EPFL

Spotlight on risk

Using “proof of personhood” to tackle social media risks

Aengus Collins and Bryan Ford

15 March 2021

The ease of creating fake virtual identities plays an important role in shaping the way information—and misinformation—circulates online. Social media platforms are increasingly prominent in shaping public debates, and the tension between online anonymity and accountability is a source of growing societal risks. This article outlines one approach to resolving this tension, with “pseudonym parties” that focus on proof of personhood rather than identity. Pseudonym parties are a low-tech approach to important digital challenges, linking online activity to anonymous digital tokens that are obtained by being physically present at an appointed time and place.

Social media platforms have led to major structural changes in the flow of information. With the growth of horizontal, peer-to-peer patterns of information exchange, these flows are relatively unconstrained by gatekeeping functions. The resulting information ecosystem can be personally, socially and professionally enriching. But it also entails risks, including fake news, conspiracy theories, echo chambers and, as a result, the potential distortion of political discourse and decisions.

Reconciling anonymity and accountability

The ease of creating fake virtual identities is an important enabler of these risks, because it makes it difficult to sanction rule-breakers. There are four issues here. First, the anonymity of fake virtual identities makes it difficult for sanctions to reach the person behind them. Second, fake virtual identities are easily replaceable, so even if you sanction a user's account, they can create another for continued abuse. Third, fake virtual identities are cheap, so determined abusers can amplify their abuse by maintaining multiple fake virtual identities at once. Fourth, abusers can use automation to amplify their power by orders of magnitude, by creating bot armies that become ever-more indistinguishable from real profiles with the help of deepfake technologies.

These issues yield an apparent conflict between anonymity and accountability. One way of resolving this would be to jettison anonymity and require that all social media accounts link to a verified identity. But anonymity is closely connected to important societal and political values, such as freedom of expression and association, that many people find in online communities and would not want weakened. Losing anonymity would be a costly way of achieving accountability on social media platforms.

However, it may be possible to guarantee both anonymity and accountability. Showing that each virtual avatar has a real-world counterpart does not necessarily require full identification. Proof of *personhood*, not *identity*, may be sufficient. A proof of personhood would need to establish digital tokens, each of which uniquely corresponds to one and only one person at some point in time. Accounts based on proof of personhood, while still anonymous, are no longer cheap and easily replaceable. Sanctions against abusive accounts thereby become more effective, and abusers lose the amplification power of creating many fake accounts, either manually or by automation. Research suggests all this can be done without compromising anonymity.

Pseudonym parties for proof of personhood

One proposal for a proof of personhood system is the idea of "pseudonym parties", first proposed more than a decade ago and recently revived in the context of broader debates about digital democracy and the relationship between technology and politics. The basic idea is simple. Pseudonym parties are in-person events, held publicly and periodically (e.g., once per month or per year), at which every attendee receives a single anonymous digital *token*—a cryptographic random number. In theory, individual events could be organised independently and locally, but shared procedures and controls would be needed if different groups wish to recognize each others' tokens. Attendance at pseudonym parties is voluntary, and attendees need not reveal anything about themselves. A token simply attests that someone is a real person who turned up at a physical location at an appointed time. It provides proof of personhood without the need for any identifying data.

An individual should be able to obtain only a single token at a single event in each cycle. This creates logistical hurdles. To prevent someone from attending multiple parties in different locations to obtain multiple tokens, pseudonym parties must be coordinated to take place simultaneously, so that it is physically possible to attend only one in each cycle. This requirement to be in a certain place at a certain time may be seen by many as a significant imposition. Moreover, simultaneous pseudonym parties would become more challenging to coordinate as their scope increases. In principle, they could be scaled globally, but simultaneity would involve significant time zone issues.

Depending on what proof of personhood tokens are used for, further precautions may be needed. For example, if tokens were used for voting, protections would be needed to prevent people from obtaining additional tokens—and therefore more voting power—through payment or coercion. Tokens used for voting would also need a limited lifespan, requiring users to obtain a fresh one each cycle. Otherwise, an individual could accumulate more valid tokens over time. Other uses for tokens would not require these additional measures. Using tokens to replace increasingly-difficult CAPTCHA puzzles, for example, could reduce the rate of abuses by orders of magnitude even without limiting token lifespan.

In addition to preventing users from obtaining multiple tokens per cycle, strong transparency provisions would be required to prevent corrupt organisers from fraudulently elevating the attendance numbers and producing additional tokens for themselves. With a small event, it would be relatively straightforward to ensure the number of people present matches the number of tokens issued. Greater vigilance would be necessary for larger events. Secure provisions for remote participation would also be required, to accommodate people who want to obtain a token but are unable to attend a pseudonym party, for example due to illness, disability or work commitments.

Proof of personhood and social media

For the social media risks outlined at the outset of this article, proof of personhood could help to rebuild trust that online social communication is driven by real people rather than by bots and sockpuppets trying to skew discussions or manipulate the algorithms that influence who sees what. One way of doing this would be to use proof of personhood tokens to produce verified counts for key social media metrics that shape online visibility and reputation, such as how many "followers" an account has or how many "likes" a piece of content receives.

A user might have multiple accounts on any given online service—such as for different personal and professional activities—but if tokens rather than accounts are the basis of verified counts, then a user's multiple accounts would not lead to an increased ability to influence or amplify online discourse. If a user "likes" the same piece of content using multiple accounts, it would increase the like count by only one, because all the accounts are tied to the same person. Similarly, if a user "follows" another user via multiple accounts, the latter user's follower count increases by only one. Rights and quotas for posting new content could be managed per person, so that users may post via any of several accounts, but more accounts do not confer a right to more posts or a "louder voice" online.

Policy implications

The idea of pseudonym parties is new, and requires further technical and logistical elaboration. Proceeding with the idea would also raise numerous questions for policymakers to resolve, including the following.

Is it effective? Would a system of proof of personhood tokens deliver a significant increase in accountability on social media platforms? For example, a social media platform could block a token that has been linked to rule-breaking, thus shutting out any accounts linked to it. But depending on how much time is left until the next cycle of pseudonym parties, that may only be a minor sanction. As noted above, it may not amount to a sanction at all if extra tokens can be purchased or coerced from other users. There is also the question of what would happen in case of significant harm or illegality: would a proof of personhood system make it easier or more difficult, relative to the status quo, for someone to be identified and held to account?

Is it feasible? There are technical and practical challenges, but they do not seem to be insuperable. However, problems of personal inconvenience should not be underestimated. Tying commonplace social media activities, such as registering a "like", to a requirement to be in a certain place at a certain time each year would be a major departure from the status quo. Concerns about convenience might diminish depending on the extent of other digital activity—such as various forms of digital democracy—that could be carried out with a proof of personhood token. Nevertheless, making it as easy as possible to attend a pseudonym party would have to be a priority.

Is it the only approach? There are a range of experimental ways of trying to establish proof of personhood. Some of these tie online activity to an individual's identity: typically using government-issued documents (a passport or similar) or biometrics (iris scans and fingerprints). Others rely on social trust principles, with participants in a digital network attesting that their connections' online identities are valid and not fake. Out of four key goals for proof of personhood—privacy, security, inclusivity and equality—preliminary analysis suggests that only pseudonym parties appear capable of achieving all four.

Should it be mandatory? The proposal for pseudonym parties assumes that the system would begin on a local and voluntary basis, establishing a proof of concept which could then be expanded. However, to have a meaningful impact on the way social media networks function, some incentives or compulsion would ultimately be required. This could be achieved by companies requiring proof of personhood for certain online activities, such as setting up an account or registering a verified "like" or "follow". It is unlikely that companies would introduce a layer of user inconvenience like this without strong external impetus. Governments might consider putting some of these requirements on a statutory footing, but depending on the details this could clash with free speech protections in some jurisdictions.

Who would manage and control it? For a system of small-scale voluntary pseudonym parties, events could be managed on a relatively informal basis. If the system is to scale nationally or internationally, then this is likely to require state and/or private-sector involvement. There are possible analogies here to the levels of administration required to conduct elections of different types and sizes, or to introduce digital IDs as some countries have done.

Pseudonym parties deal with the problem of fake virtual identities by anchoring online activity in the provable existence of real (but anonymous) humans. There are significant challenges of running this sort of system smoothly, securely and without undue individual inconvenience. But at a time of heightened regulatory interest in the risks associated with social media, this proposal deserves further attention, not least to see what other ideas it may spark for dealing with social media risks.

Acknowledgements
The authors would like to thank Arthur Petersen (University College London) for his review of an earlier draft of this article.

- Allyn, Bobby. 2020. "Researchers: Nearly Half of Accounts Tweeting About Coronavirus Are Likely Bots." NPR.Org. May 20, 2020.
- Branscomb, Anne Wells. 1995. "Anonymity, Autonomy, and Accountability: Challenges to the First Amendment in Cyberspaces." *The Yale Law Journal* 104 (7): 1639–79.
- Dzieza, Josh. 2019. "Why CAPTCHAs Have Gotten so Difficult." The Verge. February 1, 2019.
- European Commission. n.d. "The Digital Services Act: Ensuring a Safe and Accountable Online Environment." Text. European Commission – European Commission. Accessed January 25, 2021.
- Ford, Bryan. 2020. "Identity and Personhood in Digital Democracy: Evaluating Inclusion, Equality, Security, and Privacy in Pseudonym Parties and Other Proofs of Personhood." *ArXiv:2011.02412 [Cs]*, November.
- Ford, Bryan, and Jacob Strauss. 2008. "An Offline Foundation for Online Accountable Pseudonyms." In *Proceedings of the 1st Workshop on Social Network Systems*, 31–36. SocialNets '08. New York, NY, USA: Association for Computing Machinery.
- Knake, Robert. 2021. "2021: The Year We Kick the Dogs Off the Internet." *Council on Foreign Relations* (blog). January 19, 2021.
- Kumar, Srijan, Justin Cheng, Jure Leskovec, and V. S. Subrahmanian. 2017. "An Army of Me: Sockpuppets in Online Discussion Communities." *Proceedings of the 26th International Conference on World Wide Web*, April, 857–66.
- Matthews, Jeanna. 2020. "Bots and Trolls Control a Shocking Amount of Online Conversation." *Fast Company*, June 29, 2020.
- Nicas, Jack. 2020. "Why Can't the Social Networks Stop Fake Accounts?" *The New York Times*, December 8, 2020, sec. Technology.
- Vincent, James. 2019. "ThisPersonDoesNotExist.com Uses AI to Generate Endless Fake Faces." The Verge. February 15, 2019.