PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/214410

Please be advised that this information was generated on 2021-01-02 and may be subject to change.
Reducing the Noise of Reality

Abele Michela, Marieke M. J. W. van Rooij, Floris Klumpers, Jacobien M. van Peer, Karin Roelofs & Isabela Granic

To cite this article: Abele Michela, Marieke M. J. W. van Rooij, Floris Klumpers, Jacobien M. van Peer, Karin Roelofs & Isabela Granic (2019) Reducing the Noise of Reality, Psychological Inquiry, 30:4, 203-210, DOI: 10.1080/1047840X.2019.1693872

To link to this article: https://doi.org/10.1080/1047840X.2019.1693872

© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.

Published online: 04 Jan 2020.
Reducing the Noise of Reality

Abele Michela\(^a\), Marieke M. J. W. van Rooij\(^b\), Floris Klumpers\(^a,b\), Jacobien M. van Peer\(^a\), Karin Roelofs\(^a,b\), and Isabela Granic\(^a\)

\(^a\)Behavioral Science Institute, Radboud University Nijmegen, Nijmegen, The Netherlands; \(^b\)Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

A Commentary On: Causal Inference in Generalizable Environments

Miller and colleagues (this issue) propose a novel and inspiring theoretical framework that merges systematic and representative experimental design. As such, the Systematic Representative Design (SRD) framework has the potential to move psychological science forward in a significant way. Essential to the authors’ proposition is the default control group (DCG), an experimental condition that leverages the possibilities of new technologies such as Virtual Reality (VR) in order to create a close approximation to “real-life,” but with the affordances of a tightly controlled experience. With this innovation, Miller and colleagues attempt to solve the incongruous needs, common in so many experimental designs, between generalizability (to everyday life) and experimental control (to claim causality).

Although the SRD framework could potentially result in a real shift in the way psychological research is conducted, it provides far from an ‘off the shelf’ solution. In particular the use of new technologies such as VR brings with it a number of complications - of both practical and theoretical nature – that are not fully addressed in the target article. With this commentary, we aim to contribute to the discussion on improving experimental design and thus empirical psychological research in general, by drawing from our experience in designing digital, game and neurofeedback based interventions for mental health and behavioral change. Furthermore, we will suggest that there are important process similarities between SRD and game design, as well as common practical pitfalls. We will outline these links and discuss their implications in order to potentially further strengthen the SRD framework, especially for applied research purposes.

The Potential of Systematic Representative Design

SRD is a promising framework that attempts to address, and provide solutions for, both causality and generalizability requirements in experimental design. This framework allows us to maintain a high level of experimental control in environments that are usually impossible to standardize without introducing a lot of noise in the measurements. Specifically, the strength and main novelty of the SRD approach is that it offers experimental control through a carefully constructed default control group (DCG) that serves as a highly controllable virtual substitute to reality.

The first benefit of a highly controllable, yet generalizable, environment in SRD is the reduction of noise and biases in the data that are collected in the testing environment. Specifically, the authors of the target article suggest that VR, unlike the typically austere psychology laboratory, allows for the design of an experimental environment that can mimic the natural, complex context in which behaviors of interest appear. In doing so, the VR experimental design limits some of the experimental biases (e.g. due to different motivations, semantic understanding, cultural or social expectations) that otherwise emerge from studying participants acting out-of-the-ordinary in traditional lab contexts (e.g. Ceci, Kahan, & Braman, 2010). Moreover, in studies focusing on social interaction, the opportunity to use Virtual Agents (VA) in the VR setup helps researchers overcome the low internal validity and increased noise that is often introduced when human confederates are used (Kuhlen & Brennan, 2013).

On top of improving the reliability of the data being collected compared to naturalistic settings by increasing the consistency of the environment, SRD also aims to focus more precisely on the motivational core of participants’ behaviors, keeping that core as consistent as possible across the full sample of participants. A large problem in “controlled” laboratory assessments come from the range of unwanted individual differences that participants bring with them to a lab task (e.g. motivations, interpretations of the instructions, past experience with similar tasks, expectations, and so on). The SRD framework insists on taking these individual differences seriously and reducing the noise they introduce in experiments, by emphasizing the meaning of the action performed. Instead of relying on participants simply following task instructions, SRD experiments use contextualization (what is happening?) and interactive narrative (why is it happening and what should I do?) to prime participants to act consistently according to the role they are given. Thus, in participants’ experience, they behave more in accordance with their natural internal motivation rather than the instructions given.
than the external motivation of complying with the experimenter. They have a reason to act as they are asked to. Indeed, “providing a role”, has been shown to change behavior in simple setups like the ultimatum game, where being primed to imagine impersonating a banker significantly changed the way participants behaved (Lightner, Barclay, Hagen, & Hagen, 2017). In addition, the narrative context makes action matter. Stimuli become affectively salient and motivationally important as a result of these contextual enhancements (see Parsons, 2015 for a review). In many research contexts, the increases in motivation and affective salience of stimuli should not only give rise to a better approximation of reality but also improve the reliability with which these effects are captured.

**Drawbacks and Challenges**

Despite SRD being a sound framework from a theoretical perspective, there are important pitfalls that we anticipate researchers will face when they attempt to apply the framework to their own experimental work. We experienced many of these same pitfalls in our own research, in which we strove to design a suitable VR environment for studying decision-making under stress in police officers. We will review a number of these pitfalls both from an implementation and an ethical perspective, but also provide specific examples from our own research that illustrate the concrete challenges. Suggestions for solutions that we have found to address these challenges will be explained in the “proposed solutions” section.

**Implementation Concerns**

One important aim of the SRD approach that is emphasized by the authors is the need for generalizability to everyday life. This concept resembles the more widespread concept of ecological validity, and will be used interchangeably in our discussions. Ecological validity can be defined as a combination of verisimilitude and veridicality (Franzen & Wilhelm, 1996). Verisimilitude is the level of believability or the extent to which an experimental task approximates the features of everyday life. Veridicality, in contrast, is the degree to which the performance of a participant in an experimental setup accurately predicts what that person would do in reality.

**Enhancing Verisimilitude**

Enhancing the verisimilitude in an experimental setup can be deceptively appealing, but may introduce several problems. In an SRD context, enhancing verisimilitude requires isolating a target behavior, identifying the “most frequent setting” in which that “behavior of interest” appears and also the “relevant script components”, and so forth. In other words, an SRD experiment working towards maximal verisimilitude may seem to aim at approximating a simple snapshot of reality - faithfully reproduced in VR and/or using by using VA. However designing this type of VR environment leads to several problems of feasibility that need to be considered even before an SRD approach is taken, especially in VR.

First, among the most common drawbacks of VR are nausea and motion sickness: Up to 80% of participants are impacted by these physical symptoms in VR experiments (Stanney, Hale, Nahmens, & Kennedy, 2003). Most concerning for VR design using an SRD approach, nausea and motion sickness are most often a problem when the aim of the VR design is to elicit “natural behaviors.” More specifically, only a relatively small part of the population cannot physically tolerate VR that simply reacts to the natural movements of the player (e.g. an object grows bigger in order to appear closer when the user bends forward towards that object), but the proportion gets much larger when unpredictable or non-user-initiated artificial movement is experienced (e.g. a car hits a wall unexpectedly; e.g. Stanney & Hash, 1998). For example car simulation games in VR (like Project Cars, https://www.projectcarsgame.com/vr/) minimize this issue by providing wide tracks, a predictable trajectory, a reduced feeling of braking and frontward gaze fixation points. Considering the success of such games, VR could seem like a perfect environment to a researcher investigating, for example, reckless driving. The virtual environment allows for highly controlled measures to take place, and even very “risky” situations to be safely re-created. Yet, the nausea-reducing measures present in VR driving games cannot be used in most setups aiming to enhance verisimilitude, as exemplified by driving in a city: it happens in narrow tracks, frequent braking and the need to constantly monitor the surroundings, which results in gaze diversion from frontward fixation. Thus, with the current technology, enhanced verisimilitude increases unwanted nausea effects, which paradoxically, reduce the generalizability of the task, because nausea can be so unpleasant.

A second drawback with the pursuit of verisimilitude is the uncanny valley effect (Mori, 1970), which is the feeling of eeriness or revulsion experienced while interacting with robotic or virtual avatars that mimic human behavior almost, but not exactly, perfectly. In the pursuit of ecological validity, researchers tempted to faithfully reproduce a snapshot of reality have a higher risk to encounter this effect than researchers limiting themselves to simpler, less realistic, stimuli. Research on this uncanny valley effect lacks consistency regarding the prevalence and explanation of the effect (Cheetham, 2017; Lay, Brace, Pike, & Pollick, 2016; Wang & Rochat, 2017). However, it may be that the effect is related to a discrepancy between expectations raised by an anthropomorphic “entity” and its observed behavior (Zlotowski et al., 2015). Therefore, in the current state of most available technologies, the uncanny valley effect can pose a severe threat to the ecological validity of a task, as participants who feel disturbed by the perceived un-realism of the experimental setup are less likely to behave naturally.

A last, more mundane, drawback regards the limited availability of adaptable and evolved VA and complex interactive environments whether in VR or not. Their development requires large financial investments and might require niche programming expertise. These high costs and long
development times are often overlooked in grant proposals and research designs, whilst being of paramount importance for the success of any project aiming at using VA to enhance ecological validity.

Aside from the implementation concerns, enhancing verisimilitude by approximating a snapshot of reality also brings with it analytic concerns: Given the complexity and richness of VR environments, how valid are our conventional analytic methods? Most of the tests used in traditional laboratory experiments rely on specific assumptions that are easily violated in VR setups that attempt to simulate reality as closely as possible. Human behavior is complex, occurring at different levels of analysis (perception, attention, interpretation, action), and changing over time (i.e. moment-to-moment). The methodological challenge of capturing this complexity and dynamic nature is illustrated by Brehmer’s (1992) early research that attempted to re-create a generalizable, ecologically valid decision-making environment for firefighters. In his setup, a series of interdependent decisions had to be made in real-time in an environment that changed both spontaneously and as a consequence of earlier decisions. Since all the decisions made by a player were interdependent, the standard analytic strategy which aggregates all decisions across time and contexts was not feasible anymore because the assumption of independence was violated in this setup. This is why Brehmer’s focus had to move to the general tactics and strategies used by the firefighters to achieve their pursued goal rather than momentary decisions. The same could happen in SRD experiments: When broadly reproducing real-life situations in an SRD setup, actions and decisions will be embedded hierarchically, they will influence one another, and they will be contingent on prior actions and decisions. As a result, individual actions or decisions are not the correct unit of measurement and aggregating those measurements (decisions) violate most General Linear Model assumptions. More sophisticated hierarchical analysis approaches could be envisioned, but without severe constraints on the participant’s action range those analyses would quickly grow out of hand. In other words, if an SRD experiment consists of a VR-based snapshot of reality, it is more like an interactive “video” of reality than a series of independent snapshots. Thus that design does not allow analyzing single action or decision moments in isolation. Taken together, the lack of independence of measurement points prevents fundamental mechanistic questions to be addressed adequately in such a high-fidelity reproduction of reality.

**Enhancing Veridicality**

Enhancing the veridicality of an experimental setup raises another set of concerns regarding the way the data can be analyzed. Since veridicality is defined as the degree of prediction of the experimental setup on participants’ everyday behavior, it seems meaningful to enhance veridicality by incorporating “real life” elements into a validated experimental setup. It actually is the only principled way to proceed at enhancing the prediction power of a laboratory setup on everyday life’s behavior. Yet, sadly enough, every element of “reality” added introduces proportional levels of complexity. No level of technological sophistication will solve this problem; in creating a close to real-world environment, one gets the corresponding real-world complexity for free. In a situation comparable to “real life”, behaviors can have wildly different explanations, or be due to the interaction of a large number of sub-systems. Therefore, by making an experimental setup richer to enhance veridicality, the researcher might not be able to exclude potential alternative explanations of a measured effect, especially when trying to make an inference about the underlying mechanisms. Effects that are established in a controlled environment could also disappear altogether in this more complex environment.

A concrete example from our own work might be important to clarify our point. In an early iteration of an experiment on decision making under stress for police officers, we attempted creating an ecologically valid version of the laboratory shooter task standardized by Gladwin, Hashemi, van Ast & Roelofs (2016). This task requires (police) participants to perform several trials of shoot-don’t shoot decisions. The main effect we desired to reproduce from that task was a reaction time difference in shooting responses between high and low threat conditions, as found by Hashemi et al. (2019). Reaction times are measured as the latency between a target stimulus appearing on a screen and the recorded response from the participant. In the laboratory task the target stimuli appear instantaneously (an opponent with a gun or phone), thus allowing for a precise measurement of the time until the participant responds, with millisecond precision. In an ecologically valid VR setup however, even target stimuli with a seemingly sharp onset like someone opening a door or taking an item out of their pocket are not instantaneous enough for a precise reaction time measurement. In addition, the time duration of those very actions actually lasts much longer than the differences in reaction times that we wished to measure between conditions. As if that was not enough, since VR allows participants to look in any direction, target stimuli could be missed altogether. Hence, the effects of threat on response times observed in the laboratory task could disappear in the VR environment.

Also, when enhancing veridicality by adding “reality” to an established experimental setup, the problem of trial duration can also become a challenge. This concern can again be illustrated by our attempts to create an ecologically valid version of the shooter task mentioned above. In this task, participants were asked to perform 150 trials of shoot-don’t shoot decision making. The task can be completed in about an hour and contains enough trials to reliably measure the within-condition effect (i.e. high versus low threat). In this set-up, it is virtually impossible to include contextual elements to make the decision similar to what police encounter in “real life”. To have an idea of what a more ecologically valid example could be and what would be the implications for task duration, let us consider the task used by Johnson et al. (2014). In this task, deadly use of force decisions were inspired by real situations from police practice and trials lasted an average of two minutes. Transforming the shooter
task by Gladwin and colleagues according to Johnson’s example, while keeping the same number of repetitions to reliably detect the effect, would make it last almost five hours (without breaks). Thus, because of feasibility reasons, the effects reported in Gladwin and colleagues’ actual experiment may not have been discovered, and could probably not be replicated, in a setup characterized by higher veridicality. This difference raises the question of the number of repetitions allowed by an SRD experiment relying on contextualization to elicit affective reactions, and therefore the effect size needed to reliably measure an effect with very few trials.

**Ethical Concerns**

An extension to the implementation concerns we have outlined above are those related to what is ethically feasible in experimental setups. The effort to make an environment ecologically valid and able to elicit genuine behaviors of interest may raise a range of ethical concerns that may not apply to standardized, controlled studies. Specifically, as outlined by Pan and Hamilton (2018), there are several ethical risks that appear in immersive environments and interaction with VA: Enhanced personal disclosure can lead to privacy issues (Lucas, Gratch, King, & Morency, 2014; Rizzo et al., 2015), immersive environments could lead to changes in real-life behaviors through embodiment (Tajadura-Jiménez, Banakou, Bianchi-Berthouze, & Slater, 2017) and even false memories in children (Segovia & Bailenson, 2009).

As hinted by these studies, VR experiments could be considered to be emotionally charged environments because users are not only immersed in a story, with vivid graphics. It is also due to one’s whole body – including gaze, body posture, physiological arousal, and so on – being directly impacted on by this enclosed, immersed simulation. As a result, emotional experiences and associated cognitive impacts may have long-lasting effects. Those effects might linger far longer than the actual VR experience itself and generalize outside the context of the experiment. Consequently, these studies suggest the potential risk of accidentally creating traumatic experiences in VR experiments that simply aim to assess behavior. Clearly, the point of using VR in a study is to increase “immersion”, which in turn should elicit authentic emotional responses. If the responses are indeed authentic emotional experiences, they have the potential to impact the well-being of participants, as suggested by the use of VR for mood induction protocols (Baños et al., 2006), and the potential of those induction procedures to be used in clinical contexts (Herrero, García-Palacios, Castilla, Molinari, & Botella, 2014).

Moreover, the use of VR as a therapeutic intervention tool is a strong indication that it requires extra attention to potential side effects. Indeed VR is a promising tool used in the treatment of several disorders such as Post Traumatic Stress Disorder (PTSD; Rizzo & Shilling, 2017), complicated grief (Botella, Osma, Palacios, Guillén, & Baños, 2008), eating disorder (de Carvalho, Dias, Duchesne, Nardi, & Appolinario, 2017), and sexual disorders (Fromberger, Jordan, & Müller, 2018). Most of the leading scientists in those fields advise the use of extra care in considering the ethical issues specifically linked to VR. A comprehensive example of ethical guidelines can be found in the article of Madary and Metzinger (2016), where potential implications of the use of VR are reviewed extensively.

In our own work, we had similar issues around ethically designing an emotionally evocative task in VR, with high levels of ecological validity. In the previously described VR project, aimed at training decision making under stress for police officers, situations arose in which police participants could involuntarily shoot innocent bystanders. These situations seemed relatively benign and game-like to the VR simulation designers. However, our stakeholder, the Dutch Police Academy, quickly vetoed this training scenario because of the concern for triggering traumatic memories and the potential for desensitizing officers or reinforcing shooting behaviors that were against their training policies. Thus, it is important to keep these ethical considerations in mind from the start when designing VR tasks from an SRD perspective, to avoid creating unsuitable (and costly re-) designs.

**Proposed Solutions**

To overcome the previously mentioned ethical and implementation challenges in the development of SRD experiments, we propose two main directions. The first is to extend the range of techniques proposed in the target article to experimental manipulations that include using powerful tools from neuroscience. We agree that techniques proposed by the authors like fMRI neurofeedback are a good start, but we see substantial additional advantages in terms of feasibility and opportunity for making causal claims from EEG neurofeedback applications, as well as brain stimulation techniques and other psychophysiological methods. Indeed, those techniques allow for a better use of VR capacities, by preserving the participant’s head movement freedom. The second direction that we propose is the application of game-based approaches to SRD paradigms. We argue that it may be important to reconsider the importance of general verisimilitude and think about which specific elements of everyday life are necessary to claim generalizability. We suggest that less realistic game-based elements can often prove more effective in homing in on the core causal units necessary for generalizability claims.

**Extending the Use of SRD to Other Methods**

One of the core promises of SRD is to allow for causal inferences by isolating mechanisms by testing the experimental condition against a default control group (DCG). Providing the experimental condition by modifying the DCG by changing parts of the VR environment or of the VA can be relatively straightforward for a certain number of applications (like changing the gender of an interacting virtual avatar). However, as previously mentioned, for more fundamental questions (e.g., such as investigating the neural underpinning of specific decision making processes),
difficulty can grow exponentially fast. This increased difficulty is mainly due to the need to exclude alternative explanations for observed effects and to make causality claims, which is what has historically driven experiments away from ecological validity (see Burgess et al. (2006), for a narrative review). One interesting way to avoid or at least address these difficulties could come from the fields of neurofeedback, biofeedback and brain stimulation. Indeed, these techniques can provide a way to manipulate physiological parameters - and can therefore allow causal inferences - by being used to create an experimental condition without modifying the VR environment or VA behaviors.

Miller and colleagues alluded to neurofeedback as a field of applications that would benefit from the SRD framework. We agree, but argue that this benefit could go both ways, as neurofeedback has been used not only for interventions, but also as a tool to perform experimental manipulations in fundamental research (see Sitaram et al., 2017 for a review). Moreover, contrary to the authors' suggestion that fMRI neurofeedback is ideal for incorporating into SRD designs, we suggest that there is a wide range of opportunities offered by the more VR-friendly EEG neurofeedback (as exemplified by the work of Vourvopoulos et al., 2019). This latter technique has already proven its potential in investigating fundamental neuroscientific questions, like the trainability of brain plasticity (Ros, Munneke, Ruge, Gruzelier, & Rothwell, 2010), EEG biomarker’s connection to psychopathological symptoms (Ros, Baars, Lanius, & Vuilleumier, 2014), the normalization of scale-free dynamics in EEG in (Ros et al., 2016) and research on stroke patients rehabilitation (Ros et al., 2017). Moreover, the spatial resolution of EEG neurofeedback has recently been extended by using machine learning to extract deeper, non-cortical, signals like those from the amygdala (as elegantly shown by Keynan et al., 2019).

Specific brain activity could therefore be modulated either endogenously with neurofeedback or externally with brain-stimulation techniques, in controlled SRD setups. These possibilities would remove the chore of providing an experimental condition from the VR environment design. In other words, in such an experiment the VR and VA components would be constant across the control and experimental group, and the experimental manipulation would happen in terms of brain training or stimulation only. Experimenters could therefore perform targeted mechanistic manipulation of brain activity on participants interacting with a generalizable environment. If the VR environment was built in such a way that it was, indeed, reliably eliciting the behavior of interest, it would allow direct claims to be made regarding the implication of specific brain activity patterns in everyday functioning. One example could be studying the effect of disrupting (or training) posterior alpha oscillations, which has been linked to visual attention (Rihs, Michel, & Thut, 2007) in ecologically valid VR contexts. This could allow researchers to link the phase-amplitude coupling of the alpha-gamma waves (Pascucci, Hervais-Adelman, & Plomp, 2018) to behaviors generalizable to “real life”, which could in turn prove very useful to better understand the neural underpinning of the daily attentional processes.

Finally, the range of techniques used to create the experimental condition in SRD could be further expanded to biofeedback protocols controlling non-cerebral psychophysiological markers. Biofeedback could be used to investigate well-established psychophysiological effects in environments that provide a higher degree of generalizability to everyday life than traditional experimental designs. A few examples could be studying (through training) the role of easily measurable markers like heart-rate variability for stress management (Yu, Funk, Hu, Wang, & Feijs, 2018) or anticipatory bradycardia for decision making (Roelofs, 2017). In these two specific examples, studying the contribution of well-established psychophysiological markers in generalizable contexts would pave the way for a wide range of affordable interventions aiming at changing, for example, the behavior of patients suffering from anxiety disorders.

**Careful Integration of Reality: Make It a Game!**

We have argued that there are a host of potential pitfalls in designing SRD studies that aspire to maximize verisimilitude and veridicality, with the goal of increasing generalizability of research results and making stronger causal claims. An alternative and more promising approach to create an SRD experiment could come from the game design world. As we know, building the DCG requires isolating the core causal units needed for that generalizability. As long as these core units are maintained in the design, we may reduce realism in our VR simulations and still elicit the behavior of interest. We suggest that elements common in commercial games can offer such a solution by making the environment not realistic, but rather believable (Schubert, 2013). Where realism is achieved by faithfully reproducing reality, believability is achieved by engaging the player through several game mechanisms, like challenge or emotional narratives.

When Miller et al. suggest that an SRD experiment could look like “serious games” they run the risk of directing the player toward the pursuit of realism instead of believability. Serious games are developed for educational interventions (e.g. skill training) and even if they usually attempt to include a “fun” component, it does not usually compare with the entertainment value of commercial games (Baranowski et al., 2016). We propose that using conventional gamification techniques incorporated in serious games will not be enough to obtain generalizability, largely because serious games are usually simplistic simulations that pay little attention to the emotional, engagement, and motivational underpinnings of player’s experience (Schoneveld, Lichtwarck-Aschoff, & Granic, 2019). This is why we think that serious games do not offer a solution for many applications of the SRD approach. However, we do advocate for game-based approaches that include believable core concerns of participants’ motivations and engagement that are fundamentally embedded in an emotionally relevant context.

To illustrate the difference between a realistic “serious” game that feels artificial and a believable one that is
experienced as relevant and motivationally engaging, we make the analogy of an SRD “serious game” being like an empty plate on which we attempt to grow a bunch of human cells, while a believable game is a petri dish containing all the essential nutrients required for those cells to grow. The cells on the petri dish survive, whereas the cells on the empty plate quickly die. Similarly in an SRD experiment, in many cases the behavior of interest in a study could not “survive” in the DCG if the context does not provide the correct motivational, affective, and engaging conditions. A successful SRD experiment relies on a tradeoff between approximating real-life conditions (by using VR and VA in a believable way) while accommodating the scientific imperatives. We argue that providing motivational, affective and engaging elements that elicit ecologically valid behaviors in participants can be best achieved by making it into an engaging and entertaining (not “serious”) game. After all, what makes a good game greatly overlaps with the needs of SRD: An immersive experience with an interactive narrative that elicits genuine affective emotions in the participant.

Importantly, game-based interventions can provide solutions for some of the potential pitfalls of SRD studies we mentioned earlier: The narrative scaffolding directs participants’ attention to what is at the core of the experimental goals and narrows the spectrum of actions observable in the experimental setup. Moreover, a high quality narrative can also provide a believable reason for many repetitions of any particular behavior, an otherwise key potential pitfall we mentioned earlier in SRD studies. Using real games as a reference can also mitigate the ethical concerns by reducing the risk of learning transfer of unwanted aspects of the simulation. In turn, a well-designed game setup can also greatly reduce the technical difficulties like the uncanny valley effect in the development of the SRD task, thus increasing the implementation potential and generalizability to everyday life.

The final benefit of applying SRD in the form of real games, based on our own experiences in the previously described police project, is that it forces the researchers to adopt a more iterative stance in their design process, which is crucial for a successful experimental setup using VR. Working closely with game designers often means that researchers need to (at least partially) adopt design thinking principles in their development process (Scholten & Granic, 2019). Creating a suitable VR environment – with an immersive narrative and emotional experience – often requires more than just handing a list of requirements to game designers. These designers should instead be included in the experimental design process very early on, and help the researchers develop the task in an iterative procedure that will invariably challenge the scientist’s original assumptions. Indeed, as the complexity of an SRD experimental setup - sing VR, VA, and potentially other technologies combined - is orders of magnitude higher than traditional experiments, the design process may often resemble the ones used in commercial game companies. This process, called the Rapid Iterative Testing and Evaluation (RITE) method, has been advocated for more than a decade (Wixon, 2003) and can prove useful for testing assumptions and design choices in small consecutive steps, instead of designing and testing the full complex setup at once. It requires the early involvement of stakeholders, to test and refine each prototype until reaching combined scientific and design goals. This is a time consuming endeavor, but as Miller and colleagues themselves implied, isolating the behavior of interest and constructing a DCG are arduous tasks and require a very flexible mindset to be achieved.

Conclusions

The article by Miller and colleagues (this issue) suggests that the needed shift in the way we study psychology can be provided by Systematic Representative Design (SRD). We agree with them and hope that our commentary will help researchers undertaking this endeavor to be aware of the potential challenges of designing an SRD-informed study. More than a cautionary note, we aimed to raise issues about the conceptual conversion cost involved. The straightforward use of Virtual Reality and Virtual Agents to increase ecological validity by simply adding “reality” to a controlled experiment might rarely pay off due to feasibility issues, ethical complications and increased complexity in analysis methods. Our first suggestion is to widen the range of techniques used to create the experimental condition to EEG neurofeedback, biofeedback and brain stimulation. Our second suggestion is to draw inspiration from the game design world when designing an SRD experiment, both for design process and final result, and aim at creating believable environments instead of realistic ones. If these kinds of cross-disciplinary approaches are taken, we are convinced that SRD will lead to genuinely novel psychological research protocols that address many of the problems of generalizability that the field has suffered from for decades.

Funding

This work was supported by National Institute of General Medical Sciences; National Institute of Mental Health; National Institute on Drug Abuse.

References


