingham, United Kingdom
liz.dowthwaite@nottingham.ac.uk

Helen Creswick
University of Nottingham
Horizon Digital Economy Research
Nottingham, United Kingdom
helen.creswick3@nottingham.ac.uk

Virginia Portillo
University of Nottingham
Horizon Digital Economy Research
Nottingham, United Kingdom
virginia.portillo@nottingham.ac.uk

Monica Cano
University of Nottingham
Horizon Digital Economy Research
Nottingham, United Kingdom
monica.cano@nottingham.ac.uk

Elvira Perez Vallejos
University of Nottingham
Institute of Mental Health
Nottingham, United Kingdom
elvira.perez@nottingham.ac.uk

Ansgar Koene
University of Nottingham
Horizon Digital Economy Research
Nottingham, United Kingdom
ansgar.koene@nottingham.ac.uk

Menisha Patel
University of Oxford
Human Centered Computing
Oxford, United Kingdom
menisha.patel@cs.ox.ac.uk

ABSTRACT

The UnBias Youth Juries engage young people in discussion of issues that affect their online lives, especially in relation to algorithms, through the presentation of scenarios and prompts. Results from the first wave of juries, held in February 2017, produced valuable data about the concerns of young people and recommendations for improving their digital environments. Feedback regarding the jury structure suggested a series of interactive tasks to elicit more debate amongst the participants. This paper presents results from a pilot study using these tasks and discusses the plans for a second wave of juries. This includes creation of an advisory group to ensure materials are relevant to the target audience, and an open educational resource that enables other groups to run their own youth juries.

CCS CONCEPTS

• **Information systems** → Social networks • **Information systems** → Internet communication tools

KEYWORDS

Youth juries; algorithms; engagement; deliberation; social media.

1 INTRODUCTION

The UnBias project aims to understand the perspectives and concerns of internet users, to provide policy recommendations, ethical guidelines, and 'toolkits' for supporting understanding of online environments, and to raise awareness among online providers about user concerns. As part of this, the UnBias Youth Juries engage young people in discussion of issues that affect their online lives, especially in relation to algorithms. They do so through presenting relevant scenarios as prompts. This paper begins by discussing the importance of young people's agency to influence how their digital environments are regulated. It introduces youth juries as a method for engaging discussion and for eliciting concerns and recommendations on a range of topics. Plans are described for the next wave of youth juries, including results from a pilot involving new activities.

2 BACKGROUND

It is estimated that one third of internet users are under the age of 18 [1], with three quarters of children aged 10-12 having their own social media accounts despite the minimum age requirement from the platforms being 13

[2]. It is therefore important that the digital world is created with young people in mind, as a safe and supportive environment that does not exploit them. Steps have been taken to create guidelines for the ethical treatment of children online [3,4], but several recent reports suggest that there is more to be done to ensure an internet fit for children, both in terms of the regulation of online services, and at the point of their design [1,5,6,7].

Young peoples' input is crucial to ensure interventions and regulations aiming to protect them are effective. It is important to give young people the agency to influence how their online lives are regulated. Indeed, there has been a noticeable shift within social research that recognises the importance of seeing children as social actors in their own lives, although the extent to which this has affected adult agendas remains unresolved [8,9,10]. As such, the youth jury method was designed to ensure that children and young people are able to make an active and valued contribution to research, and allow their voices to be heard by feeding back to stakeholders and policy makers.

Youth juries are similar to focus groups and are designed around the concept of deliberation [11,12]. They follow a stimulus–discussion–recommendation process which allows young people to receive and exchange information, critically examine issues, and come to conclusions. The juries provide a pedagogical tool and a rigorous approach for engaging young people in programmes designed to impact behaviour. The use of scenarios builds upon the methodological research tradition of using vignettes as prompts to elicit reflective responses from participants, for which there is extensive evidence of the underlying social, cognitive and emotional processes which facilitate learning and development [e.g., 13,14]. There are critical lessons from previous research on prevention science [15,16] that has guided the development of the youth juries. The current juries are highly interactive and the scenarios have been co-produced with young people to explore their personal concerns and online experiences. This enhances engagement by making the scenarios more real, easier to relate to, and consequently maximises learning through involvement in discussions. Youth juries have previously been used to engage discussion about digital rights to great success [17,18]. The UnBias Youth Juries draw on and continue this work by focussing on the use of algorithmic decision-making processes online.

A series of 14 juries took place in February 2017, with 144 young people aged between 13 and 23 (mean age 15). The juries were discussion-based and used slides as prompts for three scenarios: personalisation through algorithms, search results and fake news, and transparency and regulation. A facilitator introduced each scenario, and the young people were encouraged to discuss, debate, and offer recommendations for improving issues that were raised. The sessions produced valuable data about participants' concerns and recommendations [19,20,21,22]. Subsequently, feedback from jurors led to a series of interactive tasks being created, allowing the juries to evolve, and ensuring that they continue to represent real issues that young people can relate to, even beyond the life of this project. The tasks were designed to promote debate and discussion between the participants rather than in response to the facilitator, and to aid in explanation of complex concepts which some participants found difficult to understand. The pilot study with these new activities will now be described.

3 METHOD

3.1 Participants

Participants were self-selected and either signed up online beforehand or "dropped-in" to one of the sessions. Recruitment materials included posters in the venue and local area, and emails to local schools. In total 13 participants took part in 2 two-hour youth juries held on the same day. There were 8 females and 5 males, aged between 13 and 17 (average 16). There was an appropriate balance of gender and ages to elicit a range of responses. All participants were thanked with a high street shopping voucher and a leaflet providing more information about online tracking and other issues.

3.2 Materials and Procedure

The study was approved by the Ethics Review Board for the Department of Computer Science at The University of Nottingham. The content of the UnBias Youth Juries is dynamic and changes in response to each group, but the overall format follows a specific path, described here. All juries were audio recorded and notes were taken. They were fully transcribed and a thematic analysis carried out.

Figure 1: An example 'map' of the young peoples' Internet.

3.2.1 Introduction. The juries started with a multiple-choice questionnaire, to establish some brief information about the jurors' level of knowledge and opinions. Participants were then asked to discuss the types of online services and applications they use on a regular basis (e.g., social media, watching videos, shopping), which were written on a large sheet of paper visible throughout the

sessions. They were provided with a sheet of stickers depicting the logos of popular online services, and asked to place the ones they used by the relevant activities they used them for. This created a "map" of their internet use (see Fig. 1) and formed the context for the juries. The facilitator then introduced the concept of algorithms with reference to the real experiences of the participants. This involved presenting a "black box" to the jurors to represent an algorithm collecting and processing data (see Fig. 2, left).

Figure 2: The "black box" algorithm. Data can be added through the slot but what is going on inside? Each black side is removed during scenario 4 to reveal a different idea about algorithm transparency.

3.2.2 Scenario 1: Online personalisation. Participants were asked to suggest data that may be collected about them online, before being presented with a series of 45 cards listing examples such as name, location, likes and shares, etc. These were used to discuss how various services might use data to personalise online experiences. The data cards were added to the "black box" as they were discussed. Finally, participants were asked to come up with possible benefits and risks of personalisation which they put on the map with post-its.

3.2.3 Scenario 2: Internet search results. After a short break, participants were presented with two real life examples of unfairness in search results: an image search for "professional hair" which discriminated against black people, and an autocomplete example which discriminated against women. They were asked to vote on the outcome, in terms of what should be done and who is responsible, respectively.

3.2.3 Scenario 3: Regulation. The third scenario used a hypothetical legal case in which the use of an algorithm caused harm, with four possible parties who could be "put on trial". The participants first discussed who was to blame, before they chose one and became jurors either for the prosecution or the defence, and then voted on whether that party was guilty or not guilty.

3.2.4 Scenario 4: Transparency. Each side of the "black box" was removed in turn to reveal a different idea for algorithm transparency (computer code, flowchart, Facebook Edgerank, and a blank side, see Fig. 2, right), prompting discussion about their clarity and usefulness.

Participants were then asked to describe their own ideas for transparency that would be meaningful to them and give them the information they desired.

3.2.5 Recommendations and Wrap-up. The sessions ended with a summary and participants were asked to provide recommendations or suggestions for increasing fairness and preventing bias in any or all of the scenarios presented. They were then given a final questionnaire, aimed at capturing any changes in opinion or knowledge.

4 PRELIMINARY RESULTS

4.1 Quantitative Results

4.1.1 Pre-Session Survey. Most (85%) used search engines several times a day, but did not know much about the way such websites ranked information (46% "Not Much" and 46% "A little"). They did however feel that this was important for people to know (23% "Really Important" and 54% "Quite Important"). Responses were roughly tied between wanting a more personalised (54%) or more neutral (46%) internet experience. The most frequent response to who "makes sure the internet and digital world is safe and neutral" was the big tech companies like Google, Twitter, or Microsoft (39%) with remaining responses split between the other options (15% each "The Police", "Nobody", "The Government").

They thought these companies had too much power over their lives (85%). Most felt that they could influence how the digital world works (46% "A little" and 46% "A lot"), but that they should have more influence (69%).
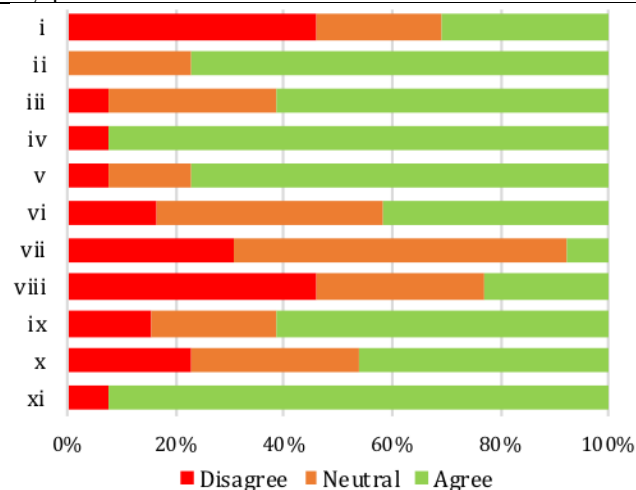
4.1.2 Post-Session Survey. Everyone felt that they had learnt something about algorithm fairness (85% "A lot" and 15% "A little") and how the Internet affects their lives (62% "A lot" and 39% "A little"). Most also came up with ideas about how the internet could be made better for people like them (54% "Lots", 39% "A few"). Participants were asked to rate their agreement with a series of statements on a scale of 1 (very little) to 10 (very much). A score between 1 and 4 can be considered negative or disagree, and a score of 7 or above positive or agree, with 5 and 6 being neutral. As shown in Table 1, participants "agreed" with 6 of these statements and did not "disagree" with any of them. They felt that recommender systems did make their lives easier, but that "social media sites should not influence the information to their users" and "it should be made easier for people to remove digital content about themselves". They also felt that they had changed their minds about how the internet should work, that "13-24 year olds should influence how digital technologies and services are run", and they would like to have more control of what happens to them and their data online.

Looking at the spread of responses (see Fig. 3), the neutral responses to "search engines like Google should produce the same results for everyone" and "nobody is going to listen to what young people say about the Internet they

want" are quite balanced but perhaps leans towards "disagree". "The big tech companies are accountable enough to users of digital technologies" elicited a similar number of neutral and positive responses.

Table 1: Mean responses to statements in the post-session survey

Statement	Mean Response
i) Search engines like Google should produce the same results for everyone	5
ii) Recommender systems make my life easier. I save time and effort	7
iii) Social media sites should not influence the information to their users	7
iv) It should be made easier for people to remove digital content about themselves	9
v) 13-24 year olds should influence how digital technologies and services are run	8
vi) The big tech companies are accountable enough to users of digital technologies	6
vii) I'm confident that I can influence the way digital technologies work for young people	5
viii) Nobody is going to listen to what young people say about the Internet they want	5
ix) I've changed my mind today about how the internet and other digital technologies should work	7
x) When I use digital technologies, I can determine what happens to me and my personal data	6
xi) When I use digital technologies, I'd like to have more control of what happens to me and my personal data.	9

Figure 3: Spread of responses to post-session survey statements.

It appears they did not feel very confident that they could "influence the way digital technologies work for young people", but that they did have some say over their own data when using digital technologies. Finally, nearly half of the respondents disagreed that "nobody is going to listen to what young people say about the Internet they want".

4.1 Qualitative Results

The qualitative data demonstrates how the addition of activities has helped to elicit valuable insights from the jurors. The ice-breaker activity, which asked participants to visualise their internet use, encouraged them to reflect on their online activity: "it was weird seeing how much, how many different things you actually use on a regular basis". They also reflected on the overall internet activity of the group, remarking that social media, and in particular Instagram, were enormously popular.

After the topic of algorithms was introduced, participants were asked if they were able to define what they thought an algorithm might be. This was met with a limited response, with some explaining that this was challenging for them: "it's hard to explain", or struggling to come up with a definition at all. However, the introduction of the "black box" was a helpful visual, provoking thoughts about the use of data and how algorithms might be involved: "some companies use what you post on social media to target adverts towards that."

The subsequent presentation of the data cards, helped to represent the breadth of data that might be collected online. Upon seeing multiple data cards distributed amongst participants, one juror described it as "scary". The data cards activity also provoked further discussion amongst participants about their privacy and the extent to which their information is shared online: "you put something on your private page and you think it's only going to be me and my followers, but actually I don't think it is." Concern about sharing sensitive information, especially financial, and the related assumptions that are made by the algorithm was also raised.

The search results scenarios elicited discussion of the challenges in regulating discrimination: "Well, it is bad but there's not really much we can do because you can't just employ thousands of people just to search through images […] because a computer has no concept of racism or sexism or anything like that." Others reflected on the need for legislation to tackle such issues: "I think there should be legislation preventing racist things coming out on the internet". This was pitched against a belief that freedom of speech should be protected: "however, I think when the government interfere it could, I don't know, it might limit the ability of people to express themselves on social media."

The various versions of algorithm transparency, represented through different images on the "black box"

offered a useful starting point for considering how algorithms might best be presented to the user, including what information would be important to know: "how they're using that data to make those assumptions, and then also where that data's going", and how this information should be presented. They unanimously felt that social media platforms releasing their code was "not at all" helpful or clear. Representations such as Facebook Edgerank were also not seen as helpful or transparent: "I mean it still doesn't make that much sense". This also generated further reflections as one juror continued, "and I feel they probably know that it doesn't make that much sense, but they've said oh here you go, here's a simple way. So that's them essentially doing their job".

5   DISCUSSION AND ONGOING WORK

This paper presented the results from a pilot study of 2 youth juries using interactive activities, created in response to a first discussion-based study. The results resemble findings from previous juries, suggesting the activities did not bias or skew opinions, and they encouraged increased participation and discussion. Young people care about how their online lives are managed and feel that it should be easier to remove content they no longer want people to see. Whilst they found recommendation algorithms useful, they also recognised the drawbacks and did not want sites to influence their opinions, seeking a balance in how information is displayed. They showed distrust towards social media platforms, who they felt were not necessarily motivated to behave ethically towards the user. It is heartening that they felt that young people like them should be able to influence how digital services are run, and that in fact they had a say about their own data, although they would like more control over it. They left the sessions having learnt about their online lives and thinking about how they may affect the way the online world works. Youth juries are a powerful method for facilitating the elicitation of the views of young people.

The activities aided participants in giving their reaction to a scenario, and also sparked deeper thoughts, insights and reflections about how their data may be being used in the digital world. The varying viewpoints indicate that the scenarios were an effective method of eliciting views and encouraging critical reflections on them. Utilising the "black box" in particular helped young people to express their views as they instantly offered feedback on the different ideas that were presented to them. Transparency was an area that previous juries found difficult to discuss. Visual representations helped to build the confidence of jurors to give their opinions, to give feedback on the current task, and to generate fruitful further responses.

As part of the UnBias project, the youth jury work package will help to ensure the internet is fit for young

people, both in terms of regulation of services and at the point of design. Using their input will ensure that interventions and regulations aiming to protect them are meaningful and effective, and help young people to influence how their online lives are regulated. Results from the two juries reported here fed into the design of another wave of youth juries in February 2018. This included creation of an advisory group of young people to ensure the materials are relevant to the target audience, and an open educational resource so that other groups can run their own youth juries.

5.1   Advisory Group

Participants in the 2 youth juries reported in this paper were invited to join an advisory group to help to plan activities to make them more engaging and relevant to young people. An email list and monthly informal meetings were set up. The group are aged between 16 and 18, and have produced valuable feedback on the overall structure and design of the juries, and on the pre- and post-session questionnaires, including simplifying the language used, making concepts more easily understood, particularly by younger participants. The activities have also been modified and strengthened: scenario 2 has been removed and integrated with scenario 1, which is now more structured in terms of the tasks and focusses on "the use of algorithms. Two activities have been added which make better use of the data cards: a "personal filter bubble" task and a "data as currency" task. Scenario 3 (regulation) has been made more relevant to young people, with two cases which can be chosen depending on the age and knowledge of the group: children being recommended inappropriate content, and teenagers being targeted by an essay writing service. These activities were also piloted with the advisory group before the next wave of juries.

5.2   Open Educational Resource

In addition to helping researchers interested in how young people consider their online lives, the youth juries are a valuable tool for engagement, increasing awareness, and giving young people agency to influence their online lives. As such, it is valuable for others, for example teachers and youth workers, to be able to use this method to run youth juries with their own groups. To this end, an online open educational resource is in development, detailing the method, and providing downloadable versions of materials, including data cards, filter bubbles, app logo stickers, and the jury scenarios. This resource can be found at http://uyj.wp.horizon.ac.uk/.

ACKNOWLEDGMENTS

Digital Economy Research Institute, University of Nottingham.

REFERENCES

[1] S Livingstone, J Carr, and J Byrne. 2016. *One in Three: Internet Governance and Children's Rights.* UNICEF Office of Research, Florence.

[2] Ofcom. 2017. *Children and Parents: Media Use and Attitudes Report.*

[3] UK Council for Child Internet Safety. 2015. *Child Safety Online: A Practical Guide for Providers of Social Media and Interactive Services.*

[4] unicef. 2010. A Summary of the UN Convention on the Rights of the Child.

[5] Children's Commissioner for England. 2017. *Life in "likes": Children's Commissioner report into social media use among 8-12 year olds.*

[6] Children's Commissioner for England. 2017. *Growing Up Digital: A report of the Growing Up Digital Taskforce.*

[7] Beeban Kidron and Angharad Rudkin. 2017. *Digital Childhood: Addressing childhood development milestones in the digital environment.* 5Rights.

[8] Berry Mayall. 2011. The sociology of childhood in relation to children's rights. In *Children's rights: Progress and perspectives.* Brill, 429–446.

[9] Gina Porter, Janet Townsend, and Kate Hampshire. 2012. Children and young people as producers of knowledge. *Child. Geogr.* 10, 2 (May 2012), 131–134.

[10] Liz Todd. 2012. Critical dialogue, critical methodology: bridging the research gap to young people's participation in evaluating children's services. *Child. Geogr.* 10, 2 (May 2012), 187–200.

[11] Julia Abelson, Pierre-Gerlier Forest, John Eyles, Patricia Smith, Elisabeth Martin, and Francois-Pierre Gauvin. 2003. Deliberations about deliberative methods: issues in the design and evaluation of public participation processes. *Soc. Sci. Med.* 57, 2 (2003), 239–251.

[12] Stephanie Solomon and Julia Abelson. 2012. Why and when should we use public deliberation. *Hast. Cent. Rep.* 42, 2 (2012), 17–20.

[13] David Davis. 2014. *Imagining the real: towards a new theory of drama in education.* Bloomsbury, London.

[14] Aysem Tombak. 2014. Importance of Drama in Pre-school Education. *Procedia - Soc. Behav. Sci.* 143, (August 2014), 372–378.

[15] G Crombie and A Trinneer. 2003. *Children and Internet Safety: An Evaluation of the Missing Program. A Report to the Research and Evaluation Section of the National Crime Preventure Centre of Justice Canada.* University of Ottowa, Ottowa.

[16] Ruth Shillair, Shelia R. Cotten, Hsin-Yi Sandy Tsai, Saleem Alhabash, Robert LaRose, and Nora J. Rifon. 2015. Online safety begins with you and me: Convincing Internet users to protect themselves. *Comput. Hum. Behav.* 48, (July 2015), 199–207.

[17] S Coleman, Kruakae Pothong, Elvira Perez Vallejos, and Ansgar Koene. 2017. *The Internet on Our Own Terms How Children and Young People Deliberated About Their Digital Rights.*

[18] Elvira Perez Vallejos, Ansgar Koene, Chris James Carter, Ramona Statache, Tom Rodden, Derek McAuley, Monica Cano, Svenja Adolphs, Claire O'Malley, and Kruakae Pothong. 2016. Juries: acting out digital dilemmas to promote digital reflections. *ACM SIGCAS Comput. Soc.* 45, 3 (2016), 84–90.

[19] Ansgar Koene. 2017. Algorithmic Decision-Making: Fairness, Bias and the Role of Ethics Standards. In *European Big Data Value Forum.* Versailles, France.

[20] Ansgar Koene, Elvira Perez Vallejos, Helena Webb, and Menisha Patel. 2017. Human Agency on Algorithmic Systems. In *AoIR2017.*

[21] Elvira Perez Vallejos, Ansgar Koene, Virginia Portillo, Liz Dowthwaite, and Monica Cano. 2017. Young People's Policy Recommendations on Algorithm Fairness. In *Proceedings of the 2017 ACM on Web Science Conference*, 247–251.

[22] Virginia Portillo, Elvira Perez Vallejos, Monica Cano, Liz Dowthwaite, and Ansgar Koene. 2017. Algorithmic taste management: young people's perspectives and recommendations.