Remote virtual reality as a tool for increasing external validity

Dejan Draschkow[1,2†]
[1]Department of Experimental Psychology, University of Oxford, Oxford, UK
[2]Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK
[†]email: dejan.draschkow@psy.ox.ac.uk

Despite increased attention to the reliability of behavioural research, efficient solutions for conducting more representative studies are lacking. Remote virtual reality offers a promising way forward by enabling researchers to study representative behaviours in large samples and diverse settings over time, without sacrificing experimental control.

Many roadblocks stand in the way of reliable research. Meta-science efforts have focused on questionable research practices, publication biases, 'publish or perish' incentive structures, and closed science. This work has had a tremendous impact on scientific culture, but important obstacles remain unaddressed. One prominent example is the difficulty of increasing the external validity of laboratory results. External validity describes the extent to which a given finding or conclusion can be generalised to other tasks, situations, people, and times[1]. External validity is an important property of a study because general conclusions are often a goal in discovery research, and interventions derived from applied research strive to be effective across settings, people, and time.

Remote behavioural studies in virtual reality (VR) promise to be a powerful and efficient tool for achieving studies that are more representative and diverse, thereby increasing the external validity of psychological findings. Specifically, remote VR studies can address three key challenges: homogenous sampling of settings, small and homogenous samples of volunteers, and measuring changes over time.

VR diversifies settings

Researchers usually want to make conclusions about findings which hold true in different settings (where 'settings' refers to stimuli, contexts, situations, sites, tasks, and other properties of the experimental or observational environment which are not a direct property of the studied organism). For example, a researcher conducting a memory experiment using photographs as stimuli will not want their findings to be specific to the set of images selected. Similarly, a researcher developing interventions for fear of public speaking would want the effects to be transferable to the everyday situations people with fear of public speaking encounter. To generalise findings to other settings, settings need to be sampled diversely, either within the constraints of the same study (for example, as random effects or variables), or by replicating the study in other settings. Furthermore, it is important to sample settings that are representative of the target settings, that is, the setting to which the researcher wishes their findings to generalise[2]. For example, a memory researcher might actually be interested in participants' memory for realistic and immersive environments, not their memory for static, isolated photographs.

Traditional laboratory constraints can obstruct diverse and representative sampling of settings. The degrees of freedom afforded by classic tools (computer screen, keyboard and mouse) can be restrictive and it can be difficult to demonstrate that the chosen study settings are, in fact, representative of the target setting. For example, researchers have learned a great deal about short-term memory from laboratory studies in which simple to-be-remembered stimuli quickly flash on the screen and participants indicate recalled information by button-press responses. However, in many real-world settings in which short-term memory is used (such as following a recipe or a manual for assembling furniture), one's surroundings remain stable across adjacent time points; critical components of the environment must be kept in memory

because of self-initiated movements away from critical sensory information; and objects are ultimately acted on directly[3]. Previously, it was nearly impossible for research studies to do justice to these types of target settings because it would have jeopardised experimental control. Testing a cognitive system adequately under such representative demands would require full information about the environment and knowledge of, for example, participants' movements, field of view and focus of attention[4].

VR´s immersive realism and the many possibilities for response generation and interaction with the experimental environment[5] make it the ideal tool for diverse sampling of settings. In VR, the researcher has full access to participants' field of view, the experimental environment, participants' eye, hand, and head movements, and their position in space. This access to rich multivariate behavioural data enables researchers to construct representative settings without sacrificing experimental control. An impressive example of VR's maturity as a research tool is the automated VR therapy developed to treat agoraphobic avoidance and distress in patients with psychosis[6]. Here, VR is used to simulate experiences in a café, shop, doctor's surgery, and bus (diverse sampling of settings), which convincingly represent the everyday situations researchers wish to generalise to (representative of the target setting). Such an immersive study is difficult to imagine within the framework of traditional laboratory constraints and showcases how VR can be a central tool for increasing the diversity and representativeness of settings.

Remote VR diversifies samples

For a study to achieve high external validity, participants need to be representative of the target sample, that is, the sample of people to which the researcher wishes their findings to generalize. Psychology has had an ongoing struggle with obtaining diverse and representative participants as it has mostly relied on undergraduate samples from Western, educated, industrialized, rich and democratic (WEIRD) societies[7]. A popular solution for increasing sample diversity has been large-scale online testing, which offers fast results, and has the long-term potential to reach representative samples effectively[8].

Online studies conducted in VR can profit from the same advantages as regular online testing. VR technology is not a recent development, and specialised laboratories have conducted research in VR for decades. High prices and poor system quality were previously major roadblocks towards adopting VR in behavioural research. However, the consumer landscape of VR technology is transforming in a similar fashion to the rapid technological development and broad product adoption of mobile phones in the early 2000s. Major technological companies, as well as many new start-ups, have developed (and keep refining) research-grade VR systems, and they are becoming increasingly affordable to consumers and researchers. Research-grade consumer systems are projected to surpass 34 million installed units in 2024.

Critically, the advantage of reaching representative samples online interacts with VR's potential for diversifying settings. Online studies will benefit from VR's immersion and novel participant-environment interaction techniques. For the first time in remote research, researchers will be able to account for the entire visual field, gaze behaviour, locomotion and navigation. This will allow researchers to construct representative, well-controlled settings, while at the same time profiting from the scalability, fast results, and more representative samples of the online world. In the long-term, it will likely be possible to conduct VR studies and administer VR interventions and treatments online.

Remote VR affords testing over time

Time is a component of external validity that can interact with the studied settings and organisms. Sampling across diverse and representative time points ensures that findings are valid (for example, are not dependent on the time of day a measure was taken), and interventions remain effective (that is, that benefits do not dissipate with exposure to the intervention, or after the intervention is discontinued). In

remote VR, similar to traditional online studies, the same participants can be repeatedly tested from the comfort of their own home.

The benefit of repeated online access to large number of participants is further amplified by VR's potential for diversifying settings. Thanks to its millisecond precision, remote VR will enable researchers to chart changes not only over longer timescales such as minutes, hours, days, months, and years, but also over very brief periods. VR affords sub-second continuous recordings across various effectors, such as pupil diameter, gaze shifts, and movement trajectories. Thus, using remote VR researchers will be able to move beyond simple button presses by remotely capturing continuous time-resolved behavioural responses in a well-controlled environment.

Outlook

The increasing affordability and quality of VR systems will make VR more common in people's homes and more accessible to researchers around the world. VR offers immersive realism without sacrificing experimental control, and conducting VR studies remotely will provide access to more representative samples at flexible time points. These benefits can lower barriers to embracing study designs with high external validity, with long-lasting consequences for research culture.

A remaining challenge is the need for technological expertise, because even simple VR studies might require the help of experienced developers. Thus, borrowing the terminology from the Center for Open Science the next step is to develop the necessary infrastructure ('make it possible') and develop the user experience ('make it easy') for conducting remote behavioural studies in VR[9,10]. An accessible VR infrastructure for online testing will enable researcher to conduct well-controlled studies with high external validity.

Acknowledgments I thank Sage Boettcher, Kia Nobre, and Daniela Gresch for helpful conversation and comments on the manuscript. The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). The work is supported by the NIHR Oxford Health Biomedical Research Centre. The funders had no role in the decision to publish or in the preparation of the manuscript.

Competing interests

The author declares no competing interests.

References

1. Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* **54**, 297–312 (1957).

2. Holleman, G. A., Hooge, I. T. C., Kemner, C. & Hessels, R. S. The 'Real-World Approach' and Its Problems: A Critique of the Term Ecological Validity. *Frontiers in Psychology* **11**, 721 (2020).

3. Draschkow, D., Kallmayer, M. & Nobre, A. C. When Natural Behavior Engages Working Memory. *Current Biology* **31**, 869-874.e5 (2021).

4. Hayhoe, M. M. Vision and Action. *Annual Review of Vision Science* **3**, 389–413 (2017).

5. Tarr, M. J. & Warren, W. H. Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience* vol. 5 1089–1092 (2002).

6. Freeman, D. *et al.* Automated virtual reality therapy to treat agoraphobic avoidance and distress in patients with psychosis (gameChange): a multicentre, parallel-group, single-blind, randomised, controlled trial in England with mediation and moderation analyses. *The Lancet Psychiatry* **9**, 375–388 (2022).

7. Henrich, J., Heine, S. J. & Norenzayan, A. Most people are not WEIRD. *Nature 2010 466:7302* **466**, 29–29 (2010).

8. Chandler, J. J. & Paolacci, G. Lie for a Dime: When Most Prescreening Responses Are Honest but Most Study Participants Are Impostors. *http://dx.doi.org/10.1177/1948550617698203* **8**, 500–508 (2017).

9. Radiah, R. *et al.* Remote VR Studies: A Framework for Running Virtual Reality Studies Remotely Via Participant-Owned HMDs. *ACM Transactions on Computer-Human Interaction (TOCHI)* **28**, (2021).

10. Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M. & Mushtaq, F. Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods* **52**, 455 (2020).