MASARYK UNIVERSITY
Faculty of Social Studies

Addressing measurement and evidence quality idealizations in gaming research

Habilitation thesis

Brno, 2024 Mgr. Ivan Ropovik, Ph.D.

Abstract

Scientific inquiry necessarily rests on layers of idealizations, yet if they misalign with reality, they impede our ability to discern the verisimilitude of our theories. The gradual elimination of idealizations thus represents a key mechanism of scientific progress. This habilitation thesis presents a collection of eight methodological and applied research studies united along a common thread –elimination of idealizations– with gaming research as the case study. The included papers address two critical domains where oversimplified assumptions potentially compromise scientific inferences: measurement (at the study level) and evidence quality (at the literature level). The thesis begins with a reflection on the ontological and psychometric implications of treating mental disorder as a latent common cause model vs a complex dynamic system (network model). Through a series of large-scale empirical studies, we examined how different symptom operationalizations change the structural properties of gaming disorder symptom network (N = 3,015 gamers, N = 801 esports players), the relationships between alternative gaming disorder operationalizations and functional impairments (N = 1,009), and the temporal dynamics of symptom networks over a six-month period (N = 1,320). The focus then shifts to tackling idealizations in evidence synthesis. In two meta-studies, we showcase statistical approaches for assessing the empirical robustness of a finite set of reported findings and for adjusting publication bias. These advances are then applied in two comprehensive meta-analyses: one examining the effectiveness of video games in changing attitudes (k = 119; N = 14,272), and another synthesizing evidence on risk and protective factors for gaming disorder (k = 1,586; N = 210,557). Both meta-analyses implemented state-of-the-art bias adjustments and a complex appraisal of evidence quality. Together, the included studies attempted to systematically examine and reduce reliance on some of the idealizations through psychometric and meta-scientific approaches, while acknowledging both the inherent complexity of studied phenomena and the provisional nature of scientific knowledge.

I couldn't have asked for better company on this academic adventure – my wonderful colleagues who became friends along the way and my happy little family with their unwavering love, care, and support. My warmest thanks. And also to you, dear reader, for choosing to spend your valuable time with these pages.

Table of Contents

Introduction

"The whole art and practice of scientific experimentation is comprised in the skillful interrogation of Nature. … Far from behaving consistently, however, Nature appears vacillating, coy, and ambiguous in her answers. She responds to the form of the question as it is set out in the field and not necessarily to the question in the experimenter's mind" (Box, 1978, p. 136). In her biography of her father Ronald Fisher, Joan Fisher Box beautifully distills the essence of the relentless struggle scientists face in attempting to interrogate nature. When studying any phenomenon, a vast array of potential research designs and analytical setups exists, each constituting a distinct research question and leading to a result — each correct within its own context. Often, however, only a small fraction of these potential setups align with the researcher's intentions. Misalignments usually leads to what is being labeled as errors, biases, or even nonsensical results.

When studying complex, causally multidetermined systems like the human mind, our designs and models necessarily rest on layers of idealizations, usually in the form of theoretical and instrumental auxiliary theories (Meehl, 1978). Idealizations are propositions that we know to be false but that we find useful for scientific purposes (Bokulich, 2011). Most scientific models involve such idealizations. Aligning the substantive research question in the researcher's mind with the design and model requires identifying the most salient —yet often implicit and non-obvious— idealizations inherent in methodological and analytic choices.

However, the sword of scientific idealization cuts both ways. On one hand, we need simplifying idealizations to acquire information (Coombs, 1964), making scientific inquiry feasible and our models interpretationally tractable. On the other hand, these idealizations may threaten the alignment with our intentions for realist interpretations. In other words, explaining how the world works using convenient fictions carries inherent risks that observed results are driven more by the idealizations than by the phenomena under study (Odenbaugh, 2011). Idealizations that are misaligned with reality also hamper progress as they deprive our data of their power to falsify wrong theories (Popper, 1962).

The nature of psychological subject matter makes it inherently difficult to develop strong, mathematically precise theories involving high-level inductions like those in physics (Meehl, 1978). As a result, much of early scientific inquiry is typically conducted within highly idealized frameworks (Vassend, 2020). Development and innovation in psychometrics, research methodology, and statistical modeling then aim to eliminate idealizations that regressively affect theory development. Assuming our commitment to epistemic realism

which arguably pervades scientific methodology (Hood, 2013; Odenbaugh, 2011), we typically aim to replace those by idealizations that are less remote from reality, gradually increasing the verisimilitude of our theories (Vassend, 2020). Findings can then be regarded as empirically robust if distinct sets of alternative idealizations lead to similar results. Truth is consequently the intersection of independent lies (Levins, 1966).

The elimination of some of the idealizations and the pursuit of empirical robustness is the common thread underlying the collection of methodological and applied research papers included in this thesis. The applied research focuses specifically on gaming, a rapidly expanding field that has received substantial cultural and academic attention. As gaming continues to attract a growing and diverse population of players, addressing issues such as measurement and evidence quality becomes critical to understanding its nature and broader implications.

Digital Gaming

Digital gaming has become one of the most pervasive forms of global entertainment, engaging individuals of all ages (Granic et al., 2014). Recent estimates suggest that over 3.4 billion people worldwide play digital games, highlighting the medium's extensive reach and cultural significance (Newzoo, 2024). The accessibility of games across multiple platforms – including computers, consoles, tablets, and mobile devices– has transformed leisure activities and social interactions (Kowert & Quandt, 2016).

While most individuals engage in gaming for enjoyment, relaxation, and social connection (Przybylski et al., 2010), a subset develops problematic patterns that may lead to negative outcomes (Király et al., 2020). Recognizing this issue, the World Health Organization included gaming disorder (GD) in the 11th revision of the International Classification of Diseases (ICD-11) in 2019 (World Health Organization, 2019). Recent meta-analyses estimate that approximately 3–4% of gamers may exhibit symptoms consistent with GD, though prevalence rates vary due to differences in diagnostic criteria and assessment methods (Stevens et al., 2021). The COVID-19 pandemic has further impacted gaming behavior, with studies suggesting increased gaming time and potential exacerbation of gaming-related problems (King et al., 2020). This highlights the need for effective prevention strategies that rely on sound measurement and a thorough understanding of the risk factors associated with GD (Paulus et al., 2018).

At the same time, video games have evolved into sophisticated narrative media capable of influencing players' attitudes and beliefs (Bachen et al., 2016). Narrative games,

which integrate storytelling elements and allow players to interact with the narrative, have the potential to influence perceptions of social, cultural, and ethical issues (Green & Jenkins, 2014). Educational and serious games are increasingly being used to promote attitudinal and behavioral change on topics such as health, environmental awareness, or social justice (Boyle et al., 2016; Soekarjo & van Oostendorp, 2015). Despite the growing interest, empirical evidence on the effectiveness of video games in promoting attitude change remains, however, limited and fragmented (Ruggiero, 2015; Baranowski et al., 2016).

In summary, digital gaming has become a dominant global phenomenon, extending its influence beyond leisure into areas such as social interaction, education, and health. However, the complexities of both the beneficial and problematic aspects of gaming behavior, as well as the potential of games to shape attitudes, underscore the need for careful scrutiny of the idealizations made in the study of gaming.

The studies included in this thesis have –apart from substantive aims related to gaming– sought to address idealizations on two fronts: measurement (at the study level) and evidence quality, especially with respect to selection biases inherent in the publication process (at the literature level).

Measurement

Implicit hypotheses about measurement relationships tend to be as crucial as structural substantive hypotheses. Misspecification of measurement models has been shown to significantly distort parameter estimates and lead to excessive error rates, especially when researchers fail to properly model the relationships between constructs and indicators, such as the direction of causality, the interchangeability of indicators, or the covariation among indicators (MacKenzie et al., 2005). Regarding the measurement aspect of gaming research, the studies included in this thesis tackle some of the critical idealizations inherent in gaming disorder measurement models. Rather than accepting the simplifying idealization of GD as a unitary latent cause of its symptoms, three of the included papers conceptualize GD as a complex dynamic system. Substantively, these papers also explore alternative operationalizations of GD symptoms, their impact on the symptom network and character of the disorder, and the relationships of different operationalizations of GD with multiple functional impairments. These psychometric explorations were conducted within the framework of the network approach –a realist-based ontological stance toward psychopathology that posits that mental disorders are relatively stable emergent states arising

from pronounced and recurrent interactions among causally linked symptoms (Borsboom, 2017).

Network approach

Coherent semantic interpretations of most measurement claims in psychology are based on realist commitments regarding the psychological attributes being measured (Cronbach & Meehl, 1955; Maul, 2013). One such model is the traditional common cause (or reflective) model, the cornerstone of psychological measurement. When applied to psychopathology, it posits that the symptoms of a mental disorder are merely manifestations of an underlying latent variable - the disorder itself (Borsboom, 2008; Kendler, 2012). It is implicitly at work whenever complex symptomatology is summarized by a single number or categorical state. This model is anchored in the assumptions of local independence and unidimensionality, implying that symptoms are conditionally independent given the latent variable and do not directly influence one another (e.g., Holland & Rosenbaum, 1986). Within this model, any observed associations among symptoms are attributed solely to their common origin in the latent disorder, effectively rendering symptoms passive indicators, i.e., solely the receivers of causal effect, rather than active agents in the psychopathological process.

However, these idealizations may oversimplify the intricate and dynamic reality of mental health phenomena. Another realist account, the network approach, challenges this paradigm by dispensing with the assumptions of a common latent cause and local independence among symptoms (Borsboom & Cramer, 2013). Instead, it conceptualizes mental disorders as emergent properties arising from direct, dynamic interactions among the symptoms themselves (Borsboom, 2017; Fried et al., 2017). Such complex system can have novel properties irreducible to more basic truths about components (Maul, 2013). This perspective recognizes symptoms as active components that can initiate, maintain, and exacerbate one another through reciprocal relationships, thereby capturing the complexity and heterogeneity inherent in psychopathology (Fried & Cramer, 2017).

For example, in the context of gaming disorder, using gaming to relieve negative moods can lead to tolerance development, requiring increasingly longer gaming sessions to achieve the same emotional relief. As more time is required for gaming, individuals experience loss of interest in previously enjoyed activities. This narrowing of interests further increases reliance on gaming as the primary source of satisfaction, strengthening the need for longer gaming sessions to maintain emotional regulation (Dong & Potenza, 2014). Such

recursive interactions contribute to the maintenance and progression of GD, a dynamic that traditional models often fail to capture due to their assumption of symptom independence (Scheffer et al., 2024; Zajac et al., 2017).

By modeling psychopathology as a network of interrelated symptoms, the network approach offers a more nuanced understanding that aligns closely with the multifaceted nature of mental disorders (Borsboom, 2017; Contreras et al., 2019). It allows for the identification of central symptoms –those that exert significant influence within the network– and critical pathways that can serve as strategic targets for intervention (Robinaugh et al., 2016; Blanken et al., 2019). For instance, if GD symptoms such as withdrawal and loss of control emerge as central nodes, interventions focused on these symptoms may have a cascading effect in reducing overall symptomatology (Kim et al., 2016).

In addition, the network approach facilitates the analysis of both contemporaneous and temporal relationships among symptoms (Bringmann et al., 2013; Epskamp et al., 2018b). Contemporaneous networks capture the co-occurrence of symptoms at a single point in time, shedding light on immediate interactions, while temporal networks elucidate how symptoms influence each other over time, providing insights into the temporal dynamics and evolution of the disorder (Haslbeck et al., 2022).

By reducing reliance on idealizations such as the common cause and local independence assumptions, the network approach seeks models that more accurately reflect clinical reality. It acknowledges the heterogeneity and individual differences in symptom patterns and trajectories inherent in psychopathology (Fried & Nesse, 2015). This aligns with the movement toward personalized medicine in psychiatry, allowing for an individualized understanding of each person's unique symptom network and facilitating tailored interventions (Fisher et al., 2018).

The network approach also addresses the pervasive issue of comorbidity in mental health disorders. Traditional models often struggle to explain the high rates of co-occurrence between disorders (Cramer et al., 2010; Kendler et al., 2011). Rather than attributing comorbidity to shared latent variables or overlapping diagnostic criteria, the network approach conceptualizes it as arising from direct relationships between symptoms of different disorders (Fried et al., 2017). This framework identifies bridge symptoms that connect symptom clusters across disorders, potentially explaining the development of comorbid conditions through symptom-symptom interactions (Jones et al., 2021).

In practice, the network approach employs advanced statistical methods to estimate the structure of symptom networks, using graphical models where nodes represent symptoms

and edges represent the strength and direction of relationships between them (Epskamp et al., 2018). These models can reveal complex patterns that are not apparent in traditional factor analyses, such as tightly connected communities or central hub symptoms (e.g., Robinaugh et al., 2016). In GD research, network analyses have identified key symptoms that may drive the disorder, such as compulsive use and withdrawal, providing valuable targets for clinical intervention (Zajac et al., 2017).

Importantly, the network approach does not dismiss the potential role of latent variables altogether but reframes them within a system of direct symptom interactions. It acknowledges that while latent variables may exist, the observable symptom interactions are critical to understanding and treating mental disorders. This perspective encourages a shift from a purely reductionist view to one that embraces complexity and dynamism, integrating both latent and observable processes (Fried et al., 2017; Bringmann & Eronen, 2018). That said, common cause model can be viewed as a specific case within a broader network framework that can also incorporate latent variables when appropriate (Epskamp et al., 2017).

In summary, the network approach represents a significant advancement in the conceptualization of psychopathology by eliminating certain idealizations inherent in traditional models. By focusing on the direct interactions between symptoms and modeling mental disorders as complex, dynamic systems, it provides a more accurate and nuanced understanding of mental illness. This approach holds considerable promise for improving diagnosis, informing targeted interventions, and ultimately potentially improving outcomes for individuals with disorders such as GD.

Methodological *Study 1* of this thesis examines the quantitative, testable implications of network models and explores why the common cause model is inappropriate for most mental disorders, given its rather unrealistic assumptions and constraints. It outlines and discusses the network approach, focusing on how models based on network theory can provide insights into the etiopathogenesis of mental disorders and support clinical intervention. Limitations and future challenges of network theory are also discussed.

Gaming disorder network structures

The empirical robustness of GD network structures in response to varying symptom operationalizations constitutes a foundational theme of the first three empirical studies presented in this thesis (*Studies 2–4*). Collectively, these studies challenge the prevalent but often unexamined assumption that all GD assessment instruments uniformly measure the same latent construct of GD (Castro-Calvo et al., 2021; Karhulahti et al., 2022). By critically examining how variations in operational definitions and measurement instruments affect the architecture of GD symptom networks and their associations with functional impairments, the research offers a nuanced perspective on the disorder's conceptualization. Specifically, the studies examine the impact of using different diagnostic criteria –rooted in different nosological frameworks, such as the DSM-5 (American Psychiatric Association, 2013) and ICD-11 (World Health Organization, 2019)– on prevalence rates, symptom severity ratings, and functional impairment identification. This careful psychometric investigation is critical for refining GD measurement methods and increasing the empirical robustness of research in this rapidly evolving field.

Study 2 studied how different operationalizations of GD symptoms influence the structural properties of the disorder's symptom network. By dissecting subtle variations in symptom definitions –such as those related to withdrawal symptoms, diminished interest in other activities, or persistent use despite negative consequences– the study addresses whether (and how minor) definitional discrepancies can significantly alter the interrelationships among symptoms and the overall network topology. If true, this would underscore the susceptibility of symptom centrality and network connectivity interpretations to measurement artifacts, which has critical implications for identifying focal points in clinical interventions (Borsboom, 2017; Fried et al., 2017). For instance, a symptom deemed central in one operationalization may not emerge as central in another, potentially leading to inconsistent intervention strategies. The study thus tackles the notion of interchangeability among validated GD measures. Building upon these insights, *Study 3* broadens the scope of analysis by incorporating additional diagnostic features beyond the core symptoms specified in the DSM-5 and ICD-11 criteria. This study investigates how ancillary features –such as disruptions in dietary habits, sleep disturbances, reduced physical activity, or increased aggression– interact with core GD symptoms within the network structure (see Burleigh et al., 2019). The question was whether certain symptoms consistently emerge as highly central nodes across different network configurations, regardless of the diagnostic framework used. If this is the case, then both DSM-5 and ICD-11-based measures are equivalent in assessing

relationships between GD and external variables such as functional impairment. These would validate the comparability and synthesis of studies that use different diagnostic criteria. To further this investigation, *Study 4* employs a longitudinal design to examine developmental trajectories of GD symptoms over time. By collecting data at three distinct time points, the study disentangles patterns of temporal symptom-symptom effects, providing critical insights into the development of GD and its associated risk factors.

Collectively, these studies systematically challenge the oversimplified conceptualization of GD as a unitary construct and demonstrate that the empirical robustness of research findings may be influenced by the specific operationalizations employed. They provide implications for both research and clinical practice, particularly with regard to the comparability of findings across studies and the generalizability of results. As the field advances, it is essential for GD research to adopt more consistent and precise operational definitions, while embracing the inherent complexity of the disorder through advanced psychometric frameforks such as network analysis (Borsboom, 2017). Such an integrative and methodologically rigorous approach may not only advance the theoretical robustness of the GD construct, but also prove practically effective in identifying and addressing gaming-related problems, ultimately improving prevention and treatment strategies.

Evidence quality

In scientific research, particularly in fields exploring complex human behaviors such as gaming and its effects, theory development and policy recommendations often rely on comprehensive evidence syntheses rather than individual studies (Borenstein et al., 2009). These syntheses integrate data from diverse set of studies, increasing statistical power, providing precise estimates of effect sizes, helping elucidate relevant moderating effects, and mapping the empirical landscape of research problems. However, assuming that the published literature represents an unbiased and comprehensive account of research is itself an idealization (Ioannidis, 2005; Rothstein et al., 2005). Systemic biases inherent in the research and publication process –such as publication bias, selective reporting, and methodological or psychometric shortcomings– can threaten the empirical robustness and integrity of conclusions drawn from published findings (Song et al., 2010; Franco et al., 2014). These biases result primarily from structural incentives that favor high publication output as a key factor for career advancement, acting as a natural selection process, not even requiring any deliberate cheating or loafing on the part of scientists (Smaldino & McElreath, 2016).

Publication bias occurs when studies with statistically significant results or those that fit a theoretical narrative are more likely to be published than studies with nonsignificant or unfavorable outcomes. Authors and editors alike are known to disfavor negative results (Emerson et al., 2010; Senn, 2012). This results in a distorted evidence base that overestimates effect sizes and propagates false positives (Carter et al., 2019; Dwan et al., 2013; Hopewell et al., 2009). Applying standard random-effects meta-analytic models to real literature with an unknown level of bias thus represents a causal misspecification. Selective reporting and the misuse of data analysis techniques to look for patterns (e.g., statistical significance) further exacerbate this problem (Simmons et al., 2011; Head et al., 2015). These practices contribute to an idealized perception of the empirical literature that may not be well aligned with reality. Implementing rigorous methods to detect and adjust for these biases is therefore essential to ensure the robustness of meta-analytic conclusions, the verisimilitude of theories, and soundness of policies based on them (McShane et al., 2016).

Two of our methodological studies, *Study 5* and *Study 6*, addressed these issues by focusing on methods for assessing and improving the empirical robustness of conclusions derived from evidence syntheses. In *Study 5*, we developed a statistical workflow using selection modeling and Monte Carlo simulations to assess the robustness of a finite set of empirical findings (Simonsohn et al., 2014; van Assen et al., 2015). Although applied to a specific research question related to the replicability of cleansing effects (backing the theoretical account of grounded procedures; Lee & Schwarz, 2021), this methodology has a generic character and can assess the robustness of any set of quantitative findings. This approach allows researchers to determine whether the observed distribution of p-values associated with a finite set of collected findings is consistent with true effects or indicative of biases such as selective reporting, thereby challenging the idealization of taking published literature at face value.

In *Study 6*, we empirically examined and identified the most common practices for dealing with publication bias in evidence syntheses and developed recommendations for appropriately addressing this bias. By critically examining existing meta-analyses, we highlighted the importance of employing state-of-the-art adjustment methods, such as selection models and regression-based techniques, which are more effective in accounting for publication bias than traditional, frequently statistically unprincipled methods (Carter et al., 2019; Stanley, 2017). Neglecting or inadequately addressing publication bias builds into the conclusions the assumption that published findings are free from bias, potentially leading to erroneous substantive inferences.

Building upon these methodological foundations, we applied rigorous bias adjustment approaches and conducted comprehensive quantitative assessments of evidence quality in two meta-analyses within gaming research: one focusing on the effects of educational video games on attitude change (*Study 7*) and the other investigating risk and protective factors for gaming disorder (*Study 8*). Recognizing the necessity of assessing the empirical robustness of the findings, we employed numerous publication bias adjustments, corrections for psychometric artifacts, evidential value assessments, and various other assessments of evidence quality.

More specifically, in both meta-analyses, we used advanced statistical methods (primarily various selection models) to directly model the selection process and adjust for publication bias. As the standard random-effects meta-analytic models assume no selection in the literature, this poses a potentially a significant causal misspecification driving the nature of the published literature. The aim in out meta-analyses was thus to ensure that our conclusions were not unduly driven by bias arising from asymmetric publication odds (Carter et al., 2019; McShane et al., 2016). Importantly, we also conducted sensitivity analyses in which we varied the assumptions of our models to test the robustness of our results under different assumed data generation processes. In *Study 8*, we also adjusted for psychometric artifacts such as attenuation of the studied associations due to measurement error (unreliability, group misclassification), distorting effects of range restriction/enhancement, and corrected for collider bias to provide less biased estimates of associations between risk factors and GD (Schmidt & Hunter, 2015; Wiernik & Dahlke, 2020).

Using permutation-based p-curve analysis, we assessed the power of the literature in both studies to determine whether selective reporting could be ruled out as the sole explanation for statistically significant findings (Simonsohn et al., 2014). We examined numerical inconsistencies in reported p-values using machine-based screening tools (Nuijten et al., 2016) and estimated the median statistical power in the literature to detect theoretically relevant effect sizes (Ioannidis, 2008). Our overall appraisal of the quality of the evidence included established Risk of Bias assessments to assess potential biases arising from study design and reporting (Higgins et al., 2019). Finally, we examined other biases arising from the publication process in the GD meta-analysis, such as decline bias and citation bias (Fanelli et al., 2017; Trikalinos & Ioannidis, 2005).

We saw these steps as essential to empirically address the frequent assumption in theory building that past empirical results can all be taken at face value. This comprehensive approach is particularly important in the relatively young field of gaming research, where

robust and reliable findings are needed to inform (still early) theoretical models, guide prevention strategies, and shape policy decisions related to gaming behavior and its impact on individuals and society (Granic et al., 2014; King et al., 2020). By challenging the assumption of empirical robustness and implementing rigorous methods to assess and improve the quality of evidence, we hoped to contribute to a more accurate and nuanced understanding of gaming and its effects.

In conclusion, what links the methodological and empirical studies included in this thesis –as the guiding theme– is the attempt to empirically address various types of idealizations. Why was this important? It is because we are only justified in believing a theory if its idealizations can be eliminated (Odenbaugh, 2011). By attempting to get past the simplified ontology of GD, by attempting to model GD rather as a dynamic complex system, and by critically evaluating and adjusting for biases that threaten the validity of evidence syntheses on gaming, we aimed to enhance the empirical robustness and verisimilitude (see Vassend, 2020) of substantive conclusions in gaming research, one step at a time.

Overview of the studies

This habilitation thesis includes eight original studies, all published in CC-indexed journals. Apart from the methodological *Study 1* intended for local readership, all studies were published in Q1 journals (by AIS). I was the first author on five, senior author on two, and once the second author on those papers. My roles are also documented in the CRediT (Contributor Roles Taxonomy) statement in most of the included articles.

List of included studies and author contribution

Study 1

Ropovik, I., Adamkovic, M., & Banik, G. (2021). Mental health as a complex dynamic system: A network approach to psychopatology. *Ceskoslovenska Psychologie, 65*(1), 31-45. doi:10.51561/cspsych.65.1.31

As the first author of this methodological study, I was responsible for its conceptualization and wrote 90% of the first draft. I also managed two rounds of revisions following peer review.

Study 2

Adamkovic, M., Martoncik, M., Karhulahti, V., & Ropovik, I. (2024). Network Structures of Internet Gaming Disorder and Gaming Disorder: Symptom Operationalization Causes Variation. *Psychology of Addictive Behaviors, 38*(4), 475–487. doi:10.1037/adb0000960

In this paper, I served as the senior (last) author. I was involved in the conceptualization of the study, the design of its methodology, and the data analyses. I have not done writing on the first draft, but contributed to manuscript revisions.

Study 3

Martoncik, M., Adamkovic, M., & Ropovik, I. (2024). Network Analysis of Additional Clinical Features of (Internet) Gaming Disorder. *International Journal of Methods in Psychiatric Research, 33*(2), e2021. doi:10.1002/mpr.2021

As the senior (last) author of this study, my contributions included the study's conceptualization and the design of its methodology. Although not involved in the initial draft, I participated in manuscript revisions.

Study 4

Martoncik*, M., Ropovik*, I., & Adamkovic, M. (2024). Development of Gaming Disorder: Underlying Risk Factors and Complex Temporal Dynamics. *Computers in Human Behavior, 153*, 108112. doi:10.1016/j.chb.2023.108112

*joint first authors

Being the first author on this study, I contributed to the conceptualization, methodological design, data curation, and statistical analyses. I wrote 50% of the first draft of the paper, specifically 100% of the Method, Analysis, and Results sections, contributed to Introduction and Discussion, and worked on several rounds of revisions.

Study 5

Ropovik, I., Sparacio, A., & IJzerman, H. (2021). The lack of robust evidence for cleansing effects. *Behavioral and Brain Sciences, 44*(E18). doi:10.1017/S0140525X20000448

As the first author of this study, I conceptualized the study and its methodology, conducted the investigation (coding of studies), performed the statistical analyses, and wrote 100% of the first draft of the paper and also took the lead on writing of the two subsequent rejoinders.

Study 6

Ropovik, I., Adamkovic, M., & Greger, D. (2021). Neglect of publication bias compromises meta-analyses of educational research. *PLOS One, 16*(6), e0252415. doi:10.1371/journal.pone.0252415

In this study, I served as the lead author, responsible for the conceptualization, methodology, investigation (data collection), data curation and statistical analyses, project administration, and writing 100% of original draft. I also handled the subsequent revisions.

Study 7

Kolek, L., Ropovik, I., Sisler, V., van Oostendorp, H., & Brom, C. (2023). Video Games and Attitude Change: A Meta-analysis. *Contemporary Educational Psychology*, 102225. doi:10.1016/j.cedpsych.2023.102225

In this meta-analytic study, I was the second author involved in the conceptualization, methodological design, statistical analyses, and writing of 100% of the Method, Analysis, and Results sections of the first draft. Additionally, I contributed to the subsequent revisions of the paper.

Study 8

Ropovik*, I., Martoncik*, M., Babincak, P., Banik, G., Vargova, L., & Adamkovic, M. (2023). Risk and protective factors for (internet) gaming disorder: A meta-analysis of pre-COVID studies. *Addictive Behaviors, 139*, 107590. doi:10.1016/j.addbeh.2022.107590. *joint first authors

In this paper, as the first author, I was responsible for the study's conceptualization, project administration, methodology, investigation (coding of studies), data curation, and statistical analyses. I wrote 100% of the first draft of the Methods, Analysis, Results sections, co-wrote (~30%) the Introduction and Discussion sections and managed all its revisions.

Summary of research questions

The thesis starts with a methodological *Study 1*, examining the psychometric implications of network models in contrast to common cause models for psychopathology. *Studies 2-4* then form a connected series investigating GD network structures: *Study 2* analyzes how different symptom operationalizations affect network properties across multiple measures, *Study 3* expands this by examining how specific GD operationalizations are related to various functional impairments, and *Study 4* employs longitudinal analysis to examine

temporal dynamics of GD symptom networks and examine the link to changes in risk factors of GD. The final four papers address methodological concerns in gaming research at the level of the literature. *Study 5* develops and showcases statistical workflow for assessing the robustness of empirical findings, *Study 6* examines practices for handling publication bias in evidence syntheses, while *Studies 7 and 8* apply these state-of-the-art methodologies in meta-analyses of gaming research.

To lay out the aims in more detail, *Study 1* tackles fundamental questions about how we conceptualize and model mental disorders, contrasting two competing psychometric theories. The paper examined whether the dominant common cause model, which underlies most current diagnostic practices and research, appropriately represents the nature of mental disorders. It reflected on how much empirically defensible are the quantitative psychometric implications of reflective latent variable models when applied on symptoms data. The study also explored whether conceptualizing mental disorders as complex dynamic systems via network models might provide a more appropriate theoretical framework. Beyond comparing these models, the paper aimed to understand the implications for clinical practice, intervention approaches, and our understanding of comorbidity.

Building on the psychometric framework laid out in *Study 1*, *Studies 2-4* form an interconnected investigation of GD networks. *Study 2* aimed to understand how GD symptoms are structurally interconnected and which symptoms play central versus peripheral roles in these networks. The study examined how DSM-5 and ICD-11 symptoms combine in network structures and identified which symptoms serve as bridges between these different diagnostic frameworks. The study further investigated how additional clinically relevant variables external to the DSM-5- and ICD-11-based symptomatology of GD, including craving, physical health neglect, and gaming time, affect the combined network structure. The key focus was understanding how different operationalizations of four critical GD symptoms –withdrawal, loss of interests, tolerance, and continued use– affect network properties. The study also sought to determine whether GD network structures differ across player groups based on play style, age, gender, gaming time, and various psychosocial characteristics.

Study 3 shifted the focus to examine how different GD operationalizations are linked to twenty additional diagnostic/clinical features or problems typically associated with GD in diagnostic manuals as a functional impairment. These additional diagnostic features were modeled in networks including the GD, where it was hypothesized, that GD will exhibit the highest degree of network centrality. Through examining multiple measures of GD including

GDS9-SF, GDSS, GDT, GAMES test, and self-assessment measures, the study aimed to understand if ICD-11-based operationalization of GD would be associated with functional impairment more strongly compared to the DSM-5-based GD operationalizations. The study further explored whether participants falling above diagnostic thresholds on different measures showed systematically different profiles of negative outcomes, seeking to understand the diagnostic utility of various screening measures.

Study 4 extended the study of GD network to a longitudinal framework, representing the first large-scale investigation of gaming disorder's temporal dynamics. The study examined how GD symptom levels and various risk factors evolved over a six-month period, mapping both the aggregate trends and individual differences in developmental trajectories. We investigated questions like how variable is the rate of change across the population, how stable will GD be over a six-month follow-up, whether temporal stability of GD depend on its operationalization, or how are the various levels of GD symptoms associated with the trend in GD going forward? The study primarily aimed to identify and describe subgroups with distinct developmental trajectories of GD and establish whether the trend in GD could be predicted by baseline levels or rates of change in thirteen measured risk factors. We also studied, which symptoms are central to the development of GD, and which are rather peripheral. Beyond examining simpler temporal patterns, the study also sought to understand GD development as a dynamic complex system, attempting to disentangle the intricate network of relationships on within-subjects temporal and contemporaneous levels as well as on a stationary between-subjects level.

The final four papers shifted focus to the issues of evidence quality in gaming research at the literature level. *Study 5* showcased a systematic approach for assessing the empirical robustness of a finite set of reported empirical findings backing a substantive claim, using literature on cleansing effects as a case study. The study empirically investigated whether the pattern of data underlying successful replications was improbable and rather consistent with selective reporting. The study thus addressed the question whether meta-analytic approaches presented by researchers in the target article on grounded procedures (Lee & Schwarz, 2021) were likely to provide support for the replicability claim even if that were false. The paper also briefly outlined the steps that are required to adjust for biases in body of evidence backing a research program.

Study 6 relatedly examined practices of handling publication bias in published research syntheses (with the field of education as a case study). Specifically, the study examined whether and how these published meta-analyses attempted to detect and correct for

publication bias. The study tackled questions like what proportion of meta-analyses employed bias correction, what methods they used, and whether they meaningfully incorporated bias-adjusted estimates into their conclusions. Beyond simply documenting current practices, the methodological part of the paper aimed to demonstrate why appropriate state-of-the-art adjustment should be carried out by default –modeling the selection-for-publication process more truthfully– and how the uncertainty inherent in meta-analytic inference under bias should be acknowledged.

Building on the methodological frameworks developed in *Studies 5* and *6*, in *Study 7*, we conducted the first comprehensive meta-analysis of video games' effects on attitudinal change, examining whether narrative games could induce changes in both explicit and implicit attitudes. The study investigated whether intervention duration strengthened these effects, whether explicit attitude changes differed based on the comparator group, how different game mechanics affected attitudes – testing whether implicit attitudes were more influenced by stereotyping and meaningful feedback while explicit attitudes were hypothesized to respond more to perspective-taking and meaningful feedback mechanics. For game genres, the study investigated whether action games had stronger effects on implicit versus explicit attitudes, and if non-action games showed larger effects on explicit attitudes. The study also explored temporal patterns in attitude change and demographic moderators like age and education, aiming to map how various game characteristics and contexts influenced attitudinal outcomes.

Lastly, *Study 8* aimed to provide the most comprehensive synthesis of risk and protective factors for GD to date, examining relationships between GD and psychological, psychopathological, demographic, maladaptive personality traits, social, and gaming-related factors. We estimated the strength of many predictive associations while investigating how effects differed across aggregate categories. This, so called psychometric meta-analysis also examined several theoretically relevant moderators of the meta-analytic effects including gender composition of the sample, mean age, sample type, and GD assessment criteria. The study aimed to understand which individual risk factors showed the strongest relationships with GD, which protective factors demonstrated reliable effects, and how these findings aligned with or challenged prominent theoretical models of GD. Through this comprehensive examination of risk and protective factors across different domains and populations, along with a rigorous mapping of the quality of the underlying evidence, the substantive objective of the study was to establish the empirical foundation for selectively targeting the at-risk

population and tailoring effective interventions at the level of these factors in the population at-large.

Together, these eight studies progressed from fundamental theoretical questions about modeling psychological disorders through specific applications in GD research, to broader methodological contributions to research synthesis and bias correction, culminating in two comprehensive evidence syntheses of research on gaming. The latter studies built upon and implemented methodological insights from earlier ones, with the empirical papers applying the psychometric frameworks established in the psychometric and methodological/meta-research studies.

Summary of methods

Study 1 had a methodological character, examining theoretical and psychometric implications of different approaches the ontology of mental disorders. While the paper utilized both our own empirical and simulated data, these served purely explanatory purposes to illustrate the theoretical arguments presented.

Study 2 employed a cross-sectional design with two large, international, and culturally diverse samples of digital game players (N = 3,015) and esports players (N = 801) recruited via an online platform Prolific. Multiple data quality control procedures were implemented, including bot detection, attention checks, and screening for careless response patterns. The study measured GD using multiple instruments: DSM-5-based Internet Gaming Disorder and ICD-11-based Gaming Disorder measures. Alternative operationalizations of four symptoms (withdrawal, loss of interests, tolerance, and continued use) were drawn from several validated measures. The analysis employed regularized partial correlation networks to model symptom relationships, with extensive stability and accuracy robustness analyses. Network comparison tests were carried out to assess structural differences between standard and alternative symptom operationalizations. The study examined network invariance across multiple subgroups through formal network comparison procedures, testing whether network structures differed based on play style, age, gender, gaming time, and various psychosocial characteristics, with continuous moderators being additionally categorized using the conditional inference trees method.

Study 3 investigated a sample of intensive digital game players (N = 1,009) who reported playing at least 13 hours weekly. The sample was screened for participants exhibiting improbable responses and careless responding patterns. The study employed six different GD measures, including standardized measures (GDS9-SF, GDSS, GDT, GAMES

test) and self-assessment items. Additionally, 18 diagnostic features and clinical problems reported in diagnostic manuals were measured to assess functional impairment. Network modeling using regularized partial correlation networks based on polychoric correlations was used to examine the complex interconnections and clusterings of GD symptoms and these additional features (as factor scores, accounting for the ordinal character of the data and modeled as individual nodes). We also conducted stability analyses using bootstrapping-based methods, and formal network comparisons to test for differences in centrality indices between different GD operationalizations.

Study 4 employed a longitudinal panel design with three measurement waves over six months. The sample comprised digital game players (N = 1,320), pre-screened using similar methods as in *Studies 2* and *3*. Attrition rates were 19.6% at wave 2 and 17.0% at wave 3. GD was measured using multiple operationalizations, along with eleven previously identified risk and protective factors. The analytical approach combined several advanced modeling techniques. First, for each combination of three GD operationalizations and 13 risk factors, latent growth curve modeling was carried out to capture the developmental trajectories of change in GD symptoms over time at an individual level, examine whether baseline level in risk factors predict the rate of change, and the association between the rates of change of GD and different risk factors. Linear mixed-effects models were used to analyze symptom-level changes over time while accounting for individual variation in both initial levels and rates of change. Second, we used growth mixture modeling to identify distinct population subgroups with different individual developmental trajectories of GD symptoms and analyzed how these classes differed in various characteristics such as gender, age, and gaming time. Lastly, for exploratory purposes, we also fitted a series of graphical vector-autoregression network models. We modeled three distinct types of symptom networks: temporal network (Granger-causal), contemporaneous network, and between-subjects network, with the aim to disentangle and shed more light on the within-subject dynamics of complex relationships among the symptoms.

In *Study 5*, we conducted a meta-research examination of the evidence robustness underlying a theoretical account of grounded procedures. We tracked and coded the exact p-values reported for 23 focal effects from replication studies of cleansing effect. These effects were presented by the authors of the target article as providing support for the replicability of cleansing effects (Lee & Schwarz, 2020). We then applied a one-parameter selection model (p-curve) to examine the skew of the distribution of significant p-values, testing whether selective reporting could be ruled out as the sole explanation of the findings. A permutation-

based implementation was used to handle dependencies between the p-values. In conjunction, we also employed a Monte Carlo simulation study systematically varying effect sizes and sample sizes to assess the cumulative probability of the observed (or more extreme) p-value distribution. We also collected data on validity evidence for measures used in the analyzed studies.

Study 6 systematically reviewed meta-analytic practices in educational research by examining all, then recent, meta-analyses (N = 62) published in two flagship educational review journals. We employed both hand-screening and systematic keyword searches and coded multiple variables related to various aspects of publication bias detection and correction practices. Apart from coding the use and reporting related to bias adjustment, we also examined whether bias-corrected estimates meaningfully influenced the substantive conclusions of the review.

In *Study 7*, we conducted a comprehensive meta-analysis of video games' effects on attitudinal change. We employed an iterative search string development, carrying out an adaptation of the relative recall technique. The strategy involved pilot database searches to establish a legacy set of reference studies, followed by iterative adjustment of the search string to optimize recall and precision. Eight academic databases were searched, yielding 3,832 studies. The final sample included 119 effect sizes from 58 independent-sample studies (N = 14,272). Prior to analysis, we conducted multiple influence diagnostics. The analysis employed multilevel random-effects models with robust variance estimation to account for both nesting of effects within studies and clustering due to shared participants. We tested for multiple moderator effects using corrections for multiple comparisons. In this meta-analysis, we made considerable effort to adjust for publication bias. We used several state-of-the-art bias adjustment methods – mainly selection models and regression-based models, implemented employing a number of statistical improvements (models wrapped within a multi-level estimation or permutation procedure). We also supplemented the analyses with a number of robustness analyses, varying the assumptions of the models. The meta-analysis also included a comprehensive appraisal of the evidence quality through multiple approaches: we assessed the risk of bias using the Risk of Bias 2 tool for randomized trials across five domains using an algorithmic approach; GRIM and GRIMMER tests to check mathematical consistency of reported means and standard deviations with the reported Ns; we carried out a machine-based screening that examined the consistency of reported p-values with test statistics; and p-curve analysis testing for evidential value using a permutation-based procedure to handle dependencies among the effects.

Study 8 synthesized evidence on risk and protective factors for GD through a systematic review and meta-analysis of pre-COVID literature. Literature search involved database searches and forward citation tracking of papers citing 13 of the most used GD measures. The synthesis involved 1,586 effects from 253 studies (N = 210,557). The analysis employed multilevel random-effects models and robust variance estimation to handle various effect dependencies. Prior to analyses, we carried out comprehensive diagnosis of the model for each correlate, screening for influential outliers using multiple influence diagnostics indices. The study implemented several layers of bias adjustments. For psychometric artifacts, we employed psychometric meta-analysis to adjust for attenuation due to measurement error (unreliability, group misclassification), and selection effects of range restriction/enhancement and collider bias using the artifact-distribution method. For publication bias adjustment, we implemented multiple approaches: permutation-based multiple-parameter selection models and Bayesian model-averaging selection models to account for uncertainty in model selection. These were complemented by multilevel regression-based PET-PEESE models. To examine the variability in adjusted estimates under different bias severity scenarios and assumptions about the selection process, we computed a series of step function selection models and robustness analyses. The synthesis also included an in-depth assessment of evidence quality through multiple approaches: we carried out machine-based screening of statistical reporting inconsistencies, and employed permutation-based p-curve analysis to test for evidential value in the literature. We also examined temporal trends in study precision, investigated potential citation bias and decline effects through covariate-adjusted models. The robustness of findings was further examined through several sensitivity analyses varying the arbitrary analytical decisions.

General Discussion

Studies included in the present habilitation thesis all aimed to empirically address some of the idealizations that potentially compromise the robustness of scientific knowledge. Within the domain of gaming research as the case study, the focus was on measurement-related idealizations at the study level and evidence quality-related idealizations at the literature level. While idealizations are necessary for the feasibility of scientific inquiry, their gradual elimination represents a key mechanism of scientific progress (Vassend, 2020; Bokulich, 2011). The studies presented here attempted to systematically examine and reduce reliance on such idealizations through psychometric and meta-scientific approaches.

At the measurement level, our investigations yielded a more nuanced picture of GD symptomatology. While *Study 2* demonstrated significant sensitivity of network structures to operational definitions - with even minor changes in symptom operationalization leading to meaningful differences in network architecture, *Study 3* revealed an important boundary condition to this finding. Although symptom-level network properties varied with operationalization, relationships between GD and external functional impairments remained largely invariant across different measurement approaches. This pattern suggests that while precise operational definitions matter for understanding the nature and internal dynamics of the disorder, broader syndrome-level relationships may be more robust to measurement variations. These findings parallel recent work in other areas of psychopathology showing that while construct content can be sensitive to measurement choices (see Fried et al., 2021; Fried et al., 2013), relationships with external criteria often show greater stability (Ko et al., 2020).

The results challenge two separate measurement-related idealizations common in psychological research. First, they show that theoretical precision in the operationalization at the item level can significantly affect the internal structure of a construct. That would imply that from the psychometrics' point of view, a deeper understanding of the nature of the disorder requires a sort of epistemic iteration, an iterative exchange between theory and measurement (Fried et al., 2022). Such a level of precision may, however, not be necessary for studying relationships with factors external to the disorder. This aligns with recent theoretical work suggesting that the pursuit of extreme precision in psychological measurement may for some research scenarios reflect misplaced priorities (Scheel et al., 2022). Second, the findings suggest that different levels of analysis (item-level versus syndrome-level) may require different measurement approaches, challenging the assumption

that a single measurement framework can adequately capture complex psychological phenomena at all levels (van Bork et al., 2022).

Our longitudinal network analysis (*Study 4*) provided additional insights by revealing some aspects in the temporal dynamics of GD symptoms. The findings showed that while symptoms generally exhibited slight negative trends, individuals with initially high symptom levels demonstrated greater stability. This suggests potential attractor states in the symptom network, with core symptoms such as escapism and social isolation perpetuating the maintenance of the disorder. If true, this represents another layer of alignment with recent theoretical developments in dynamic systems approaches to psychopathology (Robinaugh et al., 2023; Haslbeck et al., 2022; Scheffer et al., 2024) and emerging evidence that mental disorders may be better understood as stable patterns of symptom interactions rather than manifestations of discrete latent entities (Bringmann et al., 2023).

At the literature level, our work yielded several insights about both the nature of publication bias in gaming research and practices related to its adjustment. Our systematic examination of meta-analytic practices (*Study 6*) revealed concerning patterns in how the publication bias is being handled even in papers published in flagship educational review journals. Most syntheses continue to rely on outdated methods like trim-and-fill or fail-safe N, and crucially, even when bias adjustment is performed, it rarely influences substantive conclusions. This reflects a broader misunderstanding - treating publication bias as a property of a set of individual studies that can be "tested for" rather than fundamentally an inherent property of the research process. As Morey (2013) argues, just as a sequence from a biased randomizer remains biased regardless of its properties, any product of a biased publication process carries that bias, whether or not individual studies show evidence of it. Sampling a seemingly unbiased set of studies does not make the underlying process unbiased. This was the reason why we approached bias adjustment as a necessary default step rather than making it contingent on statistical bias detection (Carter et al., 2019; McShane et al., 2016). In the particular case of our two meta-analyses, this principled approach revealed that publication bias did not seem to severely distort effect estimates in either meta-analysis (*Studies 7* and *8*). It has to be noted, however, that the majority of the analyzed effects in *Study 8* were not the primary focus of the original papers and thus less subject to selective reporting pressure. This pattern aligns with recent meta-scientific work showing that publication bias tends to be more severe for headline study findings compared to non-focal findings (Mathur & VanderWeele, 2021). Other indicators of evidence quality, however, warranted attention. In *Study 7* examining video games' effects on attitude change, 56% of the included studies were judged

to be at high risk of bias across various risk domains. The statcheck analysis revealed that for 11% (*Study 7*) and 14% (*Study 8*) of the extracted results, the p-value was inconsistent with the test statistics, usually leading to the opposite conclusion regarding the presence of the effect. Relatedly, 22% (*Study 7*) of effects targeting outcomes measured on discrete scales were flagged as being based on mathematically impossible means and/or standard deviations given the reported sample sizes. The median statistical power to detect small effects (d = .20) was only .36 (*Study 7*) and .58 (*Study 8*), though power for detecting medium effects was adequate. On the other hand, the synthesized literature exhibited adequate evidential value in both meta-analyses.

The issues raised in this thesis point to broader epistemological questions about the nature and limitations of scientific knowledge in psychology. As Hayek (1974) presciently noted in his Nobel lecture "The Pretense of Knowledge", social sciences often suffer from what he termed the "scientistic" attitude – an attempt to mechanically apply methods of physical sciences to phenomena of fundamentally different complexity. In psychological science, this often manifests as a retreat to a sort of synthetic certainties – the substitution of statistical rituals and formally registered procedures for genuine theoretical understanding, providing an illusory sense of precision and control over complex psychological phenomena (Proulx & Morey, 2021). Hayek's observation thus remains remarkably relevant to contemporary psychological science, where the pressure to produce seemingly precise quantitative knowledge may lead to overlooking the inherent complexity and multideterministic nature of psychological phenomena (Devezer et al., 2021).

The replication crisis in psychology has further emphasized these limitations, revealing how seemingly robust findings can fail to replicate even under carefully controlled conditions (Nelson et al., 2018). This crisis has prompted a deeper examination of how we accumulate and validate scientific knowledge. As Meehl (1967, 1990) long argued, psychology's reliance on null hypothesis significance testing and the pursuit for universal laws may be fundamentally misaligned with the nature of psychological phenomena. However, recognizing these limitations need not lead to scientific nihilism. Rather, as Mayo (2018) argues, acknowledging the bounds of our knowledge can lead to more careful and nuanced scientific practice. This aligns with what Chambers (2017) terms "humble science" – an approach that explicitly recognizes the provisional nature of scientific claims and the importance of methodological rigor. This perspective requires maintaining both creativity in developing new hypotheses and approaches and rigorous skepticism in evaluating them (Popper, 1962). Each piece of research involves multiple layers of uncertainty – from

measurement error and construct validity through imperfect experimental control to causal misspecification. Our goal as scientists should be to systematically address these uncertainties one by one, while being explicit about those that remain.

It has to be acknowledged that the studies included in this thesis targeted a rather low-hanging fruit, as they primarily addressed what Meehl (1978) termed "technological" questions – focusing on practical matters of measurement, assessment of evidence quality, and predictive relationships rather than development of strong substantive theories. As Meehl outlined, the latter represents a far more demanding challenge in psychology, requiring much more formalized theories, able to generate risky predictions that can survive severe testing (see also van Dongen et al., 2024). By addressing auxiliary issues like measurement validity and empirical robustness of evidence that serves as building blocks for new theories, we hope, however, to have made at least an incremental contribution that may help future researchers devise substantive theories with higher verisimilitude.

The future of gaming research lies in developing richer, more nuanced understanding that explicitly acknowledges both the complexity of its subject matter and the limitations of our tools for studying it (Vazire, 2018; Simmons et al., 2021). This requires balancing the need for practical, actionable knowledge with proper scientific humility. Progress may be better characterized by gradual refinement of understanding rather than dramatic breakthroughs (Devezer et al., 2021), suggesting that improving gaming research requires not just better individual studies, but a fundamental shift in how we conceptualize, pursue, and synthesize scientific knowledge in this field and beyond.

References

Bachen, C. M., Hernández-Ramos, P. F., Raphael, C., & Waldron, A. (2016). How do presence, flow, and character identification affect players' empathy and interest in learning from a serious computer game? Computers in Human Behavior, 64, 77-87.

Baranowski, T., Blumberg, F., Buday, R., DeSmet, A., Fiellin, L. E., Green, C. S., ... & Young, K. (2016). Games for health for children—Current status and needed research. Games for Health Journal, 5(1), 1-12.

Billieux, J., Flayelle, M., Rumpf, H. J., & Stein, D. J. (2019). High involvement versus pathological involvement in video games: A critical commentary. Current Addiction Reports, 6(3), 323-330.

Blanken, T. F., Van Der Zweerde, T., Van Straten, A., Van Someren, E. J. W., Borsboom, D., & Lancee, J. (2019). Introducing Network Intervention Analysis to Investigate Sequential, Symptom-Specific Treatment Effects: A Demonstration in Co-Occurring Insomnia and Depression. Psychotherapy and Psychosomatics, 88(1), 52-54.

Bokulich, A. (2011). How scientific models can explain. Synthese, 180(1), 33-45.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. Wiley.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. Journal of Clinical Psychology, 64(9), 1089-1108.

Borsboom, D. (2017). A network theory of mental disorders. World Psychiatry, 16(1), 5-13.

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. Annual Review of Clinical Psychology, 9, 91-121.

Borsboom, D., Cramer, A. O., & Kalis, A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. Behavioral and Brain Sciences, 42, e2.

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., ... & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. Nature Reviews Methods Primers, 1(1), 1-18.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. Psychological Review, 110(2), 203-219.

Box, J. F. (1978). R. A. Fisher: The life of a scientist. Wiley.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Computers & Education, 94, 178-192.

Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model: Reconsidering the network approach to psychopathology. Psychological Review, 125(4), 606-615.

Bringmann, L. F., Elmer, T., & Epskamp, S. (2023). What do centrality measures measure in psychological networks? Journal of Abnormal Psychology, 128(8), 892-903.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. PLOS ONE, 8(4), e60188.

Burleigh, T. L., Griffiths, M. D., Sumich, A., Stavropoulos, V., & Kuss, D. J. (2019). A systematic review of the co-occurrence of Gaming Disorder and other potentially addictive behaviors. Current Addiction Reports, 6, 383-401.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. Advances in Methods and Practices in Psychological Science, 2(2), 115-144.

Castro-Calvo, J., King, D. L., Stein, D. J., Brand, M., Carmi, L., Chamberlain, S. R., Demetrovics, Z., Fineberg, N. A., Rumpf, H.-J., Yücel, M., Achab, S., Ambekar, A.,

Bahar, N., Blaszczynski, A., Bowden-Jones, H., Carbonell, X., Chan, E. M. L., Ko, C.-H., de Timary, P., ... Billieux, J. (2021). Expert appraisal of criteria for assessing gaming disorder: an international Delphi study. Addiction, 116(9), 2463-2475.
Chambers, C. (2017). The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice. Princeton University Press.
Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. Psychotherapy and Psychosomatics, 88(2), 71-83.
Coombs, C. H. (1964). A theory of data. Wiley.
Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. Behavioral and Brain Sciences, 33(2-3), 137-150.
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52(4), 281-302.
Devezer, B., Navarro, D. J., Vandekerckhove, J., & Ozge Buzbas, E. (2021). The case for formal methodology in scientific reform. Royal Society Open Science, 8(3), 200805.
Dong, G., & Potenza, M. N. (2014). A cognitive-behavioral model of Internet gaming disorder: Theoretical underpinnings and clinical implications. *Journal of Psychiatric Research, 58*, 7-11.
Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias—An updated review. PLoS ONE, 8(7), e66844.
Emerson, G. B., Warme, W. J., Wolf, F. M., Heckman, J. D., Brand, R. A., & Leopold, S. S. (2010). Testing for the presence of positive-outcome bias in peer review: A randomized controlled trial. Archives of Internal Medicine, 170(21), 1934-1939.
Epskamp, S., Borsboom, D., & Fried, E. I. (2018a). Estimating psychological networks and their accuracy: A tutorial paper. Behavior Research Methods, 50(1), 195-212.
Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. Psychometrika, 82(4), 904-927.
Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018b). The Gaussian graphical model in cross-sectional and time-series data. Multivariate Behavioral Research, 53(4), 453-480.
Fanelli, D., Costas, R., & Ioannidis, J. P. A. (2017). Meta-assessment of bias in science. Proceedings of the National Academy of Sciences, 114(14), 3714-3719.
Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. Proceedings of the National Academy of Sciences, 115(27), E6106-E6115.
Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. Science, 345(6203), 1502-1505.
Fried, E. I., & Cramer, A. O. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. Perspectives on Psychological Science, 12(6), 999-1020.
Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. Journal of Affective Disorders, 189, 314-320.
Fried, E. I., Flake, J. K., & Robinaugh, D. J. (2022). Revisiting the theoretical and methodological foundations of depression measurement. Nature Reviews Psychology, 1(6), 358-368.
Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Medicine, 13, 72.

Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. Social Psychiatry and Psychiatric Epidemiology, 52(1), 1-10.
Fried, E. I., van Borkulo, C. D., & Epskamp, S. (2021). On the importance of estimating parameter uncertainty in network psychometrics: A response to Forbes et al. (2019). Multivariate Behavioral Research, 56(2), 243-248.
Granic, I., Lobel, A., & Engels, R. C. (2014). The benefits of playing video games. American Psychologist, 69(1), 66-78.
Green, M. C., & Jenkins, K. M. (2014). Interactive narratives: Processes and outcomes in user-directed stories. Journal of Communication, 64(3), 479-500.
Haslbeck, J., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. Psychological Methods, 27(6), 930-957.
Hayek, F. A. (1974). The pretence of knowledge. Nobel Prize Lecture.
Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. PLoS Biology, 13(3), e1002106.
Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). Cochrane handbook for systematic reviews of interventions (Version 6.0). Cochrane.
Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. The Annals of Statistics, 14(4), 1523-1543.
Hood, S. B. (2013). Psychological Measurement and Methodological Realism. Erkenntnis, 78, 739-761.
Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A. D., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database of Systematic Reviews, MR000006.
Ioannidis, J. P. A. (2005). Why most published research findings are false. PLoS Medicine, 2(8), e124.
Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. Epidemiology, 19(5), 640-648.
Jones, P. J., Ma, R., & McNally, R. J. (2021). Bridge centrality: A network approach to understanding comorbidity. Multivariate Behavioral Research, 56(2), 353-367.
Karhulahti, V. M., Vahlo, J., Martončik, M., Munukka, M., Koskimaa, R., & von Bonsdorff, M. (2023). Ontological diversity in gaming disorder measurement: A nationally representative registered report. Addiction Research & Theory, 31(3), 157-167.
Kendler, K. S. (2012). Levels of explanation in psychiatric and substance use disorders: Implications for the development of an etiologically based nosology. Molecular Psychiatry, 17(1), 11-21.
Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? Psychological Medicine, 41(6), 1143-1150.
Kim, H. S., Shin, S. H., & Kim, S. E. (2016). A comparison of the symptoms of Internet gaming disorder according to DSM-5 and ICD-11. Journal of Behavioral Addictions, 5(4), 537-545.
King, D. L., Delfabbro, P. H., Billieux, J., & Potenza, M. N. (2020). Problematic online gaming and the COVID-19 pandemic. Journal of Behavioral Addictions, 9(2), 184-186.
Király, O., Potenza, M. N., Stein, D. J., King, D. L., Hodgins, D. C., Saunders, J. B., ... & Demetrovics, Z. (2020). Preventing problematic internet use during the COVID-19 pandemic: Consensus guidance. Comprehensive Psychiatry, 100, 152180.

Ko, C.-H., Lin, H.-C., Lin, P.-C., & Yen, J.-Y. (2020). Validity, functional impairment and complications related to Internet gaming disorder in the DSM-5 and gaming disorder in the ICD-11. Australian and New Zealand Journal of Psychiatry, 54(7), 707-718.
Kowert, R., & Quandt, T. (Eds.). (2016). The video game debate: Unravelling the physical, social, and psychological effects of video games. Routledge.
Kuss, D. J., Griffiths, M. D., & Pontes, H. M. (2017). Chaos and confusion in DSM-5 diagnosis of Internet Gaming Disorder: Issues, concerns, and recommendations for clarity in the field. Journal of Behavioral Addictions, 6(2), 103-109.
Lee, S. W. S., & Schwarz, N. (2021). Grounded procedures: A proximate mechanism for the psychology of cleansing and other physical actions. Behavioral and Brain Sciences, 44, e1.
Lemmens, J. S., Valkenburg, P. M., & Peter, J. (2011). Psychosocial causes and consequences of pathological gaming. Computers in Human Behavior, 27(1), 144-152.
Levins, R. (1966). The strategy of model building in population biology. American Scientist, 54(4), 421-431.
MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. Journal of Applied Psychology, 90(4), 710-730.
Mathur, M. B., & VanderWeele, T. J. (2021). Estimating publication bias in meta-analyses of peer-reviewed studies: A meta-meta-analysis across disciplines and journal tiers. Research Synthesis Methods, 12(2), 176-191.
Maul, A. (2013). On the ontology of psychological attributes. Theory & Psychology, 23(6), 752-769.
Mayo, D. G. (2018). Statistical inference as severe testing: How to get beyond the statistics wars. Cambridge University Press.
McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. Perspectives on Psychological Science, 11(5), 730-749.
Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. Philosophy of Science, 34(2), 103-115.
Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46(4), 806-834.
Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. Psychological Reports, 66(1), 195-244.
Morey, R. D. (2013). The consistency test does not—and cannot—deliver what is advertised: A comment on Francis (2013). Journal of Mathematical Psychology, 57(5), 180-183.
Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. Annual Review of Psychology, 69(1), 511-534.
Newzoo. (2024). Global Games Market Report 2024. Newzoo.
Nuijten, M. B., Hartgerink, C. H., van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985-2013). Behavior research methods, 48(4), 1205–1226.
Odenbaugh, J. (2011). True lies: Realism, robustness, and models. Philosophy of Science, 78(5), 1177-1188.
Paulus, F. W., Ohmann, S., von Gontard, A., & Popow, C. (2018). Internet gaming disorder in children and adolescents: A systematic review. Developmental Medicine & Child Neurology, 60(7), 645-659.
Popper, K. R. (1962). Conjectures and refutations: The growth of scientific knowledge. Routledge.

Proulx, T., & Morey, R. D. (2021). Beyond statistical ritual: Theory in psychological science. Perspectives on Psychological Science, 16(4), 671-681.
Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. Review of General Psychology, 14(2), 154-166.
Robinaugh, D. J., Haslbeck, J., Ryan, O., Fried, E. I., & Waldorp, L. J. (2023). Invisible hands and fine carvings: On the complex dynamics of psychological phenomena and challenges for theory construction. Psychological Review, 130(2), 440-465.
Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. Journal of Abnormal Psychology, 125(6), 747-757.
Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). Publication bias in meta-analysis: Prevention, assessment and adjustments. Wiley.
Ruggiero, D. (2015). The effect of a persuasive social impact game on affective learning and attitude. Computers in Human Behavior, 45, 213-221.
Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2022). Why hypothesis testers should spend less time testing hypotheses. Perspectives on Psychological Science, 16(4), 744-755.
Scheffer, M., Bockting, C. L., Borsboom, D., Cools, R., Delecroix, C., Hartmann, J. A., Kendler, K. S., van de Leemput, I., van der Maas, H. L. J., van Nes, E., Mattson, M., McGorry, P. D., & Nelson, B. (2024). A dynamical systems view of psychiatric disorders-theory: A review: A review. JAMA Psychiatry, 81(6), 618–623.
Schmidt, F. L., & Hunter, J. E. (2015). Methods of meta-analysis: Correcting error and bias in research findings (3rd ed.). Sage.
Senn, S. (2012). Misunderstanding publication bias: editors are not blameless after all. F1000Research, 1, 59.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22(11), 1359-1366.
Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-Curve: A key to the file-drawer. Journal of Experimental Psychology: General, 143(2), 534-547.
Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. Royal Society Open Science, 3(9), 160384.
Soekarjo, M., & van Oostendorp, H. (2015). Measuring effectiveness of persuasive games using an informative control condition. International Journal of Serious Games, 2(2), 37-56.
Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. J., ... & Harvey, I. (2010). Dissemination and publication of research findings: an updated review of related biases. Health Technology Assessment, 14(8), 1-193.
Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. Social Psychological and Personality Science, 8(5), 581-591.
Stevens, M. W., Dorstyn, D., Delfabbro, P. H., & King, D. L. (2021). Global prevalence of gaming disorder: A systematic review and meta-analysis. Australian & New Zealand Journal of Psychiatry, 55(6), 553-568.
Trikalinos, T. A., & Ioannidis, J. P. A. (2005). Assessing the evolution of effect sizes over time. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), Publication bias in meta-analysis: Prevention, assessment and adjustments (pp. 241-259). Wiley.
van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. Psychological Methods, 20(3), 293-309.
van Bork, R., Sijtsma, K., Epskamp, S., & Borsboom, D. (2022). Measurement vs. prediction in psychometrics: A tale of two stories. Psychological Methods, 27(6), 1117-1134.

van Dongen, N., van Bork, R., Finnemann, A., Haslbeck, J. M. B., van der Maas, H. L. J., Robinaugh, D. J., de Ron, J., Sprenger, J., & Borsboom, D. (2024). Productive explanation: A framework for evaluating explanations in psychological science. Psychological Review. Advance online publication.

Vassend, O. B. (2020). A verisimilitude framework for inductive inference, with an application to phylogenetics. The British Journal for the Philosophy of Science, 71(4), 1359-1383.

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. Perspectives on Psychological Science, 13(4), 411-417.

Wartberg, L., Kriston, L., & Thomasius, R. (2021). Internet gaming disorder and problematic social media use in a representative sample of German adolescents: Prevalence estimates, comorbid depressive symptoms and related psychosocial aspects. Computers in Human Behavior, 114, 106708.

Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. Advances in Methods and Practices in Psychological Science, 3(1), 94-123.

World Health Organization. (2019). International Classification of Diseases for Mortality and Morbidity Statistics (11th Revision).

Zajac, K., Ginley, M. K., Chang, R., & Petry, N. M. (2017). Treatments for Internet gaming disorder and Internet addiction: A systematic review. Psychology of Addictive Behaviors, 31(8), 979-994.