Correcting Course: Six Routes to a Productive Future of Psychological Games Research

Nick Ballou[1]

Abstract
This paper describes the need for individualized models of gaming effects.

Keywords
TODO

Introduction

Decades of research has investigated whether games are associated with, and ultimately cause, a variety of positive and negative outcomes. These potential effects are wide-ranging, with researchers having sought to understand whether violent video games cause aggression, action games have benefits for executive function, disordered or "addicted" gaming harms players or constitutes a mental disorder, and much more.

After monumental effort exerted to understand these topics, however, little is clear. Effects are nuanced, inconsistent, and generally small in magnitude. In each of these examples, a plunge into the literature reveals deeply divided opinions about the validity or importance of these effects.

In this paper, I propose 6 areas in which research can be improved.

This paper is informed largely by my own experience working on media effects research, which at its crudest conceives of games as a 'dose' that leads to a particular 'response', be it well-being, pro- or antisocial behavior, or any number of other outcomes of interest. However, the benefits of the strategies I describe are not limited solely to video games effects research, but rather any work that uses psychological methods to understand how players interact with games.

[1] Queen Mary University of London

Corresponding author:
Nick Ballou, Game AI Group, Queen Mary University of London, London, UK, E1 4NS
Email: nick.ballou@gmail.com

A Brief History of Inconsistent Effects

Before proposing solutions, it is worth reviewing the history of some key debates that have shaped the landscape of psychological research on games. The following examples are not exhaustive, but serve to illustrate how despite enormous research investment in particular topics, certain questions are yet to be satisfactorily resolved.

Violent Video Games and Aggression

Among the earliest topics of interest in psychological games research was whether games cause real-world aggression or violence. After nearly 4 decades of research attention, these debates continue to smolder. Findings supportive of a link between violent video game play and subsequent aggressive thoughts or behaviors continue to proliferate. These findings are regularly invoked when violent games are cited as a possible cause of violent tragedies like school shootings—particularly when the perpetrator is white (Markey et al. 2020). Other studies find null effects (Przybylski and Weinstein 2019), argue that statistically significant effects are likely not practically significant, or question whether laboratory measures of aggression are valid indicators of real-world violence even if meaningful effects were confirmed (Drummond et al. 2018).

Even meta-analysis has proved incapable of fully resolving these differences. A series of "warring" meta-analyses have reached radically different conclusions (Anderson et al. 2010; Ferguson 2015; Hilgard et al. 2017b). In a polarizing issue, the middle-ground opinion seems to be that evidence points to a significant, but small and potentially negligible effect (Mathur and Vander-Weele 2019). However, while meta-analyses disagree about the degree of publication bias present, the small number of preregistered studies on the topic have tended to find null effects (Ferguson 2020).

Dysregulated gaming

At the time of writing, the most active area of research on games concerns dysregulated play (variously conceptualized as [internet] gaming disorder, pathological gaming, problematic video game use, video game addiction, and more)—with over 700 papers containing one of those terms in the abstract published in 2021 alone. There is widespread agreement that a small fraction of players experience negative life consequences in relation to their gaming CITE.

There is much less agreement about whether dysregulated gaming is a symptom of other mental health conditions or a cause of them (or both) (Weinstein et al. 2017), whether the current approaches to measurement and diagnosis are appropriate (Ballou and Zendle 2021; Aarseth et al. 2017), and the degree of harm that formalizing gaming-related disorders could have for healthy players (Aarseth et al. 2017). As in other cases, we have a poor understanding of the behavioral patterns of dysregulated players: what they do in-game, when and for how long they play, and how these behaviors develop over time.

With regard to potential positive effects, some research suggests that video games, in particular action games, can have cognitive benefits such as improved executive function (Bediou et al. 2018). However, even at the meta-analytic level, findings are inconsistent and contentious (for evidence in favor, see Bediou et al. 2018; for conflicting evidence, see Hilgard et al. 2019b).

Other Examples

Similar trends exist in the video games and well-being literature. As above, some studies find positive associations between video game play and well-being CITE, some negative associations CITE, and some no significant association at all CITE.

executive function and attention - tom stafford?
games and learning literature

While the above examples are the most prominent topics in the psychological games literature, they are by no means the only ones to fail to reach firm conclusions. Researchers continue to debate topics such as whether exposure to sexualized (female) characters affects body image (Lindner et al. 2020), the contexts in which gamification is and is not successful

There is reason to suspect that beliefs about these various debates differ due to a variety of factors beyond simply the strength of evidence itself, including for example age, familiarity with commercial video games, and field of study—both for researchers (Klecka et al. 2021) and laypeople (Przybylski 2014).

In short, in all these cases and more, results have been definitively mixed. In many cases, we appear to have to reaching meaningful consensus on anything but the most basic of questions, and it appears unlikely that current approaches to research will settle these debates in the foreseeable future.

A Path Forward

It should be hopefully clear by now that the current status quo of psychological games research is at best inefficient, and at worst incapable of answering the key questions that parents, players, and policymakers have about the growing role of video games in our lives.

To be clear, this is not an argument for both-sidesism; in many of these cases, there is a pattern of more rigorous, larger sample, preregistered studies finding smaller effects. Neither, however, does this mean that all evidence for negative effects of games is invalid or the result of false positives. While it is undoubtedly true that elements of moral panic have crept into debates around games and digital technology use more broadly (Orben 2020; Markey and Ferguson 2017), acknowledging this fact does not mean that no harm, or no meaningful harm, is possible as a result of video game play.

Instead, I believe what we require is a reframing of the questions researchers are posing, the methods used to answer them, and the context in which this takes place. In this paper, I propose 6 ways that research on video games can be improved, and begin to resolve these debates. ==finish==

(Principled) Person-Specific Effects

There is increasing recognition that small, nuanced, and inconsistent effects in a general population may hide substantial variation in effects for any given individual (Johannes et al. 2021a). While most people may not be affected by violent video games, for example, some small subset may exhibit a positive or negative relationship between their exposure to violence in games and subsequent aggression.

Higher-quality evidence has indeed largely converged upon small estimates of video game effects, in some cases concluding that such effects are too small to matter at the population level. If, however, one portion of the sample is exhibiting a negative effect of, say, video games on well-being, while another portion exhibits a positive effect, the standard statistical toolkit will show that there is no effect. Such findings are valuable—a null effect of this nature provides evidence against the effectiveness of a global playtime restriction, for example—but are very different from the finding "for any given player, there is no effect of video game on well-being".

Person-specific approaches seek to understand how effects vary across people. This represents a substantial shift in approach, and suggests that researchers are often asking the wrong questions.

In other words, it may not be productive—or even tractable—to ask "do video games cause X?" in many cases. Instead, a variety of research questions would be better framed as "For whom, and under what circumstances, do video games cause X?"

==situation-level effects - more important? e.g., situation where you want distraction because you're==

bored, then turning to facebook, explains negative outcomes related to facebook

Social media research has begun to use person-specific analysis to promising effect. Verbeij et al. (2022) found that read paper and then finish sentence.

Despite being a close cousin of social media research, psychological games research has not widely adopted this analytical strategy.

It is important that person-specific research maintains as its aim the goal of understanding groups, and generalizing from sample to population. A recent paper proposes a framework, in part inspired by medical research, to unify these aims (Johannes et al. 2021a). Emphasizing that effects in the social sciences will always vary to a certain extent from individual to individual, the authors suggest specifying a 'region of practical equivalence', or ROPE, that differentiates *meaningfully large* variation from *unimportant* variation. Where variation in effects is meaningfully large, this warrants further examination, often in the form of more clearly identifying moderators that influence the strength of this effect for any one individual.

More broadly, I agree with arguments that person-specific and population-level approaches are not incompatible (Valkenburg et al. 2022). Instead, the focus in person-specific analyses should be on estimating variability in effects, and then uncovering the moderators that (in part) determine who experiences more positive relationships between media use and a certain outcome, and who experiences more negative ones (Parry et al. 2022).

Strengthening the Theoretical Derivation Chain

Many readers will recognize that theories currently used in psychological games research are typically underspecified to make the kinds of person-specific predictions discussed above. In many cases, the theories we use are underspecified even to make clear predictions on the level of the population, with substantial room for interpretation in measurement, the type of statistical model (are all effects linear?), the population, the size of the effect, and more. Prominent theories used to understand video game play are often specified at the level of broad, verbal constructs, and these constructs are often few in number. example? Even more concerning, theories are often used superficially, as background context or plausible explanations separate from the data at hand (Tyack and Mekler 2020, see).

This is a problem for our ability to make generalized predictions: what good is understanding the effect of a single video game if we have no framework to apply it to a different context?

The majority of theories in games research—and the social sciences more broadly—are underspecified to make predictions at the level of the individual CITE. We may set out to disprove that the effect in the population is 0 (which, of course, it never is exactly; Cohen 1990), but very rarely do we set out to test what proportion of people, or what particular individuals, will exhibit a particular effect.

Broadly, to solve this problem we need to strengthen the so-called derivation chain: a conjunction of theoretical and auxiliary premises that are necessary to predict observable outcomes, capable of leading to a highly specific statistical prediction (Meehl 1990).

Scheel et al. (2021) provide a useful overview of research activities that can strength the derivation chain. Among the most important of these in my opinion will be greater use and appreciation of descriptive work and exploratory data analysis. Given our current lack of knowledge about boundary conditions, auxiliary hypotheses, and other aspects of the derivation chain,

Until this course of research is more well-developed, it is likely that a (purely) confirmatory approach would be ill-advised—at the moment, we simply lack nuanced enough theory to make predictions that could be severely tested. Instead,

we will need to emphasize the value of descriptive and exploratory data analysis. Some have already recognized the importance of this: the *Journal of Quantitative Description: Digital Media* was recently launched to promote such work.

A person-specific approach can help us identify precisely these such boundary conditions.

Eventually, this may progress into a computational modelling approach, where relationships between constructs are specified at the level of equations, providing a tight link between theory and statistical predictions.

In both exploratory and confirmatory research, this process of theory-building is necessarily iterative and collective. Instead of viewing theories as wholesale explanations to accept or discard, each subsequent study should seek to improve and build upon the current version of the theory and bring it closer to the truth. Rather than having 100 competing theories with substantial overlap, a more productive game psychology field would seek to merge and synthesize the best aspects of one theory with the best aspects of its competitors to achieve the smallest, but most well-evidenced and generative subset of these theories.

Similarly, instead of viewing theories as being owned by their founder(s), we should view them as community resources.

This will be accompanied by logistical challenges. For example, when one study discovers a previously unknown boundary condition or auxiliary hypothesis and by extension updates the theory, it requires that other practitioners of the theory become aware of this and apply the updated version. Collaborative evidence monitoring webpages and projects may be useful strategies to track shifting consensus about what the most accurate specification of a theory is at any given time.

In sum, it is crucial for the progress-oriented games researcher to think carefully about how they use theory, and how we can continue to increase the specificity of the predictions we make. Greater attendance to the derivation chain, as well as to the descriptive and exploratory work that underlies this, can eventually lead to

Longitudinal Research

It is common at the end of research papers to lament that the current study is cross-sectional and call for future longitudinal research that *is* able to assess causality.

I would be remiss, however, not to restate that these studies are vitally important, especially in light of the person-specific approach for which I advocated above.

Within-person effects are also often the ultimate target of proposed interventions. Addressing dysregulated play with cognitive-behavioral therapy, for example, is a within-person intervention.

Longitudinal research is uniquely capable of understanding the person-specific effects described above, within- as opposed to between-person effects , and much more closely reflects the interests of stakeholders: the long-term well-being of players.

At the same time, I fully recognize that the relative lack of longitudinal research is not simply down to a lack of interest on the part of researchers. The focus on relatively easy to conduct, relatively fast to publish cross-sectional research is unlikely to fade until we change the ways we evaluate and hire researchers (something the Open Science movement, see below, is trying to accomplish).

In the meantime, I wish to point out some more accessible changes that researchers can adopt in their practice to slowly shift the balance in favor of more longitudinal work. The first and easiest to implement is to preferentially cite longitudinal studies. During literature reviews, appending "longitudinal", "panel" or similar to searches can help identify work that involves a temporal component. Citing these studies instead of, or in addition to, their cross-sectional analogues rewards researchers' effort in conducting them. This in turns incentivizes those researchers and others to do more work along those lines. Citing longitudinal

work will in many cases also have greater potential to support the kinds of causal claims researchers are often interested in.

My second recommendation is to consider streamlined variants to longitudinal studies. For example, many cross-sectional surveys could be extended to 2-wave studies with relatively low additional effort and short time commitments—allowing for a comparison of within- and between-person effects. Platforms such as Prolific.co have been designed with longitudinal research in mind, and have guides on how to conduct longitudinal work on their platform with easy recruitment and minimal dropout[*]. Longitudinal studies will typically require smaller sample sizes, and so the additional financial resources needed to pay participants may be balanced out.

Another option is the (re)use of existing data sets. As data sharing slowly becomes more commonplace (and required by certain agencies (Kozlov 2022)), data capable of answering one's particular research question may already exist. For example, landmark work on social media use and well-being used existing data from the Monitoring the Future, Youth Risk and Behaviour Survey and Millennium Cohort Study surveys (Orben and Przybylski 2019). Exploring publicly available data, for example on Google Dataset Search (`datasetsearch.research.google.com`) or the Open Science Framework (`osf.io/search`), has the potential of saving countless hours of work. Even when data is not publicly available, it may be possible to request access by contacting researchers directly (Ballou and Zendle 2021).

Together, changes such as these can create a generation of researchers with more skills in conducting longitudinal work, who recognize and reward the challenge in running such studies.

norm of large-scale projects that are collecting data for many researchers

piggybacking studies

Telemetry Data

terminology - behavioral data, player action data?
reinforcing problem that this particular path forward solves - e.g., actual financial harm from problem gambling/gaming, actual physical aggression
digital games, not all games

Even in a person-specific approach, games research will still be affected by the "garbage in, garbage out" rule. Even in the best case, global measures of screen time may be akin to painting Monet with a tennis racquet—lacking the granularity to understand what about the behavior may be beneficial or harmful. To improve the quality of the input, researchers need to place a much higher priority on getting access to logged play data, sometimes called telemetry data.

Telemetry data can have two key benefits. First, it can answer questions about playtime with considerably more validity. Evidence is accumulating that self-report measures of digital technology use are often inaccurate (Parry et al. 2021), and the degree of inaccuracy may even be related to well-being, creating a possible confound (Sewall et al. 2020). While data on self-report inaccuracy in games is sparse, the limited research indicates that this is a likely problem in the games domain as well (Kahn et al. 2014; Mok and Anderson 2021). Thus, research questions about whether playtime has any relation to a given outcome will likely require logged play data to give trustworthy answers.

Perhaps more valuable is its potential for use in more nuanced questions than "playtime" as a monolithic construct. With players often unable to accurately recall even the total amount of time played in a recent session or day, it is unlikely that reports of yet more detailed behaviors could be trusted. Telemetry data, on the other hand, can include details such as the number of chat

---

[*]`https://researcher-help.prolific.co/hc/en-gb/articles/360009222733-How-do-I-set-up-a-longitudinal-multi-part-study-`

interactions, the proportion of time spent idling, the win-loss rate over time, and much more ==bad examples==. In our reframed research question ("for whom, and under what circumstances, do games cause X?"), the person-specific part addresses the "for whom" part, while detailed telemetry data can help unlock the circumstances.

Getting access to telemetry data is easier said than done, but there are a variety of strategies researchers might use to do so. I am aware of at least four ways that researchers can gain access to logged telemetry data from players, each with their own strengths and weaknesses, summarized in Table 1.

==data as byproduct, e.g. twitch==

==in framing - media effects==

==potential new row - player-shared data, forums where people share play logs, boardgamegeek, twitter, wowwiki, "theorycrafting", scraping more broadly - online methods, Yelena's review paper, melissa rogerson==

The first is through industry collaborations. Some research groups have negotiated data sharing agreements with game developers and/or publishers to conduct research on their players, using data that the companies hold (Vuorre et al. 2021; Johannes et al. 2021b; Kokkinakis et al. 2017; Kahn et al. 2014).

Such agreements can provide extremely large and rich datasets, in line with the detailed data kept by most tech companies. However, industry collaborations have the downside of restricting analysis to the (likely small) number of games for which companies have made data accessible— and importantly, to data on individual players' playtime for an individual game. Although little is known about how varied gaming "diets" are, both platform-wide data and community forums indicate that players regularly alternate between a variety of games over the course of a day or week based on factors such as mood, available time, social affordances, and novelty-seeking. Thus, data

on particular games may not tell us much about how the effects of gaming overall.

==Incentivizing industry collaborations? Discussion of potential distrust relating to decades of moral panic?==

Second, researchers can use third-party tracking tools and existing time/behavior tracking systems. To date, these have been more commonly used for research studies on global screen time (Ohme et al. 2021; Sewall and Parry 2021) or social media on mobile devices

Third, there exist a handful of publicly available APIs or data interfaces that provide access to in-game data, such as OpenDota (`www.opendota.com`) for Dota 2 or RoyaleAPI for Clash Royale. These APIs often record data from individual matches, including decisions made, match results, time/duration, and more. Relatedly, some games have made "data dumps" available in the form of large downloadable data sets with in-game data (u/pants555 2020).

Fourth and finally, in certain regions, players have the legal right to request a copy of their own data. For example, GDPR affords residents of the European Union the right to request copies of the data held by data processors. Among other examples, players can therefore request data for their play history on Playstation or Xbox. A similar right exists under the California Consumer Privacy Act (CCPA).

At the time of writing, I was unable to find any examples of research using data from data requests. Instead, I provide an example of my own Playstation data, returned by via such a request in December 2021. To obtain this, I sent one request email and two brief clarifications for follow-ups; in total, the process took approximately 6 weeks. A streamlined system for players requesting this data, and sharing it with researchers alongside psychological measures, could over time contribute to the development of a large-scale, high quality dataset. I provide an email template that may serve

Access Method to Telemetry Data	Pros	Cons	Examples of Research or Available Tools
Industry Collaboration	• Offers possibility of receiving very large datasets all at once • Data is provided directly from the platform in question	• Requires individually brokered agreements, often limited to a single game • More accessible to wealthy, well-connected, and prestigious research groups	Vuorre et al. (2021); Johannes et al. (2021b); Kokkinakis et al. (2017)
APIs and data dumps	• Easy to access • Provides access to very large datasets	• Only available for a select number of games • May only be provided in anonymous form, and thus not be directly linkable to self-report data	Open Dota (`opendota.com`) RoyaleAPI (`royaleapi.com`) Steamworks (`partner.steamgames.com`)
Data requests	• Data from months or years can be accessed in a single request • Data is provided directly from the platform in question	• Users are only legally guaranteed a copy of their data in certain regions • Process may be slow and bureaucratic	• Playstation (`sony.co.uk/eu/pages/privacy/en_GB/privacy_policy.html`)
Player-facing tracking tools and features	• May be customizable to track particular behavioral data of interest	• Some implementations may violate platforms' terms of service	iOS Screen time Rescue Time Google wellbeing Microtransaction logs

Table 1. Potential methods for games researchers to access objectively-tracked telemetry data on video game play, with their associated pros and cons

to accelerate future requests. Both can be found in the ==supplementary materials.==

==standardized framework for easy and structured data requests==

==tutorials for asking for data requests==

Understanding individual features

==gamification as strong example of inconsistent effects, poorly controlled variants==

==matching/propensity scores, causal inference as statistical technique - analogue for manipulation with matched pairs - but note, how do we know what to match on until we've done the research? unknown unknown problem==

==as above, highlighting the problem and why it distorts results==

==david gundry's validity threat paper esp. lit review==

A persistent criticism of lab-based video games research is that what are often called control and experimental groups are playing entirely separate games that differ on a variety of features beyond simply the one hypothesized about (e.g., violence).

This has been particularly problematic in the violent games literature, where many prominent studies report on laboratory studies where players are assigned to play either a violent game or a non-violent one (Anderson and Dill 2000, e.g.,). However, any two distinct games will differ not only in their violent content, but also various other aspects of design. Although some of these studies

Method of controlling individual game features	Pros	Cons	Resources for Learning or Applying
Modding	• Can use existing commercial-quality games, resulting in high ecological validity • Relatively low barrier to entry, with some alterations requiring only a few lines of code	• Limited to games that support modding • Will require participants to download and install for any research outside of a lab setting	Hilgard et al. (2019a)
Building custom games	• Risks low ecological validity • Offers the greatest degree of control over each and every feature • Can use fully featured game templates	• Labor intensive (to varying degrees, depending on genre) • May require advanced programming knowledge	Peng et al. (2012)
Natural Experiments	• Directly reflects the decisions of commercial developers and how they might affect players • Is analogous to some types of industry research, and may support future career opportunities	• Requires fast action to align with planned releases or updates to a game	Zendle (2019)

Table 2. Methods that allow researchers to manipulate individual features of games to understand their effects on players

collect pilot samples to test for a significant confound between games, non-significant tests do not indicate equivalence and are particularly untrustworthy with small samples (Hilgard et al. 2017a).

Instead, to understand the effect of individual game features, design choices, or mechanics, researchers need to compare two rigorously controlled versions of the same game that differ only in that aspect.

This relates to a key challenge in games research: experiences and effects of games are subjective, holistic, emergent, situated, and dynamic (Hassenzahl 2010).

It may be impossible for games research to ever fully identify the "active ingredients" of particular game effects, as every design feature will interact with those around it to produce their final result.

Nonetheless, for a discipline that regularly invokes potential implications for designers as justification for the impact of its work, it is necessary to drill down more precisely on how particular game features are affecting players—alone and in combination.

Hilgard et al. (2019a) modded *Doom II* to create versions with additional violence and reduced violence, and found that participants who played each version did not differ in postgame aggressive behavior.

Our ability to address this issue is much improved compared to a decade ago. No longer are researchers required to build a game from scratch to manipulate a single feature, and even if one does choose to do so, software tools have made simple game development drastically more

accessible. In Table 2, I summarize 3 ways that researchers can gain control over individual game features: modding, building custom games, and using natural experiments.

There are now numerous tools for creating games entirely using a GUI with little to no coding, such as Twine for interactive stories (`twinery.org`) or GameMaker studio for 2D games (`gamemaker.io/en`).

For those with greater programming expertise, major engines such as Unity have free educational licenses. Researchers can take open source games off the shelf and make modifications to them to create experimental test beds for particular features.

An alternative strategy is to use planned changes to video games, or patches, as opportunities for natural experiments. While this can require researchers to act quickly to (Zendle 2019)

For example, a sample of World of Warcraft players could be studied before and after the release of a new expansion. Combining this further with data about achievements and equipment from the public WoW Armory could provide deep insights into how content availability changes both players' behavior and player experience. cite and fix

game templates

Open Science

don't underestimate how unfamiliar some of the target audience may be with open science

reiterating how uncontrolled research degrees of freedom can show almost any effect - file drawer, QRPs, etc etc - skews total landscape

Research, even conducted using the principles above, is only as useful as it is trustworthy and transparent in the way it is conducted and shared. Unfortunately, we have reason to believe that the status quo of research in the social sciences does not meet the standards for robustness we would like it to.

The situation, according to some, is dire. Meta-research demonstrates how so-called questionable research practices—such as reporting only one of a set of multiple dependent variables, optionally controlling for covariates, and collecting more data after a non-significant result—can balloon the false positive rate in the literature well beyond the nominal 5% (with a $p < .05$ significance level) (Simmons et al. 2011; Ioannidis 2005). Results produced through questionable research practices and "p-hacking" may not reflect a true effect, and therefore may not be replicable: a prominent project was only able to replicate 36 of 97 high-profile psychological studies, with replication yielding effect sizes half as large on average Open Science Collaboration (2015).

This becomes even more problematic when we consider publication bias. Meta-analyses on video game effects have often substantial evidence of publication bias, suggesting that many studies with non-significant results have not been published— a phenomenon known as the file drawer effect. Together, this means that the current literature is likely both biased towards particular types of results, and that many of the results that are represented are unverifiable, do not transparently reflect the analysis process used to find them, or are outright fraudulent.

The current leading movement to address issues of questionable research practices, publication bias, and more is known as open science.

The tenets of the open science movement are too varied to discuss exhaustively here. Instead, I highlight some of the key practices and the benefit they can have for games research (Table 2), and encourage interested readers to explore these ideas further. For easy entry into open science communities and topics, I recommend searching for a local ReproducibiliTea journal club (`reproducibilitea.org`), subscribing to the Open Research Calendar (`openresearchcalendar.org`), and/or exploring the entry-level reading list collated by Crüwell et al. (2019).

Given the diversity of beliefs among different researchers, adversarial collaborations offer

Open Science Practice	Benefit to Games Research	Resources for Learning or Applying
Preregistration	• Differentiates exploratory vs confirmatory research • Encourages researchers to think more thoroughly about theory and potential boundary conditions	• Przybylski and Weinstein (2019) OSF's preregistration short practical guide `osf.io/2vu7m/`
Registered Reports	• Alleviates the file drawer effect, where null results are less likely to be published	• Przybylski and Weinstein (2019)
Team Science	• Can drastically increase sample sizes • Can improve generalizability across cultures and context Offers opportunities for diverse viewpoints to influence research designs	Many Labs (`osf.io/8cd4r/wiki/home/`) • Forscher et al. (2020)
Replication studies	• TODO	TODO
Open data	• Creates opportunity for future reuse, saving drastically on time and effort needed for certain studies • Allows readers and reviewers to detect potential errors in analysis or reporting	• Soderberg (2018)

Table 3. Selected practices encouraged by the open science movement, and how these can be beneficial for research on games

another fruitful option. In these cases, researchers with differing beliefs about a particular topic agree on a research design that would (at least in part) be able to shift their opinion. These studies can be preregistered to limit analytical flexibility, providing a neutral test bed for controversial topics. In one of the few examples to date, an adversarial collaboration successfully advanced knowledge on the effects sexualized women characters in games (Lindner et al. 2020).

Discussion

In this paper, I have described the sometimes frustrating present state of psychological research on video games, in which answers to broad questions about the effects of games are scarce. I have then described 6 areas in which we can reform our research practice to understand how players and games interact more successfully: using person-specific effects and models, strengthening the derivation chain, conducting longitudinal research, collecting logged behavioral data, manipulating individual features, and applying open science methods.

The path forward I have laid out is a difficult one. In my opinion, however, it is also a hopeful one. I have described a wide range of specific actions that researchers can take to contribute to a more productive future of games research, from integrating various sources telemetry data into their research designs, to adopting open science practices, to using person-specific statistical analysis techniques, to preferentially citing existing research that uses these techniques. Even better, these 5 strategies are synergistic—a custom-designed game can collect a wide telemetry data,

or a longitudinal study could be submitted as registered report so that researchers have something to put on their CV even before the study has concluded.

As with other aspects of science reform and attempts to improve one's own research practice, *start small*. Researchers should not feel they need to conduct the one study to rule them all; instead, change will come from slowly implementing one or more of these ideas in one's research.

I have focused in this paper to the greatest extent possible on actions that individual researchers or groups can take in their upcoming projects. However, I want to make it very clear that this needs to be accompanied by large, top-down structural changes to the way we evaluate researchers and research impact, to how industry shares data, to the kinds of skills that behavioral technology researchers are taught, and much more. Rather than alleviate the need to large-scale changes to the research ecosystem, I hope that pursuing research along the 5 lines described above can empower the research community, making clear the potential gains as these practices and others become integrated into our work.

Acknowledgements

References

Aarseth E, Bean AM, Boonen H, Colder Carras M, Coulson M, Das D, Deleuze J, Dunkels E, Edman J, Ferguson CJ, Haagsma MC, Helmersson Bergmark K, Hussain Z, Jansz J, Kardefelt-Winther D, Kutner L, Markey P, Nielsen RKL, Prause N, Przybylski A, Quandt T, Schimmenti A, Starcevic V, Stutman G, Van Looy J and Van Rooij AJ (2017) Scholars' open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal. *Journal of Behavioral Addictions* 6(3): 267–270. DOI:10.1556/2006.5.2016.088.

Anderson CA and Dill KE (2000) Video Games and Aggressive Thoughts, Feelings, and Behavior in the Laboratory and in Life. *Journal of Personality and Social Psychology* 78(4): 772–790. DOI:10.1037/0022-3514.78.4.772.

Anderson CA, Shibuya A, Ihori N, Swing EL, Bushman BJ, Sakamoto A, Rothstein HR and Saleem M (2010) Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin* 136(2): 151–173. DOI:10.1037/a0018251.

Ballou N and Zendle D (2021) "Clinically significant distress" in internet gaming disorder: An individual participant meta-analysis. *Computers in Human Behavior* : 107140DOI:10.1016/j.chb.2021.107140.

Bediou B, Adams DM, Mayer RE, Tipton E, Green CS and Bavelier D (2018) Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin* 144(1): 77–110. DOI:10.1037/bul0000130.

Cohen J (1990) Things I have learned (so far). *American Psychologist* 45(12): 1304–1312. DOI:10.1037/0003-066X.45.12.1304.

Crüwell S, van Doorn J, Etz A, Makel MC, Moshontz H, Niebaum JC, Orben A, Parsons S and Schulte-Mecklenbeck M (2019) Seven Easy Steps to Open Science: An Annotated Reading List. *Zeitschrift für Psychologie* 227(4): 237–248. DOI:10.1027/2151-2604/a000387.

Drummond A, Sauer JD and Garea SS (2018) The Infamous Relationship Between Violent Video Game Use and Aggression: Uncharted Moderators and Small Effects Make It a Far Cry from Certain. In: Ferguson CJ (ed.) *Video Game Influences on Aggression, Cognition, and Attention*. Cham: Springer International Publishing. ISBN 978-3-319-95494-3 978-3-319-95495-0, pp. 23–40. DOI:10.1007/978-3-319-95495-0_3.

Ferguson CJ (2015) Do Angry Birds make for angry children? A meta-analysis of video game influences on children's and adolescents' aggression, mental health, prosocial behavior, and academic performance. *Perspectives on Psychological Science* 10(5): 646–666. DOI:10.1177/1745691615592234.

Ferguson CJ (2020) Aggressive video games research emerges from its replication crisis (Sort of). *Current Opinion in Psychology* 36: 1–6. DOI:10.1016/j.copsyc.2020.01.002.

Forscher PS, Wagenmakers EJ, Coles NA, Silan MA, Dutra NB, Basnight-Brown D and IJzerman H (2020) The Benefits, Barriers, and Risks of Big Team Science. DOI:10.31234/osf.io/2mdxh.

Hassenzahl M (ed.) (2010) *Experience Design: Technology for All the Right Reasons*. San Rafael, Calif: Morgan & Claypool. ISBN 978-1-60845-048-0 978-1-60845-047-3.

Hilgard J, Engelhardt CR, Bartholow BD and Rouder JN (2017a) How much evidence is p > .05? Stimulus pre-testing and null primary outcomes in violent video games research. *Psychology of Popular Media Culture* 6(4): 361–380. DOI:10.1037/ppm0000102.

Hilgard J, Engelhardt CR and Rouder JN (2017b) Overstated evidence for short-term effects of violent games on affect and behavior: A reanalysis of Anderson et al. (2010). *Psychological Bulletin* 143(7): 757–774. DOI:10.1037/bul0000074.

Hilgard J, Engelhardt CR, Rouder JN, Segert IL and Bartholow BD (2019a) Null effects of game violence, game difficulty, and 2D:4D digit ratio on aggressive behavior. *Psychological Science* 30(4): 606–616. DOI:10.1177/0956797619829688.

Hilgard J, Sala G, Boot WR and Simons DJ (2019b) Overestimation of action-game training effects: Publication bias and salami slicing. *Collabra: Psychology* 5(1): 30. DOI:10.1525/collabra.231.

Ioannidis JPA (2005) Why Most Published Research Findings Are False. *PLoS Medicine* 2(8): e124. DOI:10.1371/journal.pmed.0020124.

Johannes N, Masur PK, Vuorre M and Przybylski AK (2021a) How should we investigate variation in the relation between social media and well-being? Preprint, PsyArXiv. DOI:10.31234/osf.io/xahbg.

Johannes N, Vuorre M and Przybylski AK (2021b) Video game play is positively correlated with well-being. *Royal Society Open Science* 8(2): rsos.202049, 202049. DOI:10.1098/rsos.202049.

Kahn AS, Ratan R and Williams D (2014) Why We Distort in Self-Report: Predictors of Self-Report Errors in Video Game Play. *Journal of Computer-Mediated Communication* 19(4): 1010–1023. DOI: 10.1111/jcc4.12056.

Klecka H, Johnston I, Bowman ND and Green CS (2021) Researchers' commercial video game knowledge associated with differences in beliefs about the impact of gaming on human behavior. *Entertainment Computing* 38: 100406. DOI:10.1016/j.entcom.2021.100406.

Kokkinakis AV, Cowling PI, Drachen A and Wade AR (2017) Exploring the relationship between video game expertise and fluid intelligence. *PLOS ONE* 12(11): e0186621. DOI:10.1371/journal.pone.0186621.

Kozlov M (2022) NIH issues a seismic mandate: Share data publicly. *Nature* 602(7898): 558–559. DOI: 10.1038/d41586-022-00402-1.

Lindner D, Trible M, Pilato I and Ferguson CJ (2020) Examining the effects of exposure to a sexualized female video game protagonist on women's body image. *Psychology of Popular Media* 9(4): 553–560. DOI:10.1037/ppm0000251.

Markey PM and Ferguson CJ (2017) *Moral Combat: Why the War on Violent Video Games Is Wrong*. Dallas, TX: BenBella Books. ISBN 978-1-942952-98-5.

Markey PM, Ivory JD, Slotter EB, Oliver MB and Maglalang O (2020) He does not look like video games made him do it: Racial stereotypes and school shootings. *Psychology of Popular Media* 9(4): 493–498. DOI:10.1037/ppm0000255.

Mathur MB and VanderWeele TJ (2019) Finding Common Ground in Meta-Analysis "Wars" on Violent Video Games. *Perspectives on Psychological Science* 14(4): 705–708. DOI:10.1177/1745691619850104.

Meehl PE (1990) Why Summaries of Research on Psychological Theories are Often Uninterpretable. *Psychological Reports* 66: 195–244. DOI:10.2466/pr0.1990.66.1.195.

Mok L and Anderson A (2021) The Complementary Nature of Perceived and Actual Time Spent Online in Measuring Digital Well-being. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW1): 1–27. DOI:10.1145/3449160.

Ohme J, Araujo T, de Vreese CH and Piotrowski JT (2021) Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication* 9(2): 293–313. DOI:10.1177/2050157920959106.

Open Science Collaboration (2015) Estimating the reproducibility of psychological science. *Science* 349(6251): aac4716–aac4716. DOI:10.1126/science.aac4716.

Orben A (2020) The Sisyphean Cycle of Technology Panics. *Perspectives on Psychological Science* 15(5): 1143–1157. DOI:10.1177/1745691620919372.

Orben A and Przybylski AK (2019) The association between adolescent well-being and digital technology use. *Nature Human Behaviour* 3(2): 173–182. DOI:10.1038/s41562-018-0506-1.

Parry DA, Davidson BI, Sewall CJR, Fisher JT, Mieczkowski H and Quintana DS (2021) A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour* 5: 1535–1547. DOI:10.1038/s41562-021-01117-5.

Parry DA, Fisher JT, Mieczkowski H, Sewall CJ and Davidson BI (2022) Social media and well-being: A methodological perspective. *Current Opinion in Psychology* 45: 101285. DOI:10.1016/j.copsyc.2021.11.005.

Peng W, Lin JH, Pfeiffer KA and Winn B (2012) Need satisfaction supportive game features as motivational determinants: An experimental study of a self-determination theory guided exergame. *Media Psychology* 15(2): 175–196. DOI:10.1080/15213269.2012.673850.

Przybylski AK (2014) Who Believes Electronic Games Cause Real World Aggression? *Cyberpsychology, Behavior, and Social Networking* 17(4): 228–234. DOI:10.1089/cyber.2013.0245.

Przybylski AK and Weinstein N (2019) Violent video game engagement is not associated with adolescents' aggressive behaviour: Evidence from a registered report. *Royal Society Open Science* 6: 16. DOI:10.1098/rsos.171474.

Scheel AM, Tiokhin L, Isager PM and Lakens D (2021) Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science* 16(4): 744–755. DOI:10.1177/1745691620966795.

Sewall CJR, Bear TM, Merranko J and Rosen D (2020) How psychosocial well-being and usage amount predict inaccuracies in retrospective estimates of digital technology use. *Mobile Media & Communication* 8(3): 379–399. DOI:10.1177/2050157920902830.

Sewall CJR and Parry DA (2021) The Role of Depression in the Discrepancy Between Estimated and Actual Smartphone Use: A Cubic Response Surface Analysis. *Technology, Mind, and Behavior* 2(2): 1–9. DOI:10.1037/tmb0000036.

Simmons JP, Nelson LD and Simonsohn U (2011) False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11): 1359–1366. DOI:10.1177/0956797611417632.

Soderberg CK (2018) Using OSF to Share Data: A Step-by-Step Guide. *Advances in Methods and Practices in Psychological Science* 1(1): 115–120. DOI:10.1177/2515245918757689.

Tyack A and Mekler ED (2020) Self-determination theory in HCI games research – current uses and open

questions. In: *CHI*. Honolulu: ACM, p. 21. DOI: 10.1145/3313831.3376723.

u/pants555 (2020) 77 Million Runs - An STS Metrics Dump.

Valkenburg PM, Beyens I, van Driel II and Keijsers L (2022) How to Assess Variation in the Association Between Social Media Use and Well-Being: A Reply to Johannes, Masur, Vuorre, & Przybylski (2021). Preprint, PsyArXiv. DOI:10.31234/osf.io/rwc73.

Verbeij T, Pouwels JL, Beyens I and Valkenburg PM (2022) Experience sampling self-reports of social media use have comparable predictive validity to digital trace measures. *Scientific Reports* 12(1): 7611. DOI:10.1038/s41598-022-11510-3.

Vuorre M, Johannes N, Magnusson K and Przybylski AK (2021) Time spent playing video games is unlikely to impact well-being. Preprint, PsyArXiv. DOI:10.31234/osf.io/8cxyh.

Weinstein N, Przybylski AK and Murayama K (2017) A prospective study of the motivational and health dynamics of Internet Gaming Disorder. *PeerJ* 5: e3838. DOI:10.7717/peerj.3838.

Zendle D (2019) Problem gamblers spend less money when loot boxes are removed from a game: A before and after study of *Heroes of the Storm*. *PeerJ* 7: e7700. DOI:10.7717/peerj.7700.