The Tension Between Safety and Self-Expression in Adolescent Use of Social XR

MARK MCGILL and MOHAMED KHAMIS, University of Glasgow, UK

1 INTRODUCTION

As affordable, consumer-oriented mixed reality headsets find their way into the home, it becomes increasingly likely such technology will see adoption by children and adolescents, particularly for social VR given the role social experiences have played in smartphone adoption and usage. Where a new disruptive technology has entered the market, parental understanding, supervision, and controls have typically lagged, leading to a window (often years wide) where children and adolescents experience unsupervised access to new technologies. Whilst beneficial (e.g. in terms of technological literacy), historically there have been examples where this lack of safeguards has led to children experiencing new forms of bullying, harassment and abuse [7], often unbeknownst to parents.

This submission discusses the rationale behind our recently funded Facebook Reality Labs award, which will set out to explore existing and potential adoption of social XR experiences by children and adolescents; the risks posed to these groups by using these platforms with the current supervision and moderation capabilities; and the potential for parents to play a constructive role in guiding and moderating adolescent use of social XR, in ways that strive to preserve the ability for these groups to use these platforms as safe spaces for exploration, socializing and self-discovery.

2 WHY FACILITATE PARENTAL INSIGHT AND CONTROL?

As it stands, social XR experiences typically lack any sophisticated mechanisms for parental control over what is being experienced, or insight (either in real-time or after-the-fact) regarding what was experienced. At a platform level, devices like the Oculus Quest have little-to-no capacity for supporting parental restrictions on applications, and devices are tied to social media accounts that are not intended to be used by children. This ignores the reality that children will inevitably use these technologies, often with tacit parental agreement. At an application level, social XR experiences commonly lack any means of verifying the identity/age of the current user, and control over the experience is dictated solely by the current user. If a user of VR Chat for example experiences harassment, they have tools in-game to denylist/mute others, and report misconduct. However, there is no ability for a third party (i.e. the parent) to be aware of what happened, or exert control over what is allowed to happen in the future.

This lack of parental insight and control is particularly important when we consider what new forms of potential misuses and abuses are made possible by embodied social experiences. For example, our own research [5] has shown how VR can be differently affective compared to non-VR, specifically in relation to violent experiences, raising concerns regarding how content is rated, and how these ratings are presented (e.g. to parents). A plethora of research has demonstrated the potential psychological impact of VR, both through manipulation of the representation of self and others (e.g. the proteus effect [6]), and through exposure to virtual events with ever increasing realism [3], and the potential for mis-use and abuse of mixed reality technology [2]. Franks [1] remarked on the potential for VR to: induce trauma and "compassion fatigue" through exposure to e.g. violent actions or sexual harassment; enhance the

capability of others (e.g. the state) to monitor and punish our actions and the actions of others; and enable new forms of virtual-physical abuse e.g. avatar rapes, "profoundly disturbing" events where avatars could be coerced or externally controlled against the user's wishes.

3 TENSION BETWEEN SAFETY AND PRIVACY

Against this backdrop, it appears reasonable to suggest that the solution is simple: introduce strong moderation tools, backed by machine learning / social signal processing to detect aberrant behaviours, allowlists of users, and adult moderators, to create a walled garden of child/adolescent social VR experiences. And we do accept that moderation tools will play an important role in safeguarding the activities of children and adolescents in these online experiences.

However, there is an underlying tension here, between the ultimate safety of the child/adolescent user(s), and their ability to explore their own identities and enjoy the freedom of expression that online experiences can enable. Unicef developed an industry toolkit which outlined some key principles on children's online privacy and freedom of expression, noting in particular in principle 4 that "Children's privacy and freedom of expression should be protected and respected in accordance with their evolving capacities" [4]. Of particular note here is the determination that parents play an active role in what their children experience ("parents or guardians play a more active role in deciding the scope and nature of the information and content that younger children can share and consume, while also considering children's views and opinions") and that monitoring or filtering tools "take account of older children's rights and abilities to make empowered and informed decisions online". Consequently, we propose to focus on two avenues:

Parental moderation controls e.g. allowlists of verified friends to interact with, control over language used etc. This can be heavyweight, intrusive, and is effectively restrictive by default - but can be an effective tool in creating a walled garden experience.

Supporting parental insight of current/past events Can we automatically journal or self-report key sensitive events in forms parents can manage and digest, enabling parents to discuss and support their children to understand/learn from (potentially problematic) experiences and exposures. This approach inherently brings with it a greater amount of freedom, and consequently the potential for greater exploration and self-discovery - but at an acknowledged cost that some aberrant behaviours will be experienced - albeit, arguably this cost always exists as no moderation tools will be perfect in preventing such events.

In effect, our aim is to explore whether we can more actively inform parents in comprehending what happens to their children in these online virtual worlds, so that they might play a more active and supportive parental role in their child's use of these platforms, in line with guidance from the likes of unicef regarding a child's evolving right to privacy and freedom of expression.

There are a number of ethical challenges in exploring this proposed research, predominantly because exploring this tension between insight and moderation suggests that there will be children and adolescents that engage in social VR experiences where the fullest extent of the moderation tools are not utilized, in lieu of relying on real-time or post-incident parental insight as a means of intervention. We anticipate that this tension will have to be uniquely managed by parents based on the age and maturity of their children. However, we do refer back to the reality of online social experiences - moderation tools are often unavailable or poorly utilized, with adolescents in particular using platforms intended for adults with parental knowledge, despite this often being against the terms of service. In effect, we argue that moderation and walled gardens are insufficient, and that more general purpose tools and interventions are necessary to facilitate guardianship of children and adolescents in these new worlds. We welcome discussion and feedback on our proposal, and would seek to provoke discussions within the workshop around the use of social XR in vulnerable groups.

REFERENCES

[1] Mary Anne Franks. 2017. The Desert of the Unreal: Inequality in Virtual and Augmented Reality. *U.C.D. L. Rev.* (2017).

[2] Jan Gugenheimer, Mark McGill, Samuel Huron, Christian Mai, Julie Williamson, and Michael Nebeling. 2020. Exploring Potentially Abusive Ethical, Social and Political Implications of Mixed Reality Research in HCI. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (2020). https://doi.org/10/ghd5xc

[3] Mel Slater, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, Graham Breen, Raz Schwarz, William Steptoe, Dalila Szostak, Shivashankar Halan, Deborah Fox, and Jeremy Silver. 2020. The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality* (2020). https://doi.org/10/ggpvct

[4] unicef. 2018. Children's Online Privacy and Freedom of Expression. *Industry toolkit* (2018).

[5] Graham Wilson and Mark McGill. 2018. Violent Video Games in Virtual Reality: Re-Evaluating the Impact and Rating of Interactive Experiences. In *CHI PLAY '18 Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play.* https://doi.org/10/gf3ngh

[6] Nick Yee, Jeremy N. Bailenson, and Nicolas Ducheneaut. 2009. The Proteus Effect: Implications of Transformed Digital Self-Representation on Online and Offline Behavior. *Communication Research* 2 (2009). https://doi.org/10/dz7hqx

[7] Izabela Zych, David P. Farrington, and Maria M. Ttofi. 2019. Protective Factors against Bullying and Cyberbullying: A Systematic Review of Meta-Analyses. *Aggression and Violent Behavior* (2019). https://doi.org/10/dr4n