Preprint
1–22

A 6-part Manifesto for More Productive Psychological Games Research

Nick Ballou[1]

Abstract

Despite enormous research attention paid to questions of how video games do or do not affect their players, researchers continue to lack consensus on even simple questions. Research areas such as violent games and aggression, playtime and wellbeing, action game training benefits, and many more are characterized by deeply divided opinions, conflicting results even at the meta-analytic level, and a general lack of consistent or trustworthy effects. I argue that much of this disagreement stems from limitations of common epistemic approaches and methods used in the field, including use of problematic self-report measures of behavior, overreliance on cross-sectional research, and superficial use of theory, among others. To resolve debates more effectively, I describe 6 practices video games research can adopt: 1) modeling effect size variation, 2) strengthening the theoretical derivation chain, 3) conducting longitudinal studies, 4) collecting digital trace data, 5) manipulating individual game features, and 6) adopting open science. For each practice, I attempt to include practical steps toward implement them in one's own work, and highlight resources that may have gone overlooked. While critical of the current state of knowledge, the paper is hopeful: with the help of these practices, the next generation of video games research can be the most informative yet.

Keywords

video games, psychology, methods, open science, digital trace data

Introduction

Decades of research has investigated whether video games are associated with, and ultimately cause, a variety of positive and negative psychological outcomes. Hypothesized 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.

[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

After investing monumental amounts of time and research funding into understanding these topics, however, many questions—even simple ones—continue to lack clear answers. Where detected at all, effects are nuanced, inconsistent, and generally small in magnitude. As a result, many researchers hold deeply divided opinions about the trustworthiness or importance of effects. Based in part on factors such as age, familiarity with commercial video games, and field of study (Klecka et al. 2021), researchers maintain different assessments of the evidence base for these questions.

Together, this suggests that researchers are asking questions that are likely too broad to be answerable without clearer specification, and using methods while it may be the case that some of these questions are too broad to be answered without further specification, neither have we made much progress narrowing our questions to find more robust relationships.

In large part, this continued uncertainty is a result of severe limitations of the prevailing methods used in the field. Much of the literature on games is nomothetic (i.e., trying to establish general 'laws' across large groups of diverse players), cross-sectional, and uses theory superficially. Other limitations include an over-reliance on self-report measures that may not reflect actual behavior and a lack of comparable control conditions. All of this is underpinned by a research ecosystem that has undervalued transparency and openness, and whose results often cannot be reproduced or replicated. In the coming sections, I describe these issues and their consequences in more detail.

My hope in this paper is to convince you of that this characterization of the field is accurate and problematic, but that it can also act as a catalyst for change. I am hopeful that research on games can reform certain practices, and reach greater consensus on these varied topics. Below, I describe 6 ways that research can be improved, and yield more informative results:

1. Model effect size variation
2. Strengthen the derivation chain
3. Conduct longitudinal studies
4. Use digital trace data
5. Control individual features
6. Adopt open science

None of these ideas are new—instead, my goal is simply to collate them in a document that is specific to research on games, and to add practical guidance to make the ideas actionable. I drew particular inspiration from (Parry et al. 2022), who discuss a partially-overlapping methodological vision for future research on social media and well-being.

This background for 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 6 practices are not limited solely to video games effects research, but rather any work that uses psychological methods to understand how players interact with games.

A Brief History of Unresolved Questions

Before discussing solutions, it is worth briefly 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, certain questions are yet to be satisfactorily resolved.

Do violent games cause 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 study, debates continue to smolder. Findings supportive of a link between violent video game play and subsequent aggressive thoughts or behaviors continue to proliferate, and are regularly invoked 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 reached radically different conclusions (Anderson et al. 2010; Ferguson 2015; Hilgard et al. 2017b).

Are video games addictive? At the time of writing, one of the most active areas 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, but substantial disagreement about whether problematic use constitutes a mental disorder (Ferguson and Colwell 2020). Constituent areas of controversy include whether dysregulated gaming is a symptom of other mental health conditions, a cause of them, or both (Weinstein et al. 2017), whether certain types of games are inherently addictive (Lemmens and Hendriks 2016), whether the current approaches to measurement and diagnosis are appropriate (Ballou and Zendle 2022; Aarseth et al. 2017), and the degree of harm that formalizing gaming-related disorders could have for healthy players (Aarseth et al. 2017).

Do video games harm (or improve) wellbeing? Several studies report that playtime is negatively associated with certain wellbeing constructs such as depressive symptoms (Burke and Lucier-Greer 2021), sleeping problems (Wenzel et al. 2009), and anxiety (Lo et al. 2005), with at least one meta-analysis finding a small negative link with overall psychogical wellbeing (Liu et al. 2019). Other studies find associations only for certain ages or wellbeing constructs such as anxiety (Loton et al. 2016); or small and nuanced negative associations among highly engaged players (Allahverdipour et al. 2010; Colder Carras et al. 2017; Przybylski and Weinstein 2017). Yet others have found null or negligible associations (Johannes et al. 2021b; Vuorre et al. 2021).

Do action games have cognitive benefits? 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). Again, however, conclusions are inconsistent and contentious even at the meta-analytic level (for evidence in favor, see Bediou et al. 2018; for conflicting evidence, see Hilgard et al. 2019b).

While the above examples are among the most prominent, they are by no means the only ones to fall short of conclusive answers. Researchers continue to debate topics such as whether exposure to sexualized (female) game characters affects body image (Lindner et al. 2020), the contexts in which gamification is and is not successful (Dicheva et al. 2015), whether serious games can be an effective mental health intervention (Abd-Alrazaq et al. 2022), and much more.

To be clear, my intention in reviewing these examples is not to support bothsidesism: in the above conflicts, more weight should be given to the more rigorous, larger sample, preregistered studies—those that adopt many of the practices I will shortly recommend. Neither, however, does this mean that all conflicting results are 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), we can acknowledge this without dismissing all results that conflict with our own beliefs about the medium.

In short, results in a huge variety of research topics have been definitively mixed. Researchers are regularly unable to reach consensus on even seemingly basic 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 fully answering the key questions that parents, players, and policymakers have about the ever-growing role of video games in daily life. Below, I describe 6 ways that research on video games can be improved and begin to resolve these debates: 1) modeling effect size variation, 2) strengthening the theoretical derivation chain, 3) conducting longitudinal studies, 4) collecting digital trace data, 5) manipulating individual game features, and 6) adopting open science.

Model Effect Size Variation

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). This may be one of the key factors underlying the inconsistent-but-trending-toward-null findings reviewed above. While most people may not be affected by violent video games, for example, some subset may exhibit a positive or negative relationship between their exposure to violence in games and subsequent aggression.

In other words, it often may not be productive—or even tractable—to ask 'do video games cause X?'. That question represents a nomothetic approach; i.e., a generalized law that applies to the 'average person' across an entire population without necessarily describe the psychology of any actual person. Given that researchers can find evidence supporting both the existence and non-existence of almost all gaming-related effects, this suggests that the nomothetic approach may be insufficient. Instead, a variety of research questions would be better framed as 'for whom, and under what circumstances, do video games cause X?'

One way to build toward this is to attend more closely to *effect size variation*. Related to a family of approaches sometimes called 'person-specific', 'idiographic', or 'individual-level', modeling effect size variation seeks to understand how effects vary across people.

Consider the example of video games and well-being. 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 Johannes et al. (2021b); Vuorre et al. (2021). Hypothetically, however, if one portion of a sample is exhibiting a negative effect of video game play 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 meaningful effect of video game on well-being'.

Other areas of research, notably including work on social media, have begun to model effect size variation with valuable results. Beyens et al. (2020), for example, found that passive Instagram use had no meaningful relationship with wellbeing for 74% of adolescents, but that the association was positive for another 17%, and negative for the remaining 9%. This then opens the door for more detailed subgroup analyses to establish what distinguishes these groups from each other.

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

There have been debates about how best to incorporate research on effect size variation without losing sight of the goals of understanding groups, and generalizing from sample to population (cf. Valkenburg et al. 2022; Johannes et al. 2021a). 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 some 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 (in part) determine who experiences more positive relationships between media use and a certain outcome, and who experiences more negative ones (see Strengthen the Theoretical Derivation Chain below).

In sum, modeling effect size variation can help us develop more targeted predictions about how video games affect particular groups. To get started with such approaches, I recommend reading Johannes et al. (2021a)'s pedagogic description, papers from related fields that have used this approach (e.g., Beyens et al. 2020; Verbeij et al. 2022), and learning more about multilevel models, the family of statistical model most frequently used to account for effect size variation in nested data (e.g., Brown 2021).

Strengthen the Theoretical Derivation Chain

Building on the previous point, many readers will recognize that theories currently used in psychological games research are typically underspecified to make predictions at the individual level. In many cases, the theories we use are underspecified even to make clear predictions at the population level; broad, verbal statements leave substantial ambiguity about measures, statistical models (are all effects linear?), the size of the effect, boundary conditions, and more. As both cause and outcome of this problem, theories are often used superficially, as background context or plausible explanations largely separate from the data at hand (e.g., Tyack and Mekler 2020).

Without being sufficiently specific, theories risk becoming unfalsifiable (Wallach and Wallach 2010). Any failure to find an expected result can be explained away by decisions in the study design or analysis. This can either lead to failed predictions being entirely ignored, or in some cases lead to what is known as a degenerate research line: modifications are made to accommodate failed predictions without improving the theory's predictive success (Lakatos 1978). In either case, we are left with vague, unchanging, and often flimsy theories that are unable to generalize to different games, groups, or contexts.

Broadly, to solve this problem we need to strengthen the 'derivation chain': the 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 inputs to the derivation chain, as well as research activities that can strength it.

On the journey from theory to statistical hypothesis, researchers first need to clearly define the concepts, ensuring that they are coherent and differentiated from other, potentially similar concepts. These concepts will need to be measured; in the case of psychology, this is often done using a questionnaire—one that needs to be sufficiently valid for use in that context. Unfortunately, this is regularly not the case in games (Hughes and Cairns 2021) or psychology more broadly (Flake and Fried 2020).

Next, the relationship between these measured concepts must be specified; the theory should provide a causal model that includes key confounding variables with the potential to obscure or bias the

relationship. We may also need to specify situations where the theory does not hold, known as boundary conditions (see, e.g., Gilbert et al. 2021, for a recent, excellent example of this). Finally, we will need to adopt a set of auxiliary hypotheses—claims not directly derived from the theory, but that are necessary from translating from theory to observable entities. For example, a test of whether self-reported playtime increases prosocial behavior requires (among other potential auxiliaries) that self-reported play is an accurate proxy for actual play, that people define play similarly (e.g., not including time idling in the menu), and that participants respond honestly.

Finally, all these components need to be converted into a statistical test. Questions about what effect sizes are practically meaningful, the shape of the relationship (our statistical tests are overwhelmingly linear, but this may not always be appropriate), whether prior beliefs should be incorporated, and more need to be specified to ensure that the statistical test has sufficient capacity to falsify the theoretical prediction.

Laid out before us, it is clear that these steps are often not taken seriously. Luckily, we are not without recourse to improve the derivation chain. Among the most important options is simply *not testing hypotheses*. Given our current lack of knowledge about boundary conditions, auxiliary hypotheses, and other aspects of the derivation chain, a (purely) confirmatory approach is ill-advised—in many cases, we simply lack nuanced enough theory to make specific predictions that could be severely tested (Mayo 2018).

Instead, we can make greater use of descriptive work and exploratory data analysis. Both of these can play a crucial role in theory generation—identifying new effects to be tested, developing a shared understanding of a particular phenomenon, or honing terminology and methods best practices, to fleshing out existing theory by establishing boundary conditions. The importance of descriptive research is already well-appreciated in some communities: the *Journal of Quantitative Description: Digital Media* was recently launched to promote such work.

Strengthening the derivation chain implicates a more iterative and collective approach to theory-building than the current norm. Instead of viewing theories as wholesale explanations to accept or discard (often 'owned' by their founders), 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 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. Eventually, this may progress into a computational modelling approach, where relationships are specified at the level of equations, providing a tight link between theory and statistical predictions Guest and Martin (2021).

In sum, it is crucial for the progress-oriented games researcher to think carefully about how they use theory, and how we can increase the specificity of our predictions. I encourage readers to read more carefully about the relationship between theory, hypothesis, and model (e.g., Scheel et al. 2021; Guest and Martin 2021), review the basics of causal modeling (Rohrer 2018), and consider conducting more explicitly descriptive or exploratory studies (for a tutorial, see e.g., Szabelska et al. 2021; for varied examples, see e.g., Festic et al. 2021; Karakus et al. 2008; DeCamp 2017).

Conduct longitudinal studies

The most prominent type of study design in the literature on games is cross-sectional. Though I am not aware of exact figures, a Scopus search of 'video games AND (survey OR cross-sectional)' returned 3,964 results; a search of 'video games AND (panel OR longitudinal)' returned just 845.

While there is nothing inherently problematic about cross-sectional studies, an overreliance on

them can have several negative consequences. First, cross-sectional studies—especially in the absence of strong theory—are limited in their ability to support causal relationships, a key goal of our field and the social sciences more broadly (evidenced by the common lamentation at the end of papers that the current study is cross-sectional, with a call for future longitudinal work to be done). A correlational finding that video game play and wellbeing negatively correlate, for example, is interesting but leaves us unsatisfied—the result could equally be the result of video game play causing lower wellbeing, lower wellbeing causing greater play, or an unknown third variable influencing both.

Building on the above, causality is fundamentally a within-person (or within-unit) inference. We cannot apply a treatment that would make one person (or group of people) become another person (or group of people). A causal effect of video game play on wellbeing, for example, means 'if person A plays more games, they will feel [better/worse] than they did before'. While between-person comparisons can be used to support causal effects given certain assumptions, the only way to establish a within-person effect is to collect multiple data points from the same person—inherently the opposite of a cross-sectional design.

The last problem is that cross-sectional studies, by definition, capture a single moment in time (perhaps with some element of recalled past experiences). In the real world, games are often played over periods of weeks, months, or years. The effects that games have on players are unlikely to be fully visible in such snapshots. Unless our methods can reflect this ongoing engagement, our results will always be heavily qualified.

It is therefore crucial that we embrace longitudinal (or 'repeated measures') designs and aim to identify causal, within-person, and temporally relevant effects where possible. While it would be unfair not to acknowledge the incentive structures that motivate cross-sectional studies (more publications, more quickly), I am encouraged by the fact that non-cross-sectional studies are perhaps as accessible as they have ever been. In particular, I want to highlight examples and resources for two variations of longitudinal studies that I believe are underrepresented relative to their knowledge contributions: ecological momentary assessment/experience sampling method (EMA/ESM), and non-intensive longitudinal panel designs.

Ecological momentary assessment is a type of intensive longitudinal study that involves repeatedly asking participants to report on their thoughts, feelings, or behavior at that particular moment, often several times per day. A full description of EMA is well beyond the scope of this paper, but it has been used fruitfully, if sparsely, in research on games. For example, one study examined how the time at which people play Pokemon GO relates to their physical activity (Marquet et al. 2018), and another how mood improvement after playing violent games relates to aggression (Kersten and Greitemeyer 2022). There now exist numerous applications that support experience sampling, including some fully open source such as ExperienceSampler (`experiencesampler.com`; Thai and Page-Gould 2018) and PACO (`pacoapp.com`).

Non-intensive longitudinal studies, on the other hand, involve collecting fewer data points per participant, typically spaced further apart. Panel studies in particular are relatively common in the games literature, especially in research on disordered play. Some examples include Weinstein et al. (2017), who found a reciprocal relationship between disordered gaming and basic psychological need satisfaction, and Molde et al. (2019), who found that disordered gaming was linked to greater problem gambling 2 years later, but not the converse.

There is no sugarcoating the fact that longitudinal studies place high demands on researchers, but there are some resources to help. Ployhart and

Ward (2011) provide some initial tips. 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 may require smaller sample sizes to achieve sufficient statistical power, potentially balancing out financial costs.

If it remains unfeasible for researchers to participate in their own longitudinal data collection, one alternative is to (re)use existing non-cross-sectional 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 panel 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 2022).

Finally, the simplest thing researchers can do, whether they are in a position to conduct non-cross-sectional studies or not, is to preferentially cite longitudinal studies. During literature reviews, appending 'longitudinal', 'panel', 'experience sampling' or similar to searches can help identify work that involves a temporal component and may better support causal claims. Citing these studies instead of, or in addition to, their cross-sectional analogues rewards researchers' effort in conducting them. This in turn incentivizes those researchers and others to do more work along those lines.

To summarize, there are various strategies games researchers can use to conduct or support longitudinal research, to the benefit of the entire field:

- Preferentially seek out and cite non-cross-sectional work, particularly when it better supports causal claims
- Explore opportunities to collaborate 'piggy-back' on other studies where one or more measures overlap
- Investigate available resources for facilitating intensive and non-intensive longitudinal designs (see examples above for starting points)
- Search for existing data that can address one's question (see also the following section on Digital Trace Data)

Ultimately, these steps can help us foster a generation of researchers with more skills in conducting non-cross-sectional studies, and who are rewarded for the challenge in running such studies.

Use digital trace data

In trying to understand the impact of any gaming behavior on a particular outcome, research is governed by the 'garbage in, garbage out' rule: if our measures do not reflect the actual phenomenon of interest, our studies will not be informative. Endorsement of internet gaming disorder criteria may not reflect actual dysregulation (e.g., King et al. 2018); responses to a questionnaire in one gaming context may reference completely different experiences or constructs than in another gaming construct (Gundry and Deterding 2019); laboratory measures of aggression may not reflect actual verbal or physical aggression in the real world (Adachi and Willoughby 2011).

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

One fundamental example of this concerns playtime. Evidence is rapidly accumulating that self-report measures of digital technology use are often inaccurate (Parry et al. 2021), and that this is true of games in particular (Kahn et al. 2014; Mok and Anderson 2021; Johannes et al. 2021b). The degree of inaccuracy may even be related to well-being, creating a possible confound for certain research questions (Sewall et al. 2020).

In addition to potential issues with external validity, players can typically only self-report simple, global measures of behavior. 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—e.g., the number of chat interactions, the proportion of time spent idling, the win-loss rate over time—could be trusted.

One primary solution to these problems lies in gaining access to digital trace data (sometimes referred to as objective behavioral data, logged play data, and/or telemetry data). Digital trace data refers to records of activity carried out using digital technology and is widely collected by most software applications people use, including in games. This has several beneficial properties for research purposes. It eliminates the potential for biases in participants responses—due to demand characteristics, self-image, distorted memory, or any number of other factors—and replaces this with automatically collected behavioral logs. Trace data is collected non-intrusively, potentially reducing demand characteristics in environmental settings, and scales easily.

As above, we are fortunate to work in an era where collecting digital trace data is becoming increasingly accessible. I am aware of at least 5 ways researchers might do so, each with their own strengths and weaknesses (Table 1).

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, which may constitute only a small part of someone's total gaming.

Second, researchers can use third-party tracking tools and existing time/behavior tracking systems. These are apps or features designed to record some aspect of a person's behavior, such as the Android Digital Wellbeing menu for showing device usage and Apple's Health app to record physical activity. To date, these have been more commonly used for research studies on global screen time or social media on mobile devices (Ohme et al. 2021; Sewall and Parry 2021)—but are no less potentially powerful for gaming, and mobile gaming especially.

Third, 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, though grants have begun to be awarded to build infrastructure to support this (Araujo 2021). In the future, I hope to see a streamlined and standardized framework for players to request this data from various companies and share it with researchers, a project that could revolutionize how we understand gaming.

In the meantime, I provide an example of my own Playstation data, returned by via such a request in December 2021. To obtain this, I sent

Method of Accessing Digital Trace 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)
User-facing tracking tools	• May be customizable to track particular behavioral data of interest	• Can be restricted by terms of service agreements	iOS Screen time Rescue Time Google wellbeing Microtransaction logs
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)
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-as-byproduct	• (Typically) already publicly posted to the web	• May lack granularity • Ethical concerns about use of data beyond intended purpose of sharing	Game reviews (Petrovskaya et al. 2022; Lu et al. 2020) Forum/social media posts (Jørgensen and Bogers 2020) Trophy trackers (e.g., psnprofiles.com, trueachievements.com)

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

one request email and two brief clarifications for follow-ups; in total, the process took approximately 6 weeks. I provide an email template that may serve to accelerate future requests. Both can be found in the supplementary materials (osf.io/f9m8b/?view_only=8109888ab20c4886a5daced50af87bc9, *anonymized for review*). I encourage readers to investigate their own favorite games or platforms privacy agreements to explore if and how they can request data themselves.

Fourth, there exist a handful of publicly available APIs or data interfaces that provide access to in-game data, such as OpenDota (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).

Finally, I include a broad category of data generated as a byproduct of other player or community activities. Things like game reviews (Petrovskaya et al. 2022; Lu et al. 2020) or forum posts (Jørgensen and Bogers 2020) can be collected; more generally, webscraping with tools like Scrapy (`scrapy.org`)—where permissible—can generate large amounts of data about how players are behaving or engaging with games. Related data can include Twitch viewership and chat logs (Diwanji et al. 2020), or fan wikis where players aggregate information about and 'theorycraft' for games (Mittell 2009).

In short, as gaming companies continue to collect ever more data, researchers have the opportunity—obligation, even—to seek access of similar data. More than ever before, researchers have access to behavior at an extremely granular level. This can play a vital role in answering questions about how video games affect players, and we should seize the opportunity.

Control individual features

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. This means that any differences between the two groups cannot be easily attributed to any one aspect of the game. This is essentially a special case of an overarching challenge in games research: experiences and effects of games are subjective, holistic, emergent, situated, and dynamic (Hassenzahl 2010).

Not controlling individual features 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 (e.g., Anderson and Dill 2000). However, any two distinct games will differ not only in their violent content, but also various other aspects of design. Although some studies run pilots 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. Alongside proper randomization, this provides the best assurances possible that any differences observed between players, or within the same player trying both versions, is due to that particular feature.

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 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 (from templates), and using natural experiments.

Modding refers to the alteration of one or more aspects of an existing video game, often conducted at a grassroots level by players. Many games actively encourage modding to extend the variability and replayability of a game—games like *Minecraft*, *Skyrim*, *Stardew Valley* (to name but a few) have thousands of publicly shared mods on sites like `nexusmods.com`, and fervent communities where prospective modders can get started with only basic programming knowledge. In many cases, a mod that manipulates a feature of interest may already exist, and researchers can simply compare the modded and unmodded versions. Modding has been fruitfully applied in games research: Hilgard et al. (2019a), for example, modded *Doom II* to create versions with additional violence and reduced violence, and found that players of each did not differ in postgame aggressive behavior.

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 (from template)	• 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

Building custom games, by contrast, involves researchers designing and creating a new game from the ground up. This has been a relatively popular strategy in games research, but has also been criticized for lacking ecological validity: 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. Even better, many game engines have free or relatively low-cost game templates—skeletons of a game that can be used with only light customization for the particular use case.

Finally, one last strategy is to use planned changes to video games, or 'patches', as opportunities for natural experiments. Where possible, this strategy provides the most ecologically valid method of understanding individual features—patches affect players who are already engaged with a particular game made to commercial standards. Capitalizing on these opportunities can be difficult; researchers will often need to act quickly after a patch is announced (or be continuously monitoring their play already), and the researchers will not themselves have control over the changes made to the game. Zendle (2019) is one of the rare examples of a published natural experiment, finding that players spent less money on microtransactions after the removal of loot boxes from *Heroes of the Storm*. Other work has linked changes in

user reviews with the release dates of patches in *No Man's Sky* (Lu et al. 2020).

By controlling individual features, researchers can begin to understand the mechanics of how games affect people more carefully. Such work opens up exciting research possibilities such as the us 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

It is likely impossible to ever fully identify the 'active ingredients' of gaming experiences, given their systemic, emergent nature. Nonetheless, any progress we make in this area will be of major benefit to our understanding of the psychological dynamics of games. For a discipline that regularly invokes design implications as justification for the impact of its work, it is necessary to drill down more precisely on how particular game features—alone and in combination—are affecting players.

Adopt Open Science

Research 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.

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' (misuse of data analysis to find patterns in data that can be presented as statistically significant) 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. We have strong evidence that this is a problem in research on games, with studies finding that statistical reporting is often poor (Vornhagen et al. 2020), and that many topics are characterized by substantial publication bias (Hilgard et al. 2019b, 2017b).

The current leading movement to address these issues of questionable research practices, publication bias, and more is commonly known as open science. The tenets of the open science movement are too varied to discuss exhaustively here (but see Munafò et al. 2017, for an overview). 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.

The first practice is preregistration: specifying one's research plan in advance of the study, and submitting it to a timestamped registry. The Center for Open Science summarizes the benefits, writing 'Preregistration separates hypothesis-generating (exploratory) from hypothesis-testing (confirmatory) research. Both are important. But the same data cannot be used to generate and test a hypothesis, which can happen unintentionally and reduce the credibility of your results' (`cos.io/initiatives/prereg`).

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 practical guide to preregistration osf.io/2vu7m/
Registered Reports	• Alleviates the file drawer effect, where null results are less likely to be published • Allows studies to receive feedback before its too late to implement	• 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	• Provides invaluable information on the trustworthiness, consistency, and generalizability of effects • Can be a valuable pedagogic tool for levels from undergraduate to PhD	Zwaan et al. (2018) Brandt et al. (2014) Wagge et al. (2019)
Open data	• Creates opportunity for future reuse, saving drastically on 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 their benefits for research on games

Registered reports is a publishing format that extends preregistration, emphasizing the importance of the research question and the quality of methodology by conducting peer review prior to data collection. High quality protocols are then provisionally accepted for publication if the authors follow through with the registered methodology. This ensures that results will be published regardless of what they are (eliminates incentives to find positive or significant results), and allows reviewers to catch problems before it is too late to correct them.

Team science describes projects conducted by large groups of researchers, typically distributed over many regions of the world. This allows studies to obtain far larger sample sizes, varied participants, and leverage interdisciplinary expertise. Team science creates more opportunities for specialists in each aspect of the research process (e.g., statistics, survey design, game development). Prominent team science initiatives include the ManyLabs projects and the Psychological Science Accelerator (see psysciacc.org for more details and instructions on how to participate).

Replications are studies designed to test an existing finding in a new, analogous dataset. Replications give us information about the robustness and generalizability of a particular finding, and help

separate true effects from false positives. Replication studies double as an excellent pedagogical tool, having been successfully applied to research methods education as early as the undergraduate level (Wagge et al. 2019).

Finally, open data is the practice of making all data underlying the conclusions of a manuscript available freely and publicly at the time of publication. The benefits of open data are myriad: it increases confidence in findings by allowing others to reproduce values and potentially detect errors, enables reuse of the data to answer related research questions, ensures that those funding research (often the public) have access to its outputs, and can lead to greater collaboration and goodwill among researchers.

While the implementation of some open science practices continues to be debated, the movement as a whole nonetheless provides our best set of tools for increasing the rigor and accuracy of the research literature. 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) for remote seminars and events, and exploring the entry-level reading list collated by Crüwell et al. (2019), in addition to the tutorials provided in the table above for the specific practices I recommended here.

Discussion

Above, I 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: investigating effect size variation, strengthening the derivation chain, conducting longitudinal research, collecting logged behavioral data, manipulating individual features, and applying open science methods.

The path to widespread adoption of these 6 strategies is a difficult one, but not without reason to be optimistic. 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 digital trace data into their research designs, to adopting open science practices, to using multilevel models to explore effect size variation, to preferentially citing existing research that uses these techniques.

There already exist examples of how work following these practices can advance our knowledge. For example, work has found that preregistered studies (item 6) on violent video games and aggression, though few in number, have tended to find null effects (Ferguson 2020). A related study using digital trace data (point 4) following players over 6 weeks (point 3) similarly found no evidence for an effect of violent games on subsequent aggressive affect (Johannes et al. 2022).

Even better, these 6 strategies are synergistic. A custom-designed game can collect a wide range of digital trace data data. A longitudinal study could be submitted as registered report, allowing reviewers to identify opportunities for a more informative design or analysis. Modeling effect size variation can be a valuable exploratory step that helps us better specify the boundary conditions or moderators in our theories. The usefulness of each practice can be amplified through combination with others.

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 digital technology researchers are taught, and much more. For example, 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 a byproduct of 'publish or perish' culture of academia and the ways we evaluate and hire researchers (Fanelli 2010). Rather than alleviate the need for large-scale changes to the research ecosystem, I hope that pursuing research along the 6 lines described above can empower the research community, making clear the potential gains as these practices and others become integrated into our work, and using that as a platform to push for even greater change.

As with any other aspect of improving 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 each new project, where suitable. Similarly, the diversity of skills required will necessitate specialization and collaboration. Game developers, statisticians, open research experts, theory specialists, and more all have a role to play.

Conclusion

Psychological games research has been less efficient than we would like it to have been, with answers to basic questions about how games affect people proving elusive. As our knowledge of the limitations of our research designs, (statistical) analyses, and structures in the research ecosystem has advanced, however, the path forward has become clearer. We need to pay more attention to how effect sizes vary across individuals, use this to inform our theories and link them more closely to our hypotheses, and use data that reflects actual, rather than perceived, behavior. Promisingly, that road has also become better paved in recent years: there exist many more tools for running intensive or longitudinal studies, collecting digital trace data, modifying individual features of games, and adopting open science practices that increase the trustworthiness of results. Using the 6 practices outlined above, alongside others that I have either omitted or that have not yet been widely explored, I am confident that the upcoming decade of research on games can be the most productive yet.

Supplementary Materials

Supplementary materials are available on the OSF: https://osf.io/f9m8b/?view_only=8109888ab20c4886a5daced50af87bc9, *anonymous link for review*.

Acknowledgements

This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games & Games Intelligence (IGGI) [EP/S022325/1].

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.

Abd-Alrazaq A, Al-Jafar E, Alajlani M, Toro C, Alhuwail D, Ahmed A, Reagu SM, Al-Shorbaji N and Househ M (2022) The Effectiveness of Serious Games for Alleviating Depression: Systematic Review and Meta-analysis. *JMIR Serious Games* 10(1): e32331. DOI:10.2196/32331.

Adachi PJC and Willoughby T (2011) The effect of violent video games on aggression: Is it more than just the violence? *Aggression and Violent Behavior* 16(1): 55–62. DOI:10.1016/j.avb.2010.12.002.

Allahverdipour H, Bazargan M, Farhadinasab A and Moeini B (2010) Correlates of video games playing among adolescents in an Islamic country. *BMC*

Public Health 10(1): 286. DOI:10.1186/1471-2458-10-286.

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.

Araujo T (2021) Six Dutch universities start data donation project. *University of Amsterdam* .

Ballou N and Zendle D (2022) "Clinically significant distress" in internet gaming disorder: An individual participant meta-analysis. *Computers in Human Behavior* 129: 107140. DOI: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.

Beyens I, Pouwels JL, van Driel II, Keijsers L and Valkenburg PM (2020) The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports* 10(1): 10763. DOI:10.1038/s41598-020-67727-7.

Brandt MJ, IJzerman H, Dijksterhuis A, Farach FJ, Geller J, Giner-Sorolla R, Grange JA, Perugini M, Spies JR and van 't Veer A (2014) The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology* 50: 217–224. DOI:10.1016/j.jesp.2013.10.005.

Brown VA (2021) An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science* 4(1): 251524592096035. DOI:10.1177/2515245920960351.

Burke B and Lucier-Greer M (2021) Comparing video game engagement measures as related to individual and relational well-being in a community sample of adult gamers. *Computers in Human Behavior Reports* 4: 100136. DOI:10.1016/j.chbr.2021.100136.

Colder Carras M, Van Rooij AJ, Van de Mheen D, Musci R, Xue QL and Mendelson T (2017) Video gaming in a hyperconnected world: A cross-sectional study of heavy gaming, problematic gaming symptoms, and online socializing in adolescents. *Computers in Human Behavior* 68: 472–479. DOI:10.1016/j.chb.2016.11.060.

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.

DeCamp W (2017) Who plays violent video games? An exploratory analysis of predictors of playing violent games. *Personality and Individual Differences* 117: 260–266. DOI:10.1016/j.paid.2017.06.027.

Dicheva D, Dichev C, Agre G and Angelova G (2015) Gamification in Education: A Systematic Mapping Study. *Journal of Educational Technology & Society* 18(3): 75–88.

Diwanji V, Reed A, Ferchaud A, Seibert J, Weinbrecht V and Sellers N (2020) Don't just watch, join in: Exploring information behavior and copresence on Twitch. *Computers in Human Behavior* 105: 106221. DOI:10.1016/j.chb.2019.106221.

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.

Fanelli D (2010) Do Pressures to Publish Increase Scientists' Bias? An Empirical Support from US States Data. *PLoS ONE* 5(4): e10271. DOI:10.1371/journal.pone.0010271.

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.

Ferguson CJ and Colwell J (2020) Lack of consensus among scholars on the issue of video game "addiction". *Psychology of Popular Media* 9(3): 359–366. DOI:10.1037/ppm0000243.

Festic N, Büchi M and Latzer M (2021) How Long and What For? Tracking a Nationally Representative Sample to Quantify Internet Use. *Journal of Quantitative Description: Digital Media* 1. DOI:10.51685/jqd.2021.018.

Flake JK and Fried EI (2020) Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science* 3(4): 456–465. DOI:10.1177/2515245920952393.

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.

Gilbert A, Baumgartner S and Reinecke L (2021) Situational boundary conditions of digital stress: Goal conflict and autonomy frustration make smartphone use more stressful. Preprint, PsyArXiv. DOI:10.31234/osf.io/fzct9.

Guest O and Martin AE (2021) How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science* : 1–14DOI:10.1177/1745691620970585.

Gundry D and Deterding S (2019) Validity threats in quantitative data collection with games: A narrative survey. *Simulation & Gaming* 50(3): 302–328. DOI:10.1177/1046878118805515.

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.

Hughes N and Cairns P (2021) Growing Together: An Analysis of Measurement Transparency Across 15 Years of Player Motivation Questionnaires. Preprint, Open Science Framework. DOI:10.31219/osf.io/ju4s2.

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, Magnusson K and Przybylski AK (2022) Time Spent Playing Two Online Shooters Has No Measurable Effect on Aggressive Affect. *Collabra: Psychology* 8(1): 34606. DOI:10.1525/collabra.34606.

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.

Jørgensen IKH and Bogers T (2020) "Kinda like The Sims... But with ghosts?": A Qualitative Analysis of Video Game Re-finding Requests on Reddit. In: *International Conference on the Foundations of Digital Games*. Bugibba Malta: ACM. ISBN 978-1-4503-8807-8, pp. 1–4. DOI:10.1145/3402942.3402971.

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.

Karakus T, Inal Y and Cagiltay K (2008) A descriptive study of Turkish high school students' game-playing characteristics and their considerations concerning the effects of games. *Computers in Human Behavior* 24(6): 2520–2529. DOI:10.1016/j.chb.2008.03.011.

Kersten R and Greitemeyer T (2022) Why do habitual violent video game players believe in the cathartic effects of violent video games? A misinterpretation of mood improvement as a reduction in aggressive feelings. *Aggressive Behavior* 48(2): 219–231. DOI:10.1002/ab.22005.

King DL, Herd MCE and Delfabbro PH (2018) Motivational components of tolerance in Internet gaming disorder. *Computers in Human Behavior* 78: 133–141. DOI:10.1016/j.chb.2017.09.023.

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.

Lakatos I (1978) *The Methodology of Scientific Research Programmes*. His Philosophical Papers ; v. 1. Cambridge ; New York: Cambridge University Press. ISBN 978-0-521-21644-9.

Lemmens JS and Hendriks SJ (2016) Addictive Online Games: Examining the Relationship Between Game Genres and Internet Gaming Disorder. *Cyberpsychology, Behavior, and Social Networking* 19(4): 270–276. DOI:10.1089/cyber.2015.0415.

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.

Liu D, Baumeister RF, Yang Cc and Hu B (2019) Digital Communication Media Use and Psychological Well-Being: A Meta-Analysis. *Journal of Computer-Mediated Communication* 24(5): 259–273. DOI:10.1093/jcmc/zmz013.

Lo SK, Wang CC and Fang W (2005) Physical Interpersonal Relationships and Social Anxiety among Online Game Players. *CyberPsychology & Behavior* 8(1): 15–20. DOI:10.1089/cpb.2005.8.15.

Loton D, Borkoles E, Lubman D and Polman R (2016) Video Game Addiction, Engagement and Symptoms of Stress, Depression and Anxiety: The Mediating Role of Coping. *International Journal of Mental Health and Addiction* 14(4): 565–578. DOI:10.1007/s11469-015-9578-6.

Lu C, Li X, Nummenmaa T, Zhang Z and Peltonen J (2020) Patches and Player Community Perceptions: Analysis of No Man's Sky Steam Reviews. In: *Proceedings of the 2020 DiGRA International Conference: Play Everywhere*. p. 24.

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.

Marquet O, Alberico C and Hipp AJ (2018) Pokémon GO and physical activity among college students. A study using Ecological Momentary Assessment. *Computers in Human Behavior* 81: 215–222. DOI:10.1016/j.chb.2017.12.028.

Mayo DG (2018) *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. First edition. Cambridge University Press. ISBN 978-1-107-28618-4 978-1-107-05413-4 978-1-107-66464-7. DOI:10.1017/9781107286184.

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.

Mittell J (2009) Sites of participation: Wiki fandom and the case of Lostpedia. *Transformative Works and Cultures* 3. DOI:10.3983/twc.2009.0118.

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.

Molde H, Holmøy B, Merkesdal AG, Torsheim T, Mentzoni RA, Hanns D, Sagoe D and Pallesen S (2019) Are Video Games a Gateway to Gambling? A Longitudinal Study Based on a Representative Norwegian Sample. *Journal of Gambling Studies* 35(2): 545–557. DOI:10.1007/s10899-018-9781-z.

Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, Simonsohn U, Wagenmakers EJ, Ware JJ and Ioannidis JPA (2017) A manifesto for reproducible science. *Nature Human Behaviour* 1(1): 0021. DOI:10.1038/s41562-016-0021.

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.

Petrovskaya E, Deterding S and Zendle DI (2022) Prevalence and Salience of Problematic Microtransactions in Top-Grossing Mobile and PC Games: A Content Analysis of User Reviews. In: *CHI Conference on Human Factors in Computing Systems*. New Orleans LA USA: ACM. ISBN 978-1-4503-9157-3, pp. 1–12. DOI:10.1145/3491102.3502056.

Ployhart RE and Ward AK (2011) The "Quick Start Guide" for Conducting and Publishing Longitudinal Research. *Journal of Business and Psychology* 26(4): 413–422. DOI:10.1007/s10869-011-9209-6.

Przybylski AK and Weinstein N (2017) A large-scale test of the Goldilocks Hypothesis: Quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychological Science* 28(2): 204–215. DOI:10.1177/0956797616678438.

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.

Rohrer JM (2018) Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science* 1(1): 27–42. DOI:10.1177/2515245917745629.

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.

Szabelska A, Pollet TV, Dujols O, Klein RA and IJzerman H (2021) A Tutorial for Exploratory Research: An Eight-Step Approach. Preprint, PsyArXiv. DOI:10.31234/osf.io/cy9mz.

Thai S and Page-Gould E (2018) ExperienceSampler: An open-source scaffold for building smartphone apps for experience sampling. *Psychological Methods* 23(4): 729–739. DOI:10.1037/met0000151.

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.

Vornhagen JB, Tyack A and Mekler ED (2020) Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Virtual Event Canada: ACM. ISBN 978-1-4503-8074-4, pp. 4–18. DOI:10.1145/3410404.3414229.

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.

Wagge JR, Brandt MJ, Lazarevic LB, Legate N, Christopherson C, Wiggins B and Grahe JE (2019) Publishing Research With Undergraduate Students via Replication Work: The Collaborative Replications

and Education Project. *Frontiers in Psychology* 10: 247. DOI:10.3389/fpsyg.2019.00247.

Wallach L and Wallach MA (2010) Some Theories are Unfalsifiable: A Comment on Trafimow. *Theory & Psychology* 20(5): 703–706. DOI:10.1177/0959354310373676.

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.

Wenzel HG, Bakken IJ, Johansson A, Götestam KG and Øren A (2009) Excessive Computer Game Playing among Norwegian Adults: Self-Reported Consequences of Playing and Association with Mental Health Problems. *Psychological Reports* 105(3_suppl): 1237–1247. DOI:10.2466/PR0.105.F.1237-1247.

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.

Zwaan RA, Etz A, Lucas RE and Donnellan MB (2018) Making replication mainstream. *Behavioral and Brain Sciences* 41: e120. DOI:10.1017/S0140525X17001972.