

CONFERENCE PROCEEDINGS

Game Analysis Perspectives Conference

IT University of Copenhagen, April 20-24 2022



This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No 695528 - MSG - Making Sense of Games)

EXTENDED ABSTRACT

Towards Reproducibility in Game Analysis: Some Reflections on the Study Design and Methodology of Loot Box Prevalence Studies

Leon Y. Xiao*†‡

Center for Computer Games Research, IT University of Copenhagen

†: School of Law, Queen Mary University of London

‡: The Honourable Society of Lincoln's Inn

Introduction

Game studies is interdisciplinary [1] and the implication of which is that a mix of methodological approaches from various disciplines (each with different expectations) are acceptable. Many methods for studying games are available at the disposal of the 'game analyst': ranging from playing the specific games themselves, *e.g.*, close reading [2], to collecting data from other players about their gameplay experience with specific games [3] or all games in general [4], to measuring more abstract aspects of players' gameplay experience, *e.g.*, engagement with violent video games [5]. On this spectrum of available methodologies, the former (from, *inter alia*, literary studies) tend to be more subjective, whilst the latter methods (from the social sciences) tend to be more objective. Those former methodologies, which this abstract focuses on, and which seek to have individual researchers analyse specific games, often involve subjectivity because idiosyncratic researchers bring their own insights (and biases) to the analyses. This inherent subjectivity is conceded by the proponents of those former methodologies who also recommends ways to inject some degree of objectivity back into their analysis, *e.g.*, in relation to close reading [2(p. 275)].

Objectivity allows research to be more easily reproducible. Reproducibility in science is now widely accepted as of utmost importance because the credibility of the research depends entirely on it. Reproducibility is also increasingly important in game studies in relation to game analysis because video games are becoming more deeply scrutinised by society, *e.g.*, in regards to whether gambling-like game design elements known as loot boxes should be regulated [6]. Game studies research often now has wider implications: for example, a game analysis project conducted by Zendle *et al.* (2020) examining the prevalence of loot boxes in video games has been cited by, *inter alia*, the Spanish gambling regulator and is already affecting policymaking [7]. Such studies should be reproducible and adopt open science principles to allow for more transparency, easier accessibility, more active public critique, and enhanced public trust [8].

The reflections herein relate to video game loot box prevalence studies (*i.e.*, how frequently are loot boxes implemented) and mainly stem from my attempted replication of Zendle *et al.* (2020) as presented in Xiao *et al.* (2021a) [9] and Xiao *et al.* (2021b) [10] and also reported in Xiao *et al.* (2022) [11]. I plan to use the same methodology again for Xiao (2022) [12]. I cannot claim that my recommendations below are widely relevant, and some may seem obvious; however, by continually giving more thought to my own study design and seeking ways to improve further, I believe I have incrementally enhanced the quality of each attempted use of the same methodology. This abstract does *not* dismiss subjectivity as not valuable (because it is part of many disciplines that game studies interacts with) and, indeed, true objectivity is perhaps never achievable. Instead, this abstract outlines a number of important considerations for game studies researchers that might help them to more objectively design their future studies to be reproducible where appropriate, such that their research conclusions might be more convincing to wider audiences, *e.g.*, the public and policymakers.

Selecting the game analysis subjects using external sources

One manner by which subjectivity is inserted into game analysis is at the very outset: by choosing to study certain games that the analyser is already familiar with, the analyser has already biased the sample before ‘analysis’ or ‘data collection’ has even begun. Therefore, if objectivity is a goal, then the game subjects that will be studied should be derived through an external source that is not coloured by the researchers’ own selection biases. For example, the researchers might decide to study the most popular games as determined by which games’ Reddit subreddits have the most followers or by which games are grossing the most amount of money. Technically, these selections are still biased and not absolutely representative in the sense that whether less popular games or worse financially performing games would reflect the same results could not be known. However, the research team can set that out as a limitation and argue that what most stakeholders (players, parents, and regulators, *i.e.*, the target audiences of the research conclusions) would be concerned with would be the situation amongst the most popular or highest-grossing games. This shows that achieving absolute objectivity is not necessarily required and is probably impossible: some reasonable and practical concessions to subjectivity can, and should, be made.

Loot box prevalence studies generally based their sample selection on the highest-grossing lists of games obtained from authoritative external sources, *e.g.*, the app stores’ national rankings[e.g., 9,10,13]: this allowed conclusions as to the loot box prevalence rate amongst highest-grossing games on specific platforms (*e.g.*, Apple iPhone) in specific jurisdictions (*e.g.*, the UK or Mainland China) to be drawn. This also allowed for the various studies to be comparable with each other. However, in contrast, one loot box prevalence study conducted in Australia used a sample that was selected by researchers from a variety of online lists[14]: although the researcher admirably shared in detail how they selected their sample; justified their selection; and identified the relevant limitations, this relatively idiosyncratic sample selection of ‘popular games’ meant that it was not possible to conclude what the prevalence rates on specific hardware platforms in Australia were (a previous study having found that the prevalence rates on varying platforms could differ from 36% to 59%[13]). It was also more difficult to compare this particular Australian national sample to other national samples.

Using objective definitions to enhance reproducibility

Another way to improve objectivity is to use specific definitions derived from external sources, rather than to allow the analyser to rely on their own subjective decision-making. If the subject matter being studied is novel and yet undefined, the analyser can create and develop any

definitions themselves, as long as they ensure that the definitions are not changed between analyses of different games (or if a definition has been amended, that they return to any previously analysed games to conduct a reanalysis using the newer definition). This was relevant to loot box prevalence studies because there was, and remains, no complete agreement as to what game mechanic constitutes a ‘loot box.’ In the context of Xiao *et al.* (2021), this meant that the definition for a ‘loot box’ was not left to the analyser’s own judgement and was instead reproduced from a third-party, specifically, the work of Nielsen & Grabarczyk[6], which specifically considered how to define loot boxes. In contrast, Zendle *et al.* (2020) used a less detailed definition that the study itself invented. When Xiao *et al.*’s replication was compared to the original Zendle *et al.* study, there were a number of overlapping games that the two studies disagreed as to whether or not they contained loot boxes. It became evident that simulated casino games in which players can spend real-world money to buy more stakes to continue participating in simulated gambling (whose results are randomised) were not recognised as containing loot boxes by Zendle *et al.*, and this was not disclosed by that study. Zendle *et al.* therefore arguably potentially underestimated the prevalence of loot boxes[11]. It is therefore crucially important that which definition was adopted and any design decisions to include or exclude are clearly explained.

A valuable resource in this context would be a codebook or a coding manual, in which the definitions for various concepts are set out and how various concepts would be identified (*e.g.*, how long the analyser should play the game for and which aspect they need to examine) are recorded. The analyser should always refer to the codebook when experiencing games. This codebook may be treated as a living document and allowed to develop alongside the study (again, as long as the analyser would eventually return to reanalyse any games previously analysed using outdated methodologies using the final, revised methodology). This codebook should be shared alongside the publication resulting from the study to enhance reproducibility and assist in any replication. Therefore, the codebook should be written in such a way that any other person should be able to refer to it and experience the game just as the analyser originally did.

Within the study itself, the game analysis could involve multiple game analysers to assess the ‘reliability’ of the codebook and the game analysis process. A certain percentage of all games studied could be analysed by two (or even more) researchers using the same codebook. The two sets of results can be compared to calculate for inter-rater reliability (*e.g.*, Cohen's kappa coefficient). Achieving a high inter-rater reliability would allow the researchers to be confident that their results were reproducible. Inter-rater reliability could also be provisionally calculated during the codebook development process to check whether certain aspects need to be revised and clarified. For example, certain aspects might not be capable of being reliably assessed and other methods should be considered to answer research questions relating to those aspects.

‘Performing’ a specific player experience

On the point of replicating a certain gameplay experience, as documented by the codebook, it is important to recognise the limits of achievable objectivity and embrace subjectivity in the form of ‘performed subjectivity.’ It is likely not possible to reproduce an identical gameplay experience amongst two analysers if no further instruction is given to them. The codebook would set out how exactly to analyse a game for a particular study. Part of developing that codebook should involve the analyser deciding on a specific player experience that they would like to replicate. In the context of close reading, Bizzocchi and Tanenbaum (2011) refer to getting the game analyser to pretend to be the ‘naïve gameplayer’ (p. 275) or some other

‘performed player stereotype’ (p. 277). ‘Performing’ in this context means experiencing the game whilst pretending to be a specific type of player: for example, when performing an inexperienced player, choosing to view the gameplay tutorial in whole rather than skipping it. Bizzocchi and Tanenbaum (2011) suggest that these specific, potentially diverging performed experiences are subjective, as compared to the distanced game studies scholar who would analyse the game objectively (p. 275). That might be true, but this imagined stereotypical subjectivity is actually more replicable than the so-called ‘objectivity’ of the individual game studies scholar. The codebook or the publication can disclose which imaginary stereotype the analyser performed (such that someone else can also adopt and reproduce that same, imaginary experience when they re-analyse the game), whilst how the individual game studies scholar experienced the game cannot be so described and repeated. When choosing which specific, imagined player stereotype to adopt, the researchers can consider whose experience their intended audience most care about, *e.g.*, inexperienced players. Note that different analysers might not necessarily understand or perform a certain ‘player stereotype’ in exactly the same way, but having some guidance would help to make the analysers’ experiences more similar.

In the context of loot box prevalence studies, the original Zendle *et al.* (2020) study reviewed online videos recorded by other players to determine loot box presence and, if unable to decide, then through personal gameplay. In contrast, Xiao *et al.* (2021a; 2021b) determined this firstly through gameplay and, if unable to decide, then through online resources. This design change with the Xiao *et al.* studies were justified on the basis that gameplay more closely replicated how a new player would encounter the game, which is likely what relevant stakeholders (*e.g.*, potential players, parents, and policymakers) are most concerned about. Additionally, gameplay, as compared to observing other’s gameplay experience, also allowed for more in-depth scrutiny of the games in this context. Once again examining the overlapping games that Zendle *et al.* (2020) and Xiao *et al.* (2021a) both studied, it appears that Xiao *et al.* (2021a) was able to identify a few more hidden loot box implementations with complex purchasing procedures that Zendle *et al.* (2020) might have missed[11].

Publicly Sharing Screenshots and Other Records

Lastly, it must be recognised that video games, particularly online ones, are frequently updated and are therefore not immutable objects that can necessarily be studied in the exact same form at a later date[15]. Recovering older versions of the software to verify the results of a previous study is likely often impossible. Given constraints on the public deposit and sharing of video games (*e.g.*, copyright law and practical concerns, such as file sizes), taking screenshots or video recordings of gameplay and publicly depositing those as data (which copyright law would highly likely permit under fair use/fair dealing provisions) alongside academic publications would be a compromise that can help to ensure that the original game analysis experience would at least be partly preserved so that interested parties can attempt replications or at least identify differences between various versions of the game. Indeed, detailed comparisons between the results of Zendle *et al.* (2020) and Xiao *et al.* (2021a) were possible because Zendle *et al.* openly shared the names of all games it studied.

Final notes

The reflections above originated from the development of one particular research design that was used to answer specific questions. Not all of the above would be applicable to, or need to be adopted by, other studies. However, even just a bit of objectivity might be beneficial to game studies projects by making them more reproducible and authoritative. Indeed, there are

other ways to make one's game analysis more objective and reproducible that this abstract did not discuss. Although potential approaches to 'objective' or 'empirical' game analysis have been suggested herein, it must be reiterated that game studies is a multidisciplinary and interdisciplinary field (likely because game studies remains a nascent field such that most scholars came to game studies from other disciplines, rather than having been trained in game studies *per se*[16]). Therefore, it must be recognised that objectivity and reproducibility is not a goal that every project that involves game analysis necessarily requires. Indeed, research from literary disciplines is inherently subjective, and different interpretations are valuable precisely because they are subjective. Notwithstanding, when game analysis crosses with certain domains, such as regulation and governance, replicability and provability are achievable goals with appropriate study design and are, indeed, required to make convincing policy arguments. Every researcher should consider whether they can provide more information about their underlying study design choices and data, so that others can better understand the context in which the study was conducted and attempt to improve upon it.

References

1. Deterding, S. (2017). The Pyrrhic Victory of Game Studies: Assessing the Past, Present, and Future of Interdisciplinary Game Research. *Games and Culture*, 12(6), 521–543. <https://doi.org/10.1177/1555412016665067>
2. Bizzocchi, J., & Tanenbaum, J. (2011). Well read: Applying close reading techniques to gameplay experiences. In *Well played 3.0: Video games, value and meaning* (pp. 262–290). ETC Press.
3. Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N., & Nickell, E. (2006). From Tree House to Barracks: The Social Life of Guilds in World of Warcraft. *Games and Culture*, 1(4), 338–361. <https://doi.org/10.1177/1555412006292616>
4. Nicklin, L. L., Spicer, S. G., Close, J., Parke, J., Smith, O., Raymen, T., Lloyd, H., & Lloyd, J. (2021). "It's the Attraction of Winning that Draws You in"—A Qualitative Investigation of Reasons and Facilitators for Videogame Loot Box Engagement in UK Gamers. *Journal of Clinical Medicine*, 10 (10), 2103. <https://doi.org/10.3390/jcm10102103>
5. Przybylski, A. K., & Weinstein, N. (2019). Violent Video Game Engagement is Not Associated with Adolescents' Aggressive Behaviour: Evidence From a Registered Report. *Royal Society Open Science*, 6, 171474. <https://doi.org/10.1098/rsos.171474>
6. Nielsen, R. K. L., & Grabarczyk, P. (2019). Are Loot Boxes Gambling? Random Reward Mechanisms in Video Games. *Transactions of the Digital Games Research Association*, 4(3), 171–207. <https://doi.org/10.26503/todigra.v4i3.104>
7. Dirección General de Ordenación del Juego (Directorate General for the Regulation of Gambling). (2021, February 19). *Proceso participativo sobre los mecanismos aleatorios de recompensa (cajas botín)*. <https://www.ordenacionjuego.es/en/Noticia-Cajas-Botin>
8. Chin, J. M., DeHaven, A. C., Heycke, T., Holcombe, A. O., Mellor, D. T., Pickett, J. T., Steltenpohl, C. N., Vazire, S., & Zeiler, K. (2021). Improving the Credibility of Empirical Legal Research: Practical Suggestions for Researchers, Journals and Law Schools. *Law, Technology and Humans*, 3(2), 107–132. <https://doi.org/10.5204/lthj.1875>
9. Xiao, L. Y., Henderson, L. L., & Newall, P. (2021). What are the odds? Lower compliance with Western loot box probability disclosure industry self-regulation than Chinese legal regulation. In *OSF Preprints*. <https://doi.org/10.31219/osf.io/g5wd9>
10. Xiao, L. Y., Henderson, L. L., Yang, Y., & Newall, P. W. S. (2021). Gaming the system: Suboptimal compliance with loot box probability disclosure regulations in China.

- Behavioural Public Policy, Advance Online Publication*, 1–27.
<https://doi.org/10.1017/bpp.2021.23>
11. Xiao, L. Y., Henderson, L. L., & Newall, P. W. S. (2022). Loot boxes are more prevalent in United Kingdom video games than previously considered: Updating Zendle et al. (2020). *Addiction, Advance online publication*. <https://doi.org/10.1111/add.15829>
 12. Xiao, L. Y. (2022). *Belgian Loot Box Prevalence: Effectiveness of a 'Ban'? Stage 1 Version 1.pdf*. <https://osf.io/7kjs9/>
 13. Zendle, D., Meyer, R., Cairns, P., Waters, S., & Ballou, N. (2020). The prevalence of loot boxes in mobile and desktop games. *Addiction, 115*(9), 1768–1772. <https://doi.org/10.1111/add.14973>
 14. Rockloff, M., Russell, A. M., Greer, N. M., Lole, L. R., Hing, N., & Browne, M. (2020). *Loot Boxes: Are they grooming youth for gambling?* Central Queensland University. <https://doi.org/10.25946/5ef151ac1ce6f>
 15. Kaltman, E., Stacey, M., & Wardrip-Fruin, N. (2021). The Game I Mean: Game Reference, Citation and Authoritative Access. *Game Studies, 21*(3). http://gamestudies.org/2103/articles/kaltman_mason_wardripfruin
 16. Aarseth, E. (2001). Computer Game Studies, Year One. *Game Studies, 1*(1). <http://www.gamestudies.org/0101/editorial.html>



This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No 695528 - MSG - Making Sense of Games)