Comment

MATTHEW TOSTEVIN/REUTERS

Researchers have added puzzles to the game Minecraft to help study behaviour.

How games can make behavioural science better

Bria Long, Jan Simson, Andrés Buxó-Lugo, Duane G. Watson & Samuel A. Mehr

Wordle, Minecraft and Scrabble are played online by millions. Gamifying experiments can make behavioural research more inclusive, rigorous and reproducible — if it's done right.

When US cognitive scientist Joshua Hartshorne was investigating how people around the world learn English, he needed to get tens of thousands of people to take a language test. He designed 'Which English?', a grammar game that presented a series of tough word problems and then guessed where in the world the player learnt the language. Participants shared their results — whether accurate or not — on social media, creating a snowball effect for recruitment. The findings, based on data from almost 670,000 people, revealed that there is a 'critical period' for second-language learning that extends into adolescence[1].

This sort of 'gamification' is becoming a powerful research tool across fields that study humans, including psychology, neuroscience, economics and behavioural economics. By making research fun, the approach can help experiments to reach thousands or millions of participants. For instance, experiments embedded in a video game demonstrated that the layout of the city where a child lives shapes their future navigational ability[2]. Data from a digital word search showed that people who are skilled at the game do not necessarily give better advice to those trying to learn it[3]. And

© 2023 Springer Nature Limited. All rights reserved.

a dilemma game involving millions of people revealed that most individuals have reliable moral intuition[4].

Gamification can help to avoid the pitfalls of conventional laboratory-based experiments by allowing researchers to study diverse populations, to conduct more-sophisticated experiments and to observe human behaviour in naturalistic environments. It can improve statistical power and reproducibility, making research more robust. Technical advances are making gamification cheaper and more straightforward, and the COVID-19 pandemic has forced many labs to move their human experiments online. But despite these changes, most have not yet embraced the opportunities gamification affords.

To reach the full potential of this approach, researchers must dispel misconceptions, develop new gamification technologies, improve access to existing ones and apply the methods to productive research questions. We are researchers in psychology, linguistics, developmental science, data science and music who have run our own gamified experiments. We think it's time for science to get serious about games.

Games theory

Gamification motivates people to participate in experiments by incorporating point-scoring, competition, feedback about performance and the opportunity to learn about oneself through play. In early forays in 2005 and 2008, cognitive neuropsychologist Laura Germine, now at Harvard Medical School in Boston, Massachusetts, adapted psychological assessments and placed them on her citizen-science website, TestMyBrain.org. Her work showed that self-selected samples can produce high-quality data even when participants are unsupervised and unpaid[5]. As gamification has developed, it has united developmental psychology with computer science, web development and user-experience research to create exciting, immersive encounters for participants.

Gamification can include transforming experiments into bespoke games, embedding experiments in existing games and extracting data from ongoing ones. The popular game Wordle – in a sense, the world's largest psycholinguistics experiment – has already inspired investigations on topics such as optimization problems in active learning[6] and the contexts in which people cheat[7].

Participants in conventional lab-based studies of human behaviour are often few in number and WEIRD (that is, from Western, educated, industrialized, rich and democratic societies). This leads to results that are statistically imprecise or irreproducible, or that cannot be generalized to other groups.

The massive data sets enabled by gamified science can help to address questions about reproducibility and generalizability.

BRIA LONG

A kiosk at the Children's Discovery Museum of San Jose, California, collected drawings from thousands of two- to ten-year-olds to reveal how drawing changes systematically with age.

For example, small studies had shown that a person's experience of speaking a tonal language – one that uses pitch, or tones, to distinguish between words, as in Mandarin – alters their ability to perceive musical pitch. However, such work had been conducted largely in Mandarin or Cantonese. To explore tonal languages that are less commonly studied, Jingxuan Liu, now a graduate student at Columbia University in New York City, and one of us (S.A.M.) used data from a popular web-based quiz, 'Test your Musical IQ', to replicate these findings in half a million speakers of such languages, including Ewe, spoken in West Africa, and Burmese, used in Myanmar[8].

> **"Many gamified experiments have gone viral on YouTube, Reddit and other social-media platforms."**

Issues of reproducibility and generalizability are particularly acute for scientists who work with hard-to-reach study populations, such as children. Gamified experiments have the potential to encourage participation from children in settings such as schools or museums, rather than requiring a special trip to a lab, which only some families have the time for or interest in doing. For example, one of us (B.L.) installed a 'Let's Draw!' kiosk at the Children's Discovery Museum of San Jose in California. Children visiting the museum were asked to draw various things – such as a watch or a tiger – and then play games trying to recognize each other's drawings. Over the next 18 months, the kiosk collected more than 37,000 drawings from some 8,000 children aged between 2 and 10, creating the world's largest corpus of children's drawings and showing how object recognition changes with age[9].

Games played in the lab can also improve the robustness of data. Two of us (A.B.-L. and D.G.W.), working with psycholinguist Joe Toscano at Villanova University in Pennsylvania, wanted to study how people communicate through prosody – the patterns of intonation, pausing and rhythm that convey what information is important, as well as sarcasm and humour. The stilted conversations that participants usually produced in the lab limited these acoustic cues, and previous results on prosody were inconsistent. We designed puzzles in the game Minecraft that players had to solve by communicating while absorbed with playing. This produced more-natural dialogue and higher-quality data, showing that participants used speech intonation and duration to convey subtle information, such as whether a word was new to the conversation[10].

Compared with lab-based research, online gamified experiments make it easier and cheaper to recruit diverse participants, including those from groups that are under-represented in scientific research. In a study published last year, a gamified experiment on infant-directed speech and song recruited hundreds of people who self-identified across three gender categories and a wide range of household incomes and ethnicities in the United States[11]. More broadly, translating a gamified experiment and distributing it globally on the Internet can help to reduce the deeply English-centric nature

© 2023 Springer Nature Limited. All rights reserved.

of behavioural-science research[12].

The barriers to gamification are receding. Coding an experiment to run online used to require a lab to maintain and develop custom code. But the past few years have seen an explosion in free, open-source tools in mainstream programming languages, such as JavaScript and Python. One of the most common tools for basic human experiments on the Internet, the JavaScript library jsPsych[13], was launched in 2012 by Josh de Leeuw, a cognitive scientist at Vassar College in Poughkeepsie, New York. It has been used in nearly 1,000 published papers, according to Google Scholar.

The jsPsych tool provides a means for researchers to show their participants stimuli (text, images, audio, video); to ask questions about them; to collect responses in various ways (multiple choice, free text, clicking or tapping an image); and to output data in a structured format. Most importantly, de Leeuw maintains an active user forum and GitHub repository where researchers can discuss questions and collaborate on its codebase. This and similar tools – such as lab.js, Open Sesame and psychTestR – make it straightforward for researchers to start creating behavioural experiments online (see 'Get into gamification').

Ups and downs

Gamified experiments have clear weaknesses. Many scientists are used to having total control over their lab environments: they can observe participants' behaviour directly during experiments and check that people are who they say. Critics might be wary of losing this control, or might worry that people will not engage fully with the tests or will warp results by faking their identities, completing games multiple times or participating maliciously using Internet bots.

These criticisms can be partly assuaged. Gamified studies have the potential to engage participants better than lab-based ones can, because they are intrinsically motivating. Indeed, many psychologists have observed all manner of disengagement with lab experiments, such as participants looking at social media or even taking a quick nap, because tasks are often boring and low paid. The fact that many gamified experiments have gone viral on YouTube, Reddit and other social-media platforms provides a proof of concept that participants can become deeply engaged in research – perhaps more so than with conventional approaches, although we need more data to be sure.

Creative approaches to experimental design and data analysis can also give researchers more control over gamified science than might be expected. For example, in a music study aimed at children, cognitive scientist Courtney Hilton at the University of Auckland, New Zealand, and his colleagues (including S.A.M.) studied when participants played a game and compared the distribution of local times with the times adults played. The participants were less likely to play at night, implying that they were indeed children, who are more likely to be asleep at those times[14]. Concerns that participants could be bots can be mitigated by comparing the timing of spikes in recruitment to known events, such as a social-media post promoting a game, and by the use of manipulation checks, which are designed to be easy for a participant to answer, but hard for a scammer or bot. Ultimately, however, some research questions are better suited to the lab. The less-controlled environments used in web-based gamified science can and should complement those of lab work.

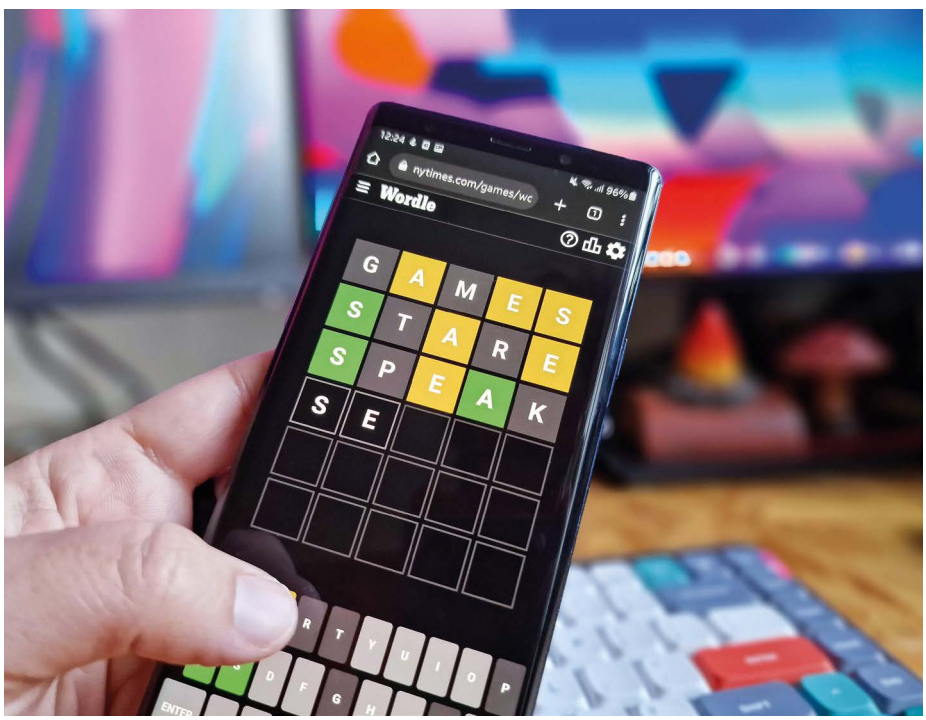
Another justified critique is that gamified science, when delivered online, leaves out people who lack smartphones, computers and Internet access. But it still offers an improvement on the status quo. Under-resourced communities are already at risk of being left out of lab experiments because labs are concentrated in a small number of places, and participants must spend time and money to visit them. By contrast, smartphones are becoming ubiquitous worldwide. In the United States, three-quarters of adults with household incomes below US$30,000 a year have a smartphone, according to a 2021 survey by the Pew Research Center in Washington DC. The fact that people from lower-income communities are more likely to connect to the Internet using smartphones than with desktop computers means it is a high priority for researchers to develop mobile-friendly experiments that work when Internet connections are slow or patchy.

Gamified experiments, especially large online ones, must also address issues concerning copyright, privacy, data storage and data accessibility. Data-protection and privacy laws differ across countries, and some research studies collect information (such as that on gender or income) that falls under protected categories. To preserve privacy and avoid legal issues, researchers can avoid collecting person-related data, such as IP addresses or other identifying information.

If data are collected by a third-party company in an existing game, the information can be proprietary. A data-use agreement is needed for research and to avoid conflicts of interest. Many universities have legal and ethical teams that can advise on gamified experiments, just as they do for lab ones. For example, when gamifying an experiment involving published music recordings, one of us (S.A.M.) received helpful guidance about US copyright law from Harvard lawyers[15].

Four future steps

For gamification to reach its full potential, four steps need to happen. First, researchers already in the community need to coordinate their development activities to widen access to technology and share technical skills. Labs should make their software publicly available, so that scientists starting out don't need to reinvent the wheel. Although open-science practices are becoming more widespread, thanks to platforms such as GitHub, Zenodo and the Open Science Framework, they are not



CLAIRE WELSH/NATURE

The game Wordle has enabled investigation of the contexts in which people cheat.

© 2023 Springer Nature Limited. All rights reserved.

yet ubiquitous. The community would benefit from pooling resources to create centralized tools and best practices for technology development, as well as running conferences and workshops aimed at increasing access to gamification and citizen-science expertise.

Second, existing tools should be improved. Many researchers use commercial software such as Qualtrics or SurveyMonkey to conduct online surveys, but no analogous software is freely available for gamified experiments. Front-end tools such as jsPsych are generally not yet optimized for mobile-friendly experiments. Back-end tools that are accessible to scientists without substantial technical expertise are uncommon – but these would jump-start scientists' ability to produce massive-scale gamified experiments, which are still relatively rare. The scientific community should work to develop these resources, and funders worldwide should make it a priority to support these efforts. The US National Science Foundation has already called for digital infrastructure projects as part of efforts to advance reproducibility (see go.nature.com/3xi6bae).

Third, scientists need to look to professions outside research to broaden their technical toolkits. Gamified experiments that produce lots of data could benefit from the use of software and practices that are standard in the technology industry but still underused in science. These include version control using Git (a way to systematically track different versions of code) and storing data in relational databases (such as those using SQL, a language for database management that is ubiquitous in tech companies). Research labs should consider consulting or recruiting artists, graphic designers, educators and others from diverse fields who can increase the robustness of experiments by improving game design. And hiring or contracting dedicated software developers ensures that technical skills are not lost when graduate students and postdocs leave a lab group. Funders should support these essential positions.

Finally, the community needs to improve the way it dispels inaccurate impressions about gamification, which by its nature risks sounding like a diversion rather than a serious research tool. We think the best way to do so is through science itself. If gamified research continues to enable substantive scientific discoveries, the results will speak for themselves. Scientists should try out gamification on their most creative, pressing and exciting research questions.

The onus is on the community to ensure that these experiments are credible and valuable. Researchers can do so by designing studies that attract deep engagement from the public, by using robust data-management practices, by adopting analytical strategies that demonstrate the validity of gamified data and by sharing data for colleagues to scrutinize. Editors and reviewers can support these efforts by encouraging the publication of high-quality experiments.

In 1986, the sociologist Lee Sproull suggested that researchers consider "a new tool for data collection – electronic mail". Decades after her work on what might be the earliest web-based study[16], collecting data on the Internet is powerful and routine. Gamified science has not yet reached such widespread acceptance, but we think it can, should and will.

Get into gamification

Using simple tools and making small tweaks to a gamified experiment's design can vastly increase engagement and expand the number of participants.

• **Take advantage of front-end tools.** Software including jsPsych, psychTestR, lab.js and OpenSesame provides an entry point to web development and gamification. Common frameworks for data collection — such as displaying a series of images and measuring the time taken to press a button — are available by default through plug-ins. A researcher gamifying a lab-based experiment needs only this, a computer and an idea.

• **Think about back-end tools.** Larger-scale experiments require cloud infrastructure that can distribute them, handle data collection and scale dynamically, so that the whole thing doesn't fall apart if your experiment goes viral. Customizable options such as Pushkin are available, although they currently require more technical expertise than do front-end tools.

• **Provide an incentive.** This could be points, scores or percentile ranks that allow participants to see how well they did, or to learn something about themselves. Pay participants a fair wage for long or onerous tasks.

• **Keep it short.** Try to restrict online games to less than 10 minutes. Some gamified experiments are even quicker.

• **Help it snowball on social media.** Make it easy for participants to share their results on social media — especially as an image or graph. This attracts more participants.

• **Use a hook.** Try adding a storyline or other form of visible progress to keep participants engaged. In one experiment on auditory perception, participants win points as they complete listening tasks to help a bird avatar stay aloft (www.themusiclab.org/quizzes/scram). Nearly 200,000 people have participated so far.

• **Visuals matter.** Putting a little effort into choosing free graphics or paying to design visuals can help to engage participants.

• **Run a pilot.** Before going public with your experiment, try it out on friends and collaborators. They'll be able to tell you what works and what could be improved.

• **Plan and test.** Ensure you have a data-management plan for handling large influxes of data (such as MongoDB or Google Cloud Firestore) if your game goes viral.

• **Collaborate and share.** Developing a project with a gamification lab can benefit everyone. If you develop new functionality, add it to collaborative codebases, as happens with jsPsych plug-ins.

The authors

Bria Long is a postdoctoral fellow in the Department of Psychology at Stanford University, California, USA. **Jan Simson** is a graduate student in the Department of Statistics at LMU Munich, Germany. **Andrés Buxó-Lugo** is an assistant professor of psychology at the University at Buffalo, New York, USA. **Duane G. Watson** is professor of cognitive science and of psychology and human development at Vanderbilt University, Nashville, Tennessee, USA. **Samuel A. Mehr** is a senior lecturer in psychology at the University of Auckland, New Zealand, a senior scientist at Haskins Laboratories, Yale University, New Haven, Connecticut, USA, and director of the citizen-science platform themusiclab.org.
e-mails: bria@stanford.edu; sam@yale.edu

1. Hartshorne, J. K., Tenenbaum, J. B. & Pinker, S. *Cognition* **177**, 263–277 (2018).
2. Coutrot, A. *et al. Nature* **604**, 104–110 (2022).
3. Levari, D. E., Gilbert, D. T. & Wilson, T. D. *Psychol. Sci.* **33**, 685–698 (2022).
4. Awad, E. *et al. Nature* **563**, 59–64 (2018).
5. Germine, L. *et al. Psychon. Bull. Rev.* **19**, 847–857 (2012).
6. Brown, K. A. *Npj Comput. Mater.* **8**, 97 (2022).
7. Wormley, A. S. & Cohen, A. B. *Perspect. Psychol. Sci.* https://doi.org/10.1177/1745691622111375 (2022).
8. Liu, J., Hilton, C. B., Bergelson, E. & Mehr, S. A. Preprint at bioRxiv https://doi.org/10.1101/2021.10.18.464888 (2021).
9. Long, B., Fan, J. E., Huey, H. Chai, Z. & Frank, M. C. Preprint at PsyArXiv https://doi.org/10.31234/osf.io/5yv7x (2021).
10. Buxó-Lugo, A., Toscano, J. C. & Watson, D. G. *Discourse Process.* **55**, 305–323 (2018).
11. Hilton, C. B. *et al. Nature Hum. Behav.* **6**, 1545–1556 (2022).
12. Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D. & Majid, A. *Trends Cogn. Sci.* **26**, 1153–1170 (2022).
13. de Leeuw, J. R. *Behav. Res. Methods* **47**, 1–12 (2015).
14. Hilton, C. B., Crowley-de Thierry, L., Yan, R., Martin, A. & Mehr, S. A. *J. Exp. Psychol. Gen.* https://doi.org/10.1037/xge0001289 (2022).
15. Mehr, S. A. *et al. Science* **366**, eaax0868 (2019).
16. Sproull, L. S. *Acad. Mgmt J.* **29**, 159–169 (1986).

The authors declare no competing interests.

© 2023 Springer Nature Limited. All rights reserved.