Experiments in games: modding the Zool Redimensioned warning system to support players' skill acquisition and attrition rate

Nemanja Vaci (n.vaci@sheffield.ac.uk)
Department of Psychology, University of Sheffield
Sheffield, UK

Tom Stafford
Department of Psychology, University of Sheffield
Sheffield, UK

Ying Ren
Department of Psychology, University of Sheffield
Sheffield, UK

Jacob Habgood and the SUMO Digital Academy team
SUMO Digital, Sheffield, UK

Abstract

The scientific potential of digital game studies in psychology is limited by the observational nature of the data that they investigate. However, digital environments present us with a perfect opportunity to incorporate experimental paradigms in complex interactive and multivariate worlds where each decision made by participants can be tracked and recorded. In this study, we demonstrate an industry-academic research collaboration that offers a proof-of-the-concept on how minor modifications of the game settings could be used to test psychological research questions. We modify the settings of the Zool platform game, where players allocated to the experimental group are provided with more information when in danger of dying in the game. Results of the study show that manipulation does not influence behaviour in the game, such as achieved score or number of deaths, but it changes the overall player's response of whether they will continue playing the game after the disappointing event of losing all their lives, game over event. In line with previous studies, the additional information provided through the experimental manipulation made death in the game more informative to the players.

Keywords: Games, Experimental manipulation, Skill acquisition, Dropout analysis, Large data

Introduction

The adoption of computer games in research has opened avenues for investigating a range of theoretical questions in cognitive science, from intricacies of skill acquisition (Stafford, Devlin, Sifa, & Drachen, 2017) and factors that support intrinsic motivation (Wang, Khoo, Liu, & Divaharan, 2008) to development of social networks (Paavilainen, Hamari, Stenros, & Kinnunen, 2013). The interactive, immersive, and competitive nature of many of these games combined with a digital footprint of every made decision, provides researchers with an extensive landscape to test psychological theories. Amidst the potential wealth of insights lies a critical challenge – the difficulty of actively manipulating the environment and falsifying the causal claims (Popper, 1963).

The possibility to randomise participants to different experimental conditions manipulated by the researchers enables the falsification of theoretical accounts and hypothetical predictions, thus providing a systematic way to study aspects of human behaviours, cognition, and emotion. Yet, the volume, variety, and velocity of digitally occurring datasets, such as social network discussions or game logs, provide new venues for research in psychology (Vaci, Edelsbrunner, et al., 2019; Vaci, Cocić, Gula, & Bilalić, 2019; Vaci & Bilalić, 2017; Bilalić, Gula, & Vaci, 2021). In previous work, we (2022) argued that the potential of games for understanding psychological theories, specifically human skill acquisition, cannot be met without more experimental studies in digital environments. In this study, we tested the feasibility of including manipulations in a digital environment and whether minor manipulations of contextual information influence players' behaviour, e.g. speed of knowledge and skill acquisition and the likelihood of discontinuing their gameplay.

Creative interrogation of big data logs from first-person shooters, such as Counter-Strike and Destiny, or simple runner games, such as Axon, confirmed and extended well-known findings in cognitive psychology. Stafford and Dewar (2014) analysed datasets consisting of several thousand players and replicated an established lab finding that spacing the practice versus massing it enables better retention of skill and performance, while players with initial higher scores reach higher levels of expertise over time. Network analysis of collaborations showed that players with stronger social relationships have better performance and tend to spend more time playing the game (Pirker, Rattinger, Drachen, & Sifa, 2018). While there seems to be a positive relationship between the fluid abilities of League of Legends players and their rank (see Kokkinakis, Cowling, Drachen, & Wade, 2017, see also; Röhlcke, Bäcklund, Sörman, & Jonsson, 2018 with null results on working memory measures).

Compared to experimental approaches in psychology, digital game records offer power in numbers and replication of findings in real-world settings. But games are also complex multivariate media with which users interact differently and whose behaviour is driven by multiple variables at once. To bridge the methodological gap between the experimental and observational paradigms in digital environments, researchers started developing their games, often making simplifications of complex digital environments (for an in-depth discussion on games as a research paradigm in psychology see Allen et al., 2023). To understand the flexibility of human tool use, Allen, Smith & Tenenbaum (2020) developed the Virtual Tools game. The goal was to place a red object into a green region, while players were asked to choose the best tool that accomplishes the goal. Information on players' performance allowed researchers to validate models of action planning, suggesting that flexibility in physical problem-solving depends on the ability to imagine the effects of actions and on the participant's previous knowledge. Van Opheusden et al. (2023) developed a digital board game, similar to the Tic-Tac-Toe, to investigate players' depth of search when making decisions. Their game reduced the decision complexity that one would see in classical board games, e.g. chess, by reducing the number of decisions that players need to consider before making a move. Yet, this simplification allowed them to test cognitive models of decision planning. They not only obtained responses in the lab setting but also collaborated with a mobile app company and gathered over 1.2 million players online to validate proposed cognitive models.

Another approach to understanding participants' behaviour in multivariate digital environments relies on game modifications or "modding" (Elson & Quandt, 2016). By introducing changes in the environment of already existing games, researchers decided to preserve the complexity of the environment but focused on investigating how game modifications influence gameplay. One example of such work came from Dubey et al.(2018), who used digital environments similar to classical Atari games and manipulated available information provided to the players. They changed the amount of prior knowledge participants can use by masking objects and terrain in the game or reversing the interactions with the elements, such as reversing the ladder functionality. These modifications showcased that players' performance depends on affordance knowledge, but more on the visual representation of objects and consistency of the environment.

In this study, we used the game modification paradigm with the aim of bridging the methodological gap between experimental and observational approaches in digital environments. We collaborated with a game development company partner to include changes in the game environment, record the players' behaviour and test their sensitivity to introduced changes.

At the start of the game, players were randomly allocated to the control or experimental group. The main difference between the groups was a change in the settings menu that enabled or disabled the health warning system. In particular, once on the final health bar, the experimental group (i.e. players with the enabled system) received an additional signal coming from their controller and screen indicating that they needed to be extra careful and avoid death. The introduction of similar warning systems has been discussed in several other domains. In horror games, it was shown that knowledge of upcoming frightening events brought by warning systems amplified the intensity of the scene (Perron, 2004) while warning systems in gambling tasks failed to change the behaviour of players (Monaghan & Blaszczynski, 2009). In our work, besides testing the feasibility of the in-game modification approach, we assumed that the warning system could make the game easier and more enjoyable by freeing up cognitive resources and allowing players to focus on other aspects of the game (Thorpe, Nesbitt, & Eidels, 2019). Therefore, we aimed to test in-game behavioural outcomes, such as the number of deaths and acquired score, as well as general outcomes, such as the likelihood of dropping out (attrition of players), assuming that the experimental group provided with additional information through warning system would outperform control group on all of these measures.

Figure 1: Zool Redimensioned (image retrieved from Games Industry.biz website)

Methods

Game environment

To test the effects of manipulations on players' behaviour, we collaborated with SUMO Digital Academy (https://www.sumo-academy.com/) in Sheffield, United Kingdom. We used the Zool Redimensioned game, a rebuilt version of the original Zool: Ninja of the Nth Dimension built for the Amiga in 1992. The Zool game is a platform game in which you control a gremlin ninja named Zool who must pass seven worlds, each with four levels and beat the boss at the end of each world (see Figure 1).

Procedure and manipulation

At the start of the first game, each player was allocated an encrypted random player ID and a randomised game setting. The manipulation consisted of an enabled (experimental group) or disabled (control group) health warning system. The enabled warning system provided more information to the experimental group, where when on the final health bar, the avatar's health bar represented as a heart icon bounced on the screen and was accompanied by a heartbeat sound and vibration of the PlayStation controller. Contrary to this, the control group did not receive any further information when on their last health bar.

The decision to manipulate the warning system of the game was not grounded in the previous psychological research but was reached jointly with the game developers and focused on minimising disruptive changes in the environment. In particular, we decided against making local changes, such as interactions between objects and enemies in the environment and the avatar or adding warning information to specific aspects of the game. Besides being potentially more disruptive for their gameplay, additional game elements could also be recognised by players and introduce additional confounding in measured behaviour. In addition to manipulating global options through the warning system, the toggle option that switches the warning system on or off was enabled for all players. In other words, players were left with an option to change the warning information and move between the control and experimental groups. Our decision to enable this transition relied on increasing the ecological validity of our manipulation, as we would expect that some players explore game settings and available options that would change their game experience. Not only that, but the switch between the groups also created additional factorial combinations, where we expected to observe players that start their gameplay without warning system information (control group) and enabled it in later levels, actively switching to the experimental group and the other way around. However, due to very few players switching between options in the final dataset, we decided only to analyse responses from non-switching players.

After the initial randomisation of the setting, players entered a specific world and level when data collection started. The data collection was aggregated and recorded after players finished each level. In situations when players quit the game during the level progression, the data collection would stop and reset, while in the case of a game over event, the data was aggregated and recorded up to that moment.

The data collection procedure was online and integrated into the game in the PlayStation environment. Due to the data being automatically collected by the system, it was not possible to guarantee that the same person played under their allocated unique anonymised player ID. Players in-game behaviour was the only information that was collected in this game. This allowed us to test the effect of our manipulation without changing players' game immersion. However, this also meant that we could not measure and test the effect of experimental manipulation on gameplay experience or cognitive workload load, as measured by standard psychological tools (e.g. Flow scales or NASA TLX), or any other psychological construct.

Ethical approval for the data collection conducted in this study was applied for and granted by the University of Sheffield Department of Psychology Ethics Committee, under application number 051044.

Variables

Multiple measures were collected during players' gameplay. We recorded their anonymised ID, world, level, and time when the data was recorded and whether the warning was turned off (control) or on (experimental setting). Information on the level outcomes game included the number of deaths, total seconds spent on each level, the number of times players were left with one health bar, and the total number of health bars they recovered when on final health bar. We also collected global game-level outcomes, such as collection of collectable badges and the overall achieved score. Information about the environment was logged as well, such as how many health pickups were spawned in the environment. Finally, we collected information on players' playstyle, from how many inputs players made when on final health to how much time they spent on the final health bar at each level.

Dataset and cleaning procedure

The collection of the data spanned a period between May 16th and June 6th, which resulted in a dataset of 520 players with 6500 observations. Out of the total number of players, 273 (52%) were allocated to the control group and 247 (48%) to the experimental group at the beginning of their first game. To clean the data, we checked how many players changed the warning setting, switching between the experimental and control group, and only kept the players without warning setting changes in the analysis. Most of our players (484 or 93%) stayed in the initially assigned condition, while the majority of players who changed the setting did it only once (15). We excluded all observations where players spent more than 20 minutes in a level (1200 seconds), assuming that they were inactive in the environment, which resulted in the deletion of 6 observations (0.1% of the data). This resulted in the dataset of 484 players (252 versus 232 with warning turned off or on) and 5592 observations (see Table 1 for descriptive statistics). In the final step of data preparation, we divided the dataset into a developmental set and validation dataset using participants as a level of analysis. The developmental set held 85% (411 players and 4814 observations), while the validation set consisted of 15% (73 participants and 778 observations) of the collected data.

Analysis

In the first step of the analysis, we tested whether additional information provided to the experimental group through the warning system influenced their in-game behaviour. We compared averaged values on three main measures that relate to final health bar behaviour: how many times players entered the final health bar and observed warning from the system, how many health bars did they recover when on the final health bar, and how many inputs per second did they produce when on the final health bar. These measures were averaged for each player and experimental condition and compared using an unpaired two-sample Wilcoxon test.

In the second step of the analysis, we tested the effect of experimental manipulation on the measures of skill acquisition

Table 1: Descriptive statistics for selected variables variables (M and SD)

Group	Deaths	Seconds played	Recovered bars	Entered bars	Inputs	Game over	Score
Control	0.97 (1.48)	164.04 (119)	0.33 (0.74)	1.43 (1.82)	40.26 (65)	263	91212 (120448)
Experimental	1.13 (1.58)	174.64 (129)	0.38 (0.76)	1.64 (1.93)	47.30 (70)	298	87374 (104393)

Note. Descriptive measures were calculated across all participants and played levels. Deaths: average number of deaths; Seconds played: the average amount of time (seconds) spent on each level; Recovered and Entered bars: average number of times players were left with only one health bar and average number of recovered health bars when on final health bar; Inputs: average number of controller inputs when on final health bar; Game over: total number of game over events. Score: achieved score

(number of deaths and score per level) and the likelihood of players discontinuing their play (dropout/attrition rate). In all models, we tested the main effect of experimental manipulation (warning setting turned off or on) and the interaction between experimental manipulation and changes over time (level number). We controlled for the effect of the total number of seconds spent in a level, total collected score, time of play, and total number of inputs per level. In the case of dropout analysis, we included game over the outcome as a predictor, as we expected that this might be one of the main reasons behind players' decisions to quit the game, as well as its interaction with the warning setting. Finally, we included players' unique IDs as random intercepts and changes over levels as random slopes in the model. Given that three outcome measures follow different probability functions, we used three linear regression models to analyse the effect of the warning system. To model the counted variable (number of deaths) which theoretically follows Poisson distribution, we used mixed-effect Poisson regression. The collected score per level was continuous and was modelled with linear mixed-effect regression. Finally, in the case of time-to-event data (dropout rate), we are not only interested in whether players discontinue their engagement but also when they stop playing. To model such events, we used Cox proportional hazards regression.

Independent variables and their interactions, as well as controls, were iteratively added to the model. In the cases when models failed to converge, we used a less complex version of the model, with a reduced number of controls or a simpler random effect structure. We used Akaike's and Bayesian information criterion and likelihood differences to compare the models. Finally, all models were tested on the validation dataset, where we used the Brier score to test the predictions of Poisson models, mean absolute error (MAE) to test the prediction of linear models, and area under the ROC curve (AUC) to test the performance of Cox regression models.

The additional model results with full R code are reported in the supplemental materials (see `https://osf.io/6ptbk/`)

Results

The effect of health warning manipulation on the behaviour in the game

We see no differences in the mean number of times players enter the final health bar (W = 20571, p = 0.65) and how much health they recover when on the final health bar (19897, p = 0.30) between the two conditions. However, players allocated to the experimental group with an activated warning system show a higher number of controller inputs per second when on the final health (M = 2.17, SD=1.07) versus the control group (M = 1.88, SD = 1.00; W = 17073, p $\leq$ 0.01).

The effect of health warning on number of deaths

Results show that the number of deaths in a game is not being impacted by the experimental manipulation (B = 0.13, SE = 0.10, z = 0.12, p = .90), nor does it modify the slope of how the death count changes over the levels (B = -0.005, SE = 0.006, z = -0.86, p = .387). The only significant effect was a level progression, where expectedly players on later levels tend to have a higher number of deaths (B = 0.07, SE = 0.004, z = 14.18, p $\leq$ .01). This model also scored the lowest Brier score (see Death count: model validation in supplemental materials).

The effect of health warning on score acquisition

Results show that the level score is not being impacted by the experimental manipulation (B = 6565, SE = 4905, t = 1.33, p = .18), nor does it modify the slope of how death count changes over the levels (B = -1568, SE = 984 t = -1.593, p = .112). The only significant effect was a level progression, where expectedly players playing later levels tend to score higher scores (B = 6338, SE = 703, t = 9.01, p $\leq$.01).

The effect of health warning on player dropout

In the case of survival analysis, the dropout was defined as players discontinuing their gameplay before reaching level 28 (world 7 and Level 4). The distribution of attrition events is illustrated in Figure 2.

Contrary to previous analyses, we only included the first attempt at playing each level by each player, which left us with 3,543 observations in the developmental dataset. Results show the main effect of the number of inputs and game over predictors (see Table 2). Looking at the hazard ratio, players with a higher number of inputs have a significant even though a very small decreased likelihood of discontinuing their gameplay. Players who experience a game over event have more than 5 times higher likelihood of stopping their engagement with the digital environment.

As the main effect of game over effect was strong, we tested it in interaction with our main experimental manipulation

Table 2: Table 2: Longitudinal Cox regression coefficients

	Estimate	Hazard ratio	Standard error	z statistics	p-value	95% CI (hazard ratio)
$HealthWarning_{exp}$	-0.008	0.99	0.11	-0.06	0.94	0.78 – 1.25
$gameover_{yes}$	1.738	5.68	0.17	9.71	<.01	4.00 – 8.05
$TimeOfPlay$	0.096	1.10	0.20	0.47	0.63	0.73 – 1.64
$Inputs$	-6.4e-04	0.99	3.0e-04	-2.15	<.05	0.99 – 0.99
$SecondsInLevel$	7.8e-04	1.00	8.6e-04	0.91	0.35	0.99 – 1.00
$Score$	-1.0e-06	1.00	5.6e-07	-1.83	<.06	1.00 – 1.00
$HealthWarning_{exp} : gameover_{yes}$	-0.53	0.58	0.24	-2.15	<.05	0.35 – 0.95

Note. The reference level reflects players allocated to the control group who did not experience game over event

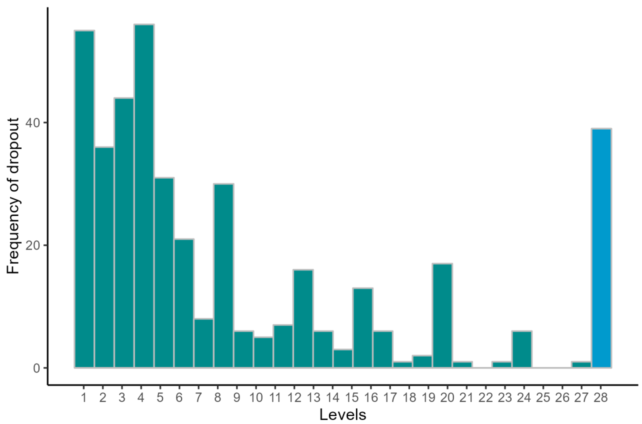

Figure 2: Histogram of players discontinuing their gameplay. Players that reached level 28 were defined as censored events where the event (dropout) did not happen.

(Gelman & Hill, 2006). Results indicate that experimental manipulation indeed moderated the effect of the game over event, where players with warning setting turned on (experimental group) have a smaller hazard ratio (5.68 – 0.58 = 5.10) and a slightly higher likelihood of continuing their play after the game over event (see Figure 3).

In the final step of the analysis, we validated the proportionality of the hazard ratio assumption of the Cox regression and the performance of all developed models on the validation dataset. Our model shows no violations of assumptions (see Cox regression: validation of assumptions in supplemental materials), while the area under the curve value is highest (0.64) for the reported model with interaction between game over event and warning setting (see Cox regression: model validation).

Discussion

Digital games offer unparalleled opportunities to understand cognitive processes, decision-making patterns, and emotional responses during gameplay. This approach not only offers insights into individual gaming habits but also contributes to the broader understanding of social dynamics (Paavilainen et al., 2013), skill development (Stafford & Dewar, 2014), and

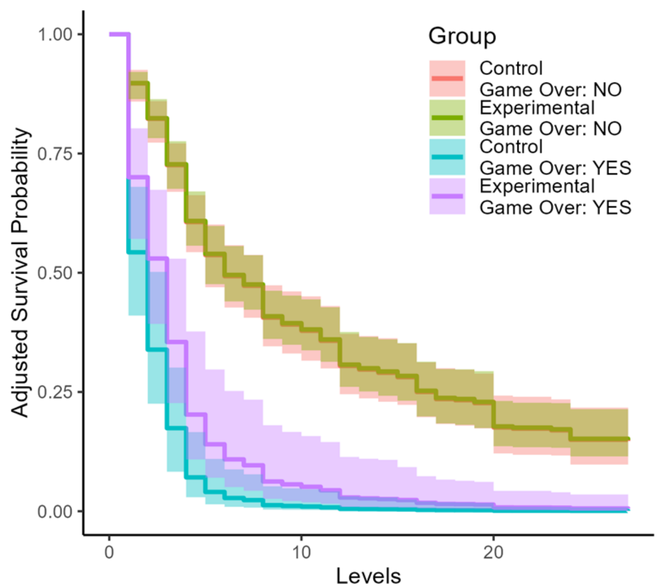

Figure 3: Figure 3: Survival probability for an interaction between warning setting and game over event. Groups illustrated in red (Control group without game over outcome) and green (Experimental group without game over outcome) colour overlap and are least likely to stop their engagement with the game. Groups illustrated in purple (Experimental group with a game over outcome) and blue (Control group with a game over outcome) have a higher likelihood of dropping out from the game.)

the impact of gaming on cognitive and emotional well-being (Pallavicini, Ferrari, & Mantovani, 2018). In this study, we tested one approach that can be used to bridge the methodological gap between experimental paradigms (Elson & Quandt, 2016) and observational data collected by digital environments. Using a modified game setting, we provided additional information during crucial aspects of the game to a subset of our players and tested the effect of this modification on several in-game outcomes, from skill-relevant measures, such as the number of deaths and acquired score, to global game outcomes, such as dropout rate.

By introducing experimental paradigms to the games, we not only support the falsification of psychological theories but also gain the benefits of the big data approaches. In the short span of data collection time, we have managed to collect over 500 players with more than 6000 observations allowing us to isolate the effect of our manipulation on the players' behaviour. Digital games are a complex multivariate environment that can replicate the complexity of our decision-making in everyday life. They are also fun, attention-grabbing, and cognitively rewarding, which all help to keep players motivated and willing to explore the digital environment (Allen et al., 2023). However, this is exactly where challenges were identified in our work. The decision behind which aspect of the game to manipulate was deliberate and through several discussions with our industry partner, we explored different aspects of the game where changes could be introduced. Changing interaction of specific game elements and the main character (avatar) could be recognised by players and introduce additional confounding in outcome measures, while more disruptive changes could impede players' enjoyment of the game. Our final decision focused on changes in the global setting of the game that enabled or disabled the warning system when on the final health bar across all played levels. We assumed that such change allows for greater ecological validity, but it also made data processing more challenging as a subset of players switched between experimental and control groups. When thinking about experiments aimed to test complex cognitive theories, the manipulation of the environment will equally be more complex and would require careful trade-offs between introducing changes that minimise disruptions in the gameplay but allow testing of the hypothetical assumption (Rafferty, Zaharia, & Griffiths, 2014).

The multidimensional aspect of games allows players to optimise different metrics and strategies in their gameplay, posing a data scientist challenge when working with the collected data. Besides being interested in the effect of experimental manipulation, we are left with a decision on which outcome measures are the best proxies of the underlying cognitive process. In the Zool game, we collected dozens of measures that players could optimise, such as level clearing speed or acquired score, collectables, or minimisation of deaths. In the initial plan of the analysis, we expected that the last health bar warning system would likely influence play focused on avoidance of death and free up cognitive resources to focus on other aspects of the game (Thorpe et al., 2019). Yet, this is unlikely a final list of outcomes that players explored and optimised in their gameplay and where the effect of our experimental manipulation could be identified. Analysis of all of these outcomes requires different modelling strategies (Vaci, Cocić, et al., 2019), as evidenced by us using three different regression models, but also careful consideration of controls, covariates, and even engineered measures (e.g. time difference between play attempts, see Stafford & Dewar, 2014).

In this study, we show that players are sensitive to the additional signal introduced by the warning system. They do not change their playstyle by avoiding the last health bar instances or by collecting more health bars but the experimental group observes a higher number of controller inputs per second when on the last health bar. Similarly, results from the skill acquisition measures (number of deaths and achieved score per level) indicated that the warning manipulation did not change the way players optimised their level progression. However, players were less likely to discontinue their exploration of the game once they encountered the disappointing event of losing all their lives and having to repeat the world, i.e. game over event. This interaction between experimental manipulation and game over event on the likelihood of discontinuing the gameplay is small but validated on the held-out dataset and potentially indicates that our manipulation changed how players experience the game over event in the game. In particular, we assume that the warning system made a game over event informative and gave players a sense that they have learned something after experiencing it, thus increasing the chances of players having another try at the game. This interpretation agrees with previous work on game deaths, where studies showed that players tend to feel bored if the in-game death is inconsequential, as they cannot learn from their mistakes (Bartle, 2004; Juul, 2009), while also smiling after challenging deaths (Van den Hoogen, Poels, IJsselsteijn, & De Kort, 2012). The best illustrations of this mechanism are the games that use it as their main learning tool such as Geometry Dash or Dark Souls where players die hundreds of times at each level and still come back to the environment ready to take on the same challenge but now equipped with the skill honed on the previous run.

Through industry partnership and experimental manipulation of an already existing game platform, we illustrated one way of bridging the methodological gap between experimental paradigms and observational studies in digital games. This allowed us to get benefits of both worlds, big data and experimental controls, and to demonstrate the potential of in-game experiments to reveal small effects. This is potentially important for game design where small effects might lead to unbalances in the game or where players rack up 000s of hours and small effects can be very consequential. Large-scale complex experiments also push the frontiers of cognitive science, where our research is often restricted to the domain of large effects because experiments are not powered enough for smaller effects.

References

Allen, K. R., Brändle, F., Botvinick, M., Fan, J., Gershman, S. J., Griffiths, T. L., ... others (2023). Using games to understand the mind.

Allen, K. R., Smith, K. A., & Tenenbaum, J. B. (2020). Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *Proceedings of the National Academy of Sciences*, *117*(47), 29302–29310.

Bartle, R. A. (2004). *Designing virtual worlds*. New Riders.

Bilalić, M., Gula, B., & Vaci, N. (2021). Home advantage mediated (ham) by referee bias and team performance during covid. *Scientific Reports*, *11*(1), 21558.

Dubey, R., Agrawal, P., Pathak, D., Griffiths, T. L., & Efros, A. A. (2018). Investigating human priors for playing video games. *arXiv preprint arXiv:1802.10217*.

Elson, M., & Quandt, T. (2016). Digital games in laboratory experiments: Controlling a complex stimulus through modding. *Psychology of Popular Media Culture*, *5*(1), 52.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Juul, J. (2009). Fear of failing? the many meanings of difficulty in video games. *The video game theory reader*, *2*(237-252).

Kokkinakis, A. V., Cowling, P. I., Drachen, A., & Wade, A. R. (2017). Exploring the relationship between video game expertise and fluid intelligence. *PloS one*, *12*(11), e0186621.

Monaghan, S., & Blaszczynski, A. (2009). Electronic gaming machine warning messages: Information versus self-evaluation. *The Journal of Psychology*, *144*(1), 83–96.

Paavilainen, J., Hamari, J., Stenros, J., & Kinnunen, J. (2013). Social network games: Players' perspectives. *Simulation & Gaming*, *44*(6), 794–820.

Pallavicini, F., Ferrari, A., & Mantovani, F. (2018). Video games for well-being: A systematic review on the application of computer games for cognitive and emotional training in the adult population. *Frontiers in psychology*, *9*, 2127.

Perron, B. (2004). Sign of a threat: The effects of warning systems in survival horror games. In *Cosign 2004 proceedings* (pp. 132–141).

Pirker, J., Rattinger, A., Drachen, A., & Sifa, R. (2018). Analyzing player networks in destiny. *Entertainment Computing*, *25*, 71–83.

Popper, K. R. (1963). Science as falsification. *Conjectures and refutations*, *1*(1963), 33–39.

Rafferty, A. N., Zaharia, M., & Griffiths, T. L. (2014). Optimally designing games for behavioural research. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *470*(2167), 20130828.

Röhlcke, S., Bäcklund, C., Sörman, D. E., & Jonsson, B. (2018). Time on task matters most in video game expertise. *PloS one*, *13*(10), e0206555.

Stafford, T., Devlin, S., Sifa, R., & Drachen, A. (2017). Exploration and skill acquisition in a major online game. In *The 39th annual meeting of the cognitive science society (cogsci)*.

Stafford, T., & Dewar, M. (2014). Tracing the trajectory of skill learning with a very large sample of online game players. *Psychological science*, *25*(2), 511–518.

Stafford, T., & Vaci, N. (2022). Maximizing the potential of digital games for understanding skill acquisition. *Current Directions in Psychological Science*, *31*(1), 49–55.

Thorpe, A., Nesbitt, K., & Eidels, A. (2019). Assessing game interface workload and usability: A cognitive science perspective. In *Proceedings of the australasian computer science week multiconference* (pp. 1–8).

Vaci, N., & Bilalić, M. (2017). Chess databases as a research vehicle in psychology: Modeling large data. *Behavior research methods*, *49*, 1227–1240.

Vaci, N., Cocić, D., Gula, B., & Bilalić, M. (2019). Large data and bayesian modeling—aging curves of nba players. *Behavior research methods*, *51*, 1544–1564.

Vaci, N., Edelsbrunner, P., Stern, E., Neubauer, A., Bilalić, M., & Grabner, R. H. (2019). The joint influence of intelligence and practice on skill development throughout the life span. *Proceedings of the National Academy of Sciences*, *116*(37), 18363–18369.

Van den Hoogen, W., Poels, K., IJsselsteijn, W., & De Kort, Y. (2012). Between challenge and defeat: Repeated player-death and game enjoyment. *Media Psychology*, *15*(4), 443–459.

van Opheusden, B., Kuperwajs, I., Galbiati, G., Bnaya, Z., Li, Y., & Ma, W. J. (2023). Expertise increases planning depth in human gameplay. *Nature*, *618*(7967), 1000–1005.

Wang, C. K. J., Khoo, A., Liu, W. C., & Divaharan, S. (2008). Passion and intrinsic motivation in digital gaming. *Cyberpsychology & behavior*, *11*(1), 39–45.