



OPEN Increased user engagement on YouTube for loot box content and its potential relevance for behavioural addictions

Elke Smith^{1,3}✉, Yannik Poth^{1,3}, Kara S. Z. Sohns¹, Kai Kaspar² & Jan Peters¹

Video games frequently contain loot boxes, i.e. virtual in-game items sharing structural similarities with gambling. On YouTube®, there are multi-million subscriber channels prominently featuring loot box-related content. A gambification of digital games may increase player engagement, and we tested if user engagement on YouTube is linked to loot box content. We extracted aggregate user engagement measures from more than 22 thousand YouTube gaming videos with and without focused display of loot boxes. Principal component analysis was used to reduce dimensionality and derive components reflecting overall and sustained, and relative user engagement, respectively. Confirming our pre-registered hypothesis (see <https://osf.io/nh7zr>), a significant effect of loot box content on the first principal component was found, reflecting higher overall and sustained user engagement for videos featuring loot box content. This increased engagement may be linked to the gambling-like properties of the reward structure conveyed by loot boxes. Publicly available user data may serve as an early indicator of potential changes in problematic internet use and gambling-related behaviour.

Keywords Loot box, User engagement, YouTube, Gaming, Gambling-like mechanisms, Gambling

During the last decade, developers of many video games have implemented a new method to generate revenue in video games. So-called *loot boxes* are digital containers that contain virtual items that are useful to the player. In most video games, loot boxes can be acquired in three (not mutually exclusive) ways: directly using real money, indirectly by purchasing in-game currency, which is then used to buy loot boxes, and through in-game currency earned by playing and completing tasks within the game. The content of loot boxes is not predictable and unknown to the player at the time of purchase^{1,2}. Loot boxes offer players a way to obtain specific game items, such as weapons, special abilities, or cosmetic objects. While cosmetic items usually only serve as a status symbol or expression of one's identity within the video game, functional items such as weapons have a direct influence on the gameplay and can improve a player's position in the game or enable faster progress³.

Loot boxes represent a special form of in-game purchases, and can be distinguished from *direct sales*, in which a specific item is purchased directly. In-game purchases are also referred to as microtransactions, describing the seemingly insignificant purchase amount of such options⁴. In-game purchases constitute an important source of revenue for game developers in the face of stagnating retail prices in the video game segment and increasingly expensive productions⁵. While for some video games, loot boxes generate additional post-sales revenue, many free-to-play video games generate revenue entirely from loot boxes and other microtransactions. In 2020, more than 50% of the top games in the Google Play Store and Apple's App Store games contained loot boxes². The implementation of loot boxes seems to be especially profitable for developers of mobile games since adolescents exposed to loot boxes spend more on mobile video games⁶. Revenue from loot boxes is not evenly distributed among users. Around 5% of buyers generate more than half of the revenue⁷. Some authors suggest that these high-consumption gamers, also known as "whales", are primarily pathological gamblers, pathological video gamers, or other at-risk groups⁸. Individuals who gamble appear to spend more on loot boxes compared to individuals who don't⁹. The appeal and utility of loot boxes, especially those containing rare or legendary items, depend on a player's opportunity to use the items in meaningful ways. This dynamic may create a feedback loop, where acquiring powerful items encourages prolonged gameplay to maximise an item's potential. The

¹Department of Psychology, Biological Psychology, University of Cologne, Cologne, Germany. ²Department of Psychology, Social and Media Psychology, University of Cologne, Cologne, Germany. ³Elke Smith and Yannik Poth contributed equally to this work. ✉email: e.smith@uni-koeln.de

value that players attribute to loot box content, driven by perceived higher power and in-game achievement and progression may reinforce involvement in the game and are likely key motivators for loot box purchases.

Loot boxes are discussed as having gambling-like properties, since they share structural similarities with games of chance, such as randomness, unpredictability of the outcome, and hope for a valuable prize^{10–13}. As in classical gambling formats, the chance-based reward structure of loot boxes represents a form of intermittent reinforcement¹¹, which is thought to be a primary driver of the addictive potential of gambling¹⁴. The specific implementation of loot boxes in video games is also often reminiscent of classic gambling environments, such that rewards and reward anticipation are accompanied by exciting sounds, visual effects, and other dramatisations during opening³. Obtaining items from loot boxes triggers reward-related responses and arousal comparable to those elicited by slot machines¹⁵.

While many manufacturers of video games claim that loot boxes are harmless items with no real value^{16,17}, studies demonstrate links between loot-box purchasing behaviour and gaming addiction, with meta-analyses reporting correlations in the small to medium range^{7,18}. Further, time and money spent on loot boxes are correlated with problematic gambling behaviour and maladaptive gambling-related beliefs¹⁹. The use of loot boxes appears to influence the frequency of video gaming and gambling, and loot boxes pose a risk of developing symptoms of problem gambling by promoting greater involvement in video games²⁰. A recent study reports loot box use to be a strong predictor of initiating gambling behaviour⁹.

On video platforms such as YouTube, gaming channels often publish so-called gameplay videos, i.e. recordings of players playing video games and commenting on the gameplay as they play. Platforms such as YouTube offer consumers ways to engage with and respond to content. For instance, it is possible to rate a video with a *like*, or to comment on a video. The Consumers' Online Brand-Related Activities (COBRA) model provides a framework for describing user engagement, differentiating between the three levels of consumption, contribution and creation, with consumption as minimal level of engagement (e.g. viewing), contribution as medium level (e.g. liking, commenting), and creation as highest engagement level (e.g. publishing content)^{21,22}.

In the last few years, the trend to create videos that focus on loot box openings has emerged. YouTube is one of the most visited websites²³, and there are multi-million subscriber video streams focusing on both gameplay content and on the opening of loot boxes. This again resonates with online gambling content, where a similar trend to watch others gamble has emerged²⁴. There is an online market for people who enjoy watching others gamble, and some gamblers report that watching others gamble inspires one's own gambling²⁵. With regard to video games, buyers of loot boxes report feelings of excitement during openings²⁶, and viewers of gaming videos, possibly identifying themselves with the streamer, might experience some excitement from simply watching loot boxes being opened.

Based on these observations, the present work aimed to assess whether user engagement with gaming videos on YouTube is in part modulated by the display of loot boxes in such videos. (Behavioural) addiction is characterised by repeated and compulsive engagement with rewarding stimuli²⁷, frequently accompanied by impulsive responding^{28,29}. Assuming that loot boxes resemble gambling, they may trigger addiction-like behaviour, manifesting as stronger and faster engagement with such content. Therefore, we predicted that user engagement would be higher for videos that contain loot boxes compared to videos without, and preregistered this prediction (<https://osf.io/nh7zr>). To test this hypothesis, we scraped publicly available video statistic data from selected YouTube gaming channels and quantified user engagement for videos with and without loot boxes. We operationalised user engagement via dimensionality reduction (principal component analysis) across a set of various user engagement measures (see “[User engagement variables](#)” section), and predicted higher component values for videos with compared to without loot box content (see “[Principal components](#)” section).

Methods

Procedure

Selection of gaming channels

For the analysis, English-speaking YouTube gaming channels were selected that featured video games prominently featuring loot boxes as a significant component of gameplay¹¹. The video game titles were gathered through literature reviews and internet searches. In a subsequent step, we collected video statistics data for the videos of the respective channels. Hereto, a fresh instance of the Chromium web browser (version 103.0.5040.1, <https://www.chromium.org/>) was used to ensure that the browser was cookie-free, thereby preventing any pre-existing user behaviour from influencing the YouTube search results. Further, a Virtual Private Network (VPN) tunnel provided by NordVPN (version 6.45.10.0, <https://nordvpn.com/>) was employed to obscure the IP address of the device used for the search. Next, channel searches using specific keywords were conducted on the YouTube website. The term “opening” in the context of video games is commonly used to describe the action of opening a loot box, for instance, “pack opening” in the game FIFA or “mega box opening” in the game Brawl Stars. By inputting the game titles along with the keyword “opening” into the YouTube search bar, relevant search results were retrieved. After sorting the search results by relevance, verified English-language channels (i.e. official channels of creators, artists, companies, or public figures) featuring at least five videos showcasing the opening of loot boxes and five gaming videos devoid of loot box video content were considered for the analysis. Of the 17 channels pre-registered for analysis (see <https://osf.io/nh7zr>), 3 channels were excluded due to insufficient focus on loot box content. These channels did not prominently feature loot box openings and hence were not considered. This resulted in the inclusion of 14 channels (see Table 1).

Data collection

The video statistics data were collected by means of web scraping, a technique used to extract data from the internet and store it for later retrieval or analysis³⁰. The data were downloaded using Google's YouTube Data API with a custom Python script (version 3.10.0, <https://www.python.org/>), and the *Requests* library (version 2.26.0,

Channel no. ^b	Main game	Genre	N ^a			%
			Subscriber	Views	Videos	
1	Brawl Stars	Battle arena	912,000	202,126,570	960	7.81
2	Brawl Stars	Battle arena	1,400,000	444,832,418	710	11.83
3	Brawl Stars	Battle arena	453,000	120,029,407	1727	22.76
4	Brawl Stars	Battle arena	2,570,000	668,131,822	732	5.87
5	Counter-Strike	FP shooter	3,110,000	1,040,472,967	578	19.03
6	Counter-Strike	FP shooter	2,110,000	699,151,593	1791	8.71
7	FIFA	Football	3,670,000	524,840,746	206	33.01
8	Forza	Racing	1,370,000	548,491,472	3899	0.41
9	Forza	Racing	4,870,000	2,139,844,721	3223	0.19
10	NBA 2K	Basketball	403,000	125,285,641	2828	30.09
11	NBA 2K	Basketball	3,120,000	1,037,875,186	1162	14.63
12	Overwatch	FP shooter	349,000	126,262,461	836	1.44
13	World of Tanks	Tank MMO	668,000	417,738,767	1752	0.40
14	World of Tanks	Tank MMO	230,000	117,293,702	2293	0.44
Total			25,235,000	8,212,377,473	22,697	

Table 1. Selected YouTube gaming channels. *FP* first-person, *MMO* massively multiplayer online game.

^aMay 2022. ^bThe channels have been assigned numbers for confidentiality. The names of the channels can be provided upon reasonable request.

<https://pypi.org/project/requests/>). The script collected channel statistics, including the number of channel subscribers, video statistics, including the number of views, likes and comments for videos, and timing of likes for videos of a channel, and comment statistics, including the timing of comments, the number of comment likes, comment replies and reply likes. At the time of data collection (6 days, 5 h and 3 min, from 5th to 20th of May 2022), the channels had an average of 1,802,500 subscribers ($SD = 1,462,884.39$) and 586,598,390.93 video views ($SD = 548,136,071.00$).

Data exclusion

Videos that did not contain information regarding comments or likes (presumably due to the deliberate concealment of such content by channel owners) were excluded. Further, we excluded videos for which the fit of the comment latency model was below $R^2 \leq 0.6$ (see “[Comment latency model](#)” section).

Classification of loot box video content

Videos were categorised as containing substantial loot box content versus minimal or no loot box content. If a video did not contain any action of loot box opening or the act of opening a loot box was incidental and not relevant to the video's content (i.e. the opening of a loot box happened without any focus on it), it was labelled as containing minimal or no loot box content. The classification was done manually, whereby each video was screened using YouTube's progress bar to determine the presence and significance of loot boxes. Following the data collection procedure as described in “[Data collection](#)” section resulted in a total of 22,697 videos, of which 2001 were classified as loot (9%), and 20,696 (91%) as non-loot videos.

Measures

User engagement variables

The primary objective of the study was to examine whether user engagement differed between videos with and without loot box content. User engagement was defined based on various metrics available on YouTube, including video views, likes, comments, comment likes, and comment replies. Since the number of likes, comments and views increases over time and differs between channels, they are not suitable for comparing videos from different time points and channels. Therefore, we calculated two relative measures from the absolute numbers of likes, comments and views per video: the *like frequency* as video likes per 1000 video views, and the *comment frequency* as the number of comments per video in relation to 1000 video views. Further, means of comment likes, comment replies, and reply likes were calculated for each video to provide additional insight into user engagement. The video statistics and the derived user engagement variables for loot and non-loot videos, respectively, are listed in Table 2. Data within each channel were *z*-standardised to facilitate comparisons between channels with varying levels of engagement. We also considered the temporal pattern of user engagement, specifically the frequency of comments over time within the first 7 days after video publication (*comment latencies*). We therefore calculated the interval between the time a video was uploaded and the time a comment was published. We binned the *comment latencies* into 8-h time windows and modelled the temporal evolution of the comment frequencies within the first 21 time bins after video release as exponential decay:

$$y(x) = y_0 - a * e^{(-k*x)} + a$$

	Loot	Non-loot	Difference in % ^a
Views	431,189.48 (942,077.81)	248,866.16 (536,951.45)	53.62
Likes	12,268.65 (23,892.17)	7095.74 (14,226.19)	53.43
Comments	860.32 (1451.25)	647.53 (1164.23)	28.22
Comment likes	2.36 (2.62)	2.80 (2.56)	− 17.05
Comment replies	0.45 (0.32)	0.44 (0.27)	2.23
Reply likes	0.43 (0.37)	0.6 (0.50)	− 33.01
Comment frequency ^b	9.75 (19.50)	5.54 (9.14)	55.07
Like frequency ^b	38.58 (21.47)	34.19 (15.89)	12.07
Latency model y0	11,783.66 (55,482.65)	6955.05 (62,565.52)	51.54
Latency model slope	2.33 (1.69)	1.98 (1.34)	16.24
Latency model asymptote	11,783.66 (55,482.65)	6955.05 (62,565.52)	51.54

Table 2. YouTube video statistics (means and standard deviations) for loot and non-loot videos. ^aDifference for loot minus non-loot videos. ^bCounts per 1000 views.

with intercept y_0 , asymptote a , and decay rate k , using `optimize.curve_fit` of the *SciPy* package for *Python* (version 1.9.0, <https://scipy.org/>). The intercept y_0 reflects how many comments were posted in the first eight hours following the publishing of a video. The decay rate k models drop in the number of comments per eight hours. The asymptote a reflects the frequency of comments per 8 h, to which the function converges to after 7 days. Goodness-of-fit was evaluated based on R^2 .

Principal components

The present data set consisted of many interrelated variables (e.g. a video with more views will tend to have more comments). Principal component analysis (PCA) provides a method to summarise a large set of related variables into a smaller set of uncorrelated principal components, by transforming the data such that the resulting components explain a maximal amount of variance. To reduce redundancy among the user engagement measures, we therefore conducted a PCA, including the variables view count, like count, comment count, comment likes, comment replies, reply likes, comment frequency, like frequency, and the three parameters from the comment latency model (comment latency intercept, comment latency asymptote, and comment latency slope). PCA was conducted using the *scikit-learn* package for *Python* (version 1.1.0, <https://scikit-learn.org/>)³¹. The number of components for statistical analysis was determined by visual inspection of the scree plot (see Fig. S2 of the Supplementary Material) and identifying the maximum position of the elbow bend.

Statistical analysis

To examine the relationship between loot box video content and user engagement while accounting for variability related to specific channels, a linear mixed model (LMM) was computed for each of the two components (see “User engagement variables” section), with loot box video content as fixed factor and channel as random factor. The LMMs were specified as a random-intercepts and random-slopes model for the variable “channel”, to account for baseline differences in user engagement across channels and differences in the magnitude of effects of loot box content across channels. The model was implemented with R (version 4.2.1, <https://www.r-project.org/>) and the R libraries *lme4* (version 1.1-15, <https://cran.r-project.org/web/packages/lme4/>) and *lmerTest* (version 3.1-3, <https://cran.r-project.org/web/packages/lmerTest/>), using Restricted Maximum Likelihood (REML) estimation, with an adjusted significance level of $\alpha_{\text{adjusted}} = 0.025$ (significance level of $\alpha = 0.05$ with Bonferroni correction). Since the Levene test indicated deviations from homogeneity of variances for the first ($W = 106.69, p = 0.00$) and second component ($W = 25.56, p = 0.00$), the data were square-root transformed before entering the LMMs. Visual inspection of the data suggested minor deviations from normality after transformation (see Figs. S2 and S3 of the Supplementary Material).

Results

Comment latency model

We excluded 410 videos (1.8%) due to poor variance explanation by the comment latency model ($R^2 \leq 0.6$, see “Data exclusion” section), leaving 22,287 videos for consideration in further analyses. Goodness-of-fit was good for videos with loot box content (median $R^2 = 0.99, MAD = 0.006$), and without loot box content (median $R^2 = 0.99, MAD = 0.007$). The channels contained between 6 and 827 videos with loot box content, amounting to 1964 videos (9%).

Intercorrelations

The full correlation matrix of user engagement measures is provided in Fig. 1. Moderate positive correlations ($0.4 \leq r < 0.6$) were observed between views and comments, likes and comments, likes and the asymptote of the comment latency model, and between the intercept and decay rate of the comment latency model. Very strong positive correlations ($r \geq 0.8$) were observed between views and likes, and between comments and the asymptote of the comment latency model.

	1	2	3	4	5	6	7	8	9	10	11
1: View count	1.000	0.883	0.660	-0.209	-0.318	0.478	-0.001	0.238	0.020	0.580	-0.182
2: Like count	0.883	1.000	0.755	0.200	-0.130	0.543	-0.092	0.301	0.053	0.679	-0.194
3: Comment count	0.660	0.755	1.000	0.178	0.354	0.288	-0.123	0.159	0.139	0.872	-0.225
4: Like frequency	-0.209	0.200	0.178	1.000	0.495	0.118	-0.174	0.156	0.043	0.174	-0.017
5: Comment frequency	-0.318	-0.130	0.354	0.495	1.000	-0.233	-0.077	-0.136	0.136	0.296	-0.048
6: Comment likes	0.478	0.543	0.288	0.118	-0.233	1.000	0.212	0.435	-0.045	0.248	-0.093
7: Comment replies	-0.001	-0.092	-0.123	-0.174	-0.077	0.212	1.000	-0.024	-0.086	-0.203	0.004
8: Reply likes	0.238	0.301	0.159	0.156	-0.136	0.435	-0.024	1.000	-0.067	0.144	-0.090
9: Comment latency intercept (y0)	0.020	0.053	0.139	0.043	0.136	-0.045	-0.086	-0.067	1.000	0.127	0.685
10: Comment latency asymptote (a)	0.580	0.679	0.872	0.174	0.296	0.248	-0.203	0.144	0.127	1.000	-0.158
11: Comment latency slope (k)	-0.182	-0.194	-0.225	-0.017	-0.048	-0.093	0.004	-0.090	0.685	-0.158	1.000

Fig. 1. Correlation matrix for the user engagement variables. y0, a and k are parameters of an exponential decay model for the comment latencies.

	Loot			Non-loot		
	M	MD	SD	M	MD	SD
PC1	0.0872	-0.0281	2.1885	-0.0084	-0.0571	1.8860
PC2	-0.0152	-0.0399	1.5931	-0.0152	-0.1768	1.3926

Table 3. Component values for the first principal components. The data were square-root transformed and within-channel z-standardised. PC1 reflects overall (active and passive) and ongoing engagement, PC2 reflects relative engagement. PC principal component, M mean, MD median, SD standard deviation.

Principal component analysis

A PCA was performed on the z-standardised data. Based on inspection of the scree plot (see Fig. S1 of the Supplementary Material), we considered the first two components for further analysis. The component values for loot and non-loot videos are listed in Table 3, the distributions are depicted in Fig. 3. Component loadings for all principal components are provided in Fig. 2. The first component (PC1) exhibited moderate to strong positive loadings for video views, likes and comments, and asymptote of the comment latency model. The component reflects, in absolute figures, how often a video is viewed, liked and commented on, and how long the video is interacted after its publication. We therefore term PC1 overall (active and passive) and ongoing engagement. The second component (PC2) was primarily constituted by like frequency and comment frequency, reflecting how many likes and comments a video received in relation its views. We therefore term PC2 relative engagement.

Linear mixed models (LMMs)

Results from the LMMs revealed a positive and significant effect of loot box video content for the first principal component (PC1, $\beta=0.873$, $SE=0.79$, $p=0.008$, see Table 4). This implies that, on average, videos featuring loot boxes exhibited higher overall and ongoing engagement than those without loot boxes, thus confirming the predicted engagement effect of loot box content. Notably, the size of the effect is rather small (see Fig. 3). However, looking at the raw video statistics, differences between loot and non-loot videos also manifest descriptively for all but two measures. For example, videos with loot box content had 54% more (absolute) views, 53% more (absolute) likes and 12% more likes per 1000 video views (like frequency), and 55% more comments per 1000 video views (comment frequency) compared to videos without loot box content (see Table 2). With regard to the second principal component, no significant effect of loot box content was observed ($\beta=0.167$, $SE=0.256$, $p=0.529$, see Table 4).

Discussion

Building on the assumption that loot boxes resemble gambling, we tested the preregistered prediction that user engagement on YouTube gaming channels would be increased by loot box content in videos. To this end, we used web scraping to collect video statistics reflecting user engagement such as the number of views, likes and comments for videos posted in selected YouTube gaming channels. We condensed the measures using principal component analysis, and tested for differences in user engagement components for videos with and without loot box content, respectively.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
View count	0.443	-0.232	0.167	-0.228	0.006	-0.067	0.207	0.079	0.383	0.378	-0.572
Like count	0.490	-0.058	0.077	-0.011	0.046	-0.282	0.310	0.079	0.147	0.075	0.734
Comment count	0.458	0.226	-0.036	-0.179	-0.154	0.165	-0.021	0.025	0.109	-0.787	-0.149
Like frequency	0.094	0.409	-0.230	0.532	0.060	-0.491	0.355	-0.016	-0.150	-0.026	-0.306
Comment frequency	0.017	0.582	-0.267	0.072	-0.308	0.248	-0.216	0.078	0.495	0.338	0.126
Comment likes	0.305	-0.270	0.096	0.448	-0.137	-0.266	-0.730	0.027	0.029	-0.007	-0.021
Comment replies	-0.056	-0.270	0.049	0.187	-0.869	0.105	0.315	-0.038	-0.138	0.007	0.006
Reply likes	0.204	-0.169	-0.008	0.573	0.311	0.684	0.191	0.007	0.020	0.019	0.015
Comment latency intercept (y0)	0.014	0.336	0.632	0.037	-0.042	0.082	-0.023	0.624	-0.283	0.079	-0.024
Comment latency asymptote (a)	0.431	0.242	-0.019	-0.197	-0.055	0.169	-0.128	-0.453	-0.600	0.320	-0.011
Comment latency slope (k)	-0.145	0.200	0.654	0.153	0.008	-0.046	0.050	-0.619	0.302	-0.086	0.021

Fig. 2. Eigenvector matrix of the user engagement variable loadings onto the principal components (PC). y0, a and k are parameters of an exponential decay model for the comment latencies.

	Estimate (β)	SD	z value	p value	Random effects		R ²	
					Variance	SD	Marginal	Conditional
PC 1							0.016	0.038
Intercept	−0.020	0.034	−0.589	0.567	0.012	0.113		
Loot box video content	0.873	0.279	3.127	0.008*	0.970	0.985		
PC 2							0.001	0.036
Intercept	−0.015	0.015	−0.999	0.358	0.001	0.036		
Loot box video content	0.167	0.256	0.652	0.529	0.839	0.916		

Table 4. Linear mixed models (LMMs) for the first and second principal component. The data were square-root transformed and within-channel z-standardised. PC1 reflects overall (active and passive) and ongoing engagement, PC2 reflects relative engagement. PC principal component. *p < 0.025, one-sided.

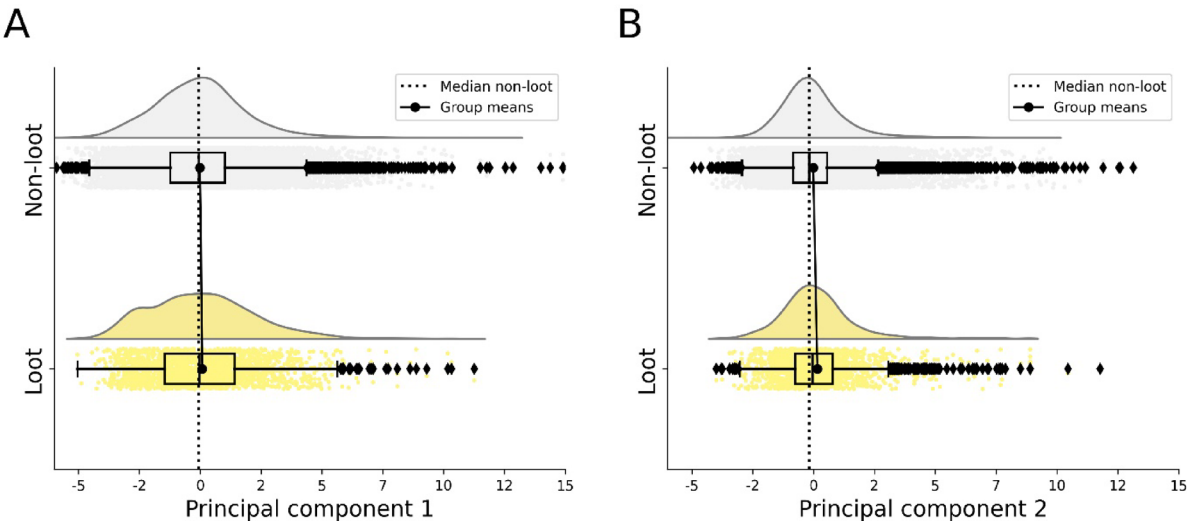


Fig. 3. Distribution (cloud), scores (rain) and central tendency (box plot) of the eigenvalues for the first two principal components (panel A and B) for videos with and without loot boxes in grey and yellow, respectively. PC1 reflects overall and ongoing engagement, PC2 reflects relative engagement with videos.

As hypothesised, we found significant differences in component values for the first component, labelled as overall and ongoing engagement. For the second component, we found no evidence for differences in component values between videos with and without loot box content, respectively. The variables number of views, likes, comments and the asymptote of the comment latency model loaded most prominently on the first principal component. This component therefore reflects general quantitative user engagement, both passive (viewing), and active (commenting), as well as aspects of long-term interaction. A higher asymptote of the comment latency model for videos with compared to without loot box content indicates ongoing consumer engagement over longer time periods after a video is published.

For some games, players create videos of themselves opening large numbers of loot boxes (or card packs, depending on the game) shortly after a publisher has released new game content, e.g., an expansion pack. Some of this new content may have a significant impact on gameplay strategies, often altering or breaking the current most effective and dominant strategies in the game. Part of the enjoyment for viewers may come from watching whether the video creator acquires the most significant or impactful assets. Thus, interest in loot box opening should peak just after the release of new content, and should gradually decline as the expansion ages. Following this, the popularity and engagement with gameplay videos should be more consistent across the lifetime of an expansion when comparing it to the game's overall interest timeline, which may include downtimes between expansions.

Not all forms of actions on social media reflect the same level of engagement. The COBRA framework differentiates between the three levels of consumption, contribution and creation, consumption reflecting minimal level of engagement (e.g. viewing), contribution reflecting medium level (e.g. liking, commenting), and creation highest level of engagement (e.g. publishing content)^{21,22}. Like and comment counts load most strongly onto the first principal component, thus reflecting stronger medium-level engagement for videos with compared to without loot box content. Overall, the results suggest that passive, active and ongoing forms of user engagement are higher for videos with compared to without loot box content, confirming our hypothesis that loot box content is associated with elevated user engagement.

One possibility is that higher user engagement for loot compared to non-loot box content is linked to the gambling-like properties of loot boxes. Loot boxes are designed to keep players involved and investing money in a game and promote greater involvement in video games²⁰. Thereby, potentially harmful content is incorporated into games in order to generate revenue. Critically, loot boxes are often found in games considered suitable for children, both in the mobile and PC/console games sector^{2,11}. Our data show that greater involvement is not only found for games with loot boxes, but also when consuming gaming related video content on YouTube. We don't know how old the users are, but we assume that many of them are teenagers³². This is a vulnerable phase for mental disorders, which is why loot content may be particularly problematic here.

The "gateway hypothesis" posits that video gaming and loot boxes may serve as a gateway activity into gambling and problem gambling^{33–35}. While gaming and social media use may not be problematic per se^{36,37}, the exposure to videos featuring loot box openings may increase the frequency of gaming and/or gambling. In a longitudinal study of people aged 18–26, spending and risky loot box use predicted the start of gambling at a later point in time⁹. Our data do not provide any information on this, but this should be further investigated in future studies. Brooks and Clark's⁹ analyses suggest that the relationship is driven by the exposure to randomised rewards rather than by performing microtransactions per se.

Some gamblers often report that not only gambling itself gives them pleasure, but watching other people gamble does also³⁸. Receiving items from loot boxes in video games may trigger arousal¹⁵, and along similar lines, this may also apply to observing other people who receive items from loot boxes. Our analysis demonstrates that the gambification of video games through loot boxes may drive increased user engagement with YouTube gaming videos. In light of the evidence for a positive relationship between engagement with loot boxes and problematic gaming and gambling³⁹, the current analyses underline the need for recognising and analysing these links using also digital behaviour. While internet platforms such as YouTube collect massive amounts of user data to generate revenue from knowledge about correlations between content and user engagement with videos, these data contain information potentially relevant for research on problematic internet use and gambling behaviour. While the relationship between gambling-related content and social media engagement may be relatively straightforward as users engage with content directly related to gambling, using such data for research on problematic internet use requires more targeted approaches due to the broader nature of problematic internet use. Here, user-level data would be more informative in identifying problematic usage patterns. Still, social media activity may offer insights into problematic usage patterns on a broader scale by tracking excessive time spent on platforms or excessive engagement with specific content (e.g. by looking at posting frequency and timing patterns). Future research could benefit from combining platform-level data with user-level data.

Limitations, strengths and perspectives

The current approach also has some limitations. First, it should be noted that the definition of views is not entirely transparent. YouTube does not disclose what duration of video viewing is counted as a view, or whether users are counted multiple times when watching a video repeatedly⁴⁰. To reduce the influence of age of the videos, we also calculated relative user engagement measures. However, more likes will result in better findability of the video determined by the YouTube algorithm (which is not precisely known). Furthermore, more comments give more reasons for further comments. Therefore, ratios such as likes or comments per 1000 views are also subject to an influence of time.

The results of our work are to be viewed as specific for the games included in the analysis for several reasons. First, the selected games may have a more compelling display of loot boxes and reward structures than other games. Also, it may be the case that the gameplay videos of the games included are less compelling compared to loot box openings, or less compelling compared to games not included. Further, the ratio of games to YouTube

content creators may have influenced the results. For some games, there may be a greater concentration in terms of the number of content creators that hold a certain percentage of viewership or hours viewed, while for other games, there may be a larger number of content creators that divide the fan base more evenly. This affects the popularity of certain games and content creators on YouTube, and in turn, the selection of games and channels for the current work. The results are therefore to be viewed as specific for the games included in the analysis and several prominent games could have been included (e.g. Genshin Impact and Clash Royale) but were simply overlooked during study preparation. Future research should explore a wider variety of games and genres.

A general shortcoming of the present approach (and similar web-scraping-based approaches in general) is that we cannot differentiate which people interact with the videos, and that the mechanisms underlying the difference in engagement for loot and non-loot content remain unclear. Coming from an observational study, the results do not allow any conclusions to be drawn about causality between loot box content and user engagement on the gaming channels. While increased engagement may reflect the gambling-like properties of loot boxes, it is also conceivable that higher engagement reflects enjoyment from watching other's reacting to loot content. Loot box openings may generate tension and excitement and the mechanism driving higher engagement may, in part, resemble that of watching so-called reaction videos, for instance a desire to observe authentic reactions of others⁴¹. Nonetheless, regardless of the motivation to engage with loot box content, the present results highlight that loot box content is linked to higher engagement, which may drive exposure to and normalisation of gambling-like content in the context of gaming.

Data collection using web scraping offers many advantages for psychological research. It enables extracting large amounts of data from the internet in a comparatively short time, and data collection has no influence on the data itself. An advantage compared to survey-based research is the high ecological validity, since behaviour can be studied in a real environment. The current study demonstrates that such data contain information meaningful for behavioural addiction research. Data analysis should not only be left to the providers of these services, who adjust the content to maximise revenue, which may have negative effects for users. Finding higher user engagement for loot compared to non-loot box content may reflect the gambling-like properties of loot boxes. Building upon this, warning notices could be designed more effectively, since consumers perceive harm resulting from engaging with loot boxes lower than for many gambling activities⁴². Publicly available user data may serve as an early indicator of potential changes in problematic internet use and gambling behaviour. For instance, it would be valuable to identify individual factors associated with higher user engagement on gaming channels, since previous studies showed a particularly high increase in video game activities during the COVID-19 lockdown among male subjects⁴³.

Legal statement

The study was an observational study in which publicly available social media data were collected by means of web scraping. In Germany, web scraping for independent scientific purposes is generally legal, and no consent is required under the following conditions: The information to be evaluated must be generally accessible; technical measures designed to prevent web scraping may not be disregarded; the research may only serve non-commercial purposes; the use of web scraping technologies must not cause technical damage to the operator of the website^{44,45}. The study's approach aligned with these regulations. The application programming interface (API) and the respective key used for data collection granted access to publicly available data only. We did not disregard any technical measures designed to prevent web scraping and the procedure did not cause any damage to the website operator. Our research serves non-commercial purposes only. We did not collect data on individual user profiles and the data do not allow the identification of natural persons.

Data availability

The datasets generated and analysed during the current study are not publicly available, since in compliance with German laws, sharing data collected by means of web scraping is not permitted. Upon request, information on the data to reproduce the analyses will be shared from the corresponding author on reasonable request. The code for analysis is openly available at <https://osf.io/bjndc/>.

Received: 17 July 2024; Accepted: 6 May 2025

Published online: 15 May 2025

References

1. Rockloff, M. et al. *Loot Boxes: Are They Grooming Youth for Gambling* Vol. 10 (Central Queensland University, 2020).
2. Zendle, D., Meyer, R., Cairns, P., Waters, S. & Ballou, N. The prevalence of loot boxes in mobile and desktop games. *Addiction* **115**, 1768–1772 (2020).
3. Cerulli-Harms, A., Münsch, M., Thorun, C., Michaelsen, F. & Hausemer, P. Loot boxes in online games and their effect on consumers, in particular young consumers. In *Publication for the Committee on the Internal Market and Consumer Protection (IMCO)*, vol. 202 (Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2020).
4. Muhshoff, M. *Mikrotransaktionen und Videospiele* (2019).
5. Agarwal, P. *Microtransactions in Video Games*. Intelligent Economist. intelligenteconomist.com.
6. Shibuya, A., Teramoto, M., Shoun, A. & Akiyama, K. Long-term effects of in-game purchases and event game mechanics on young mobile social game players in Japan. *Simul. Gaming* **50**, 76–92 (2019).
7. Close, J. et al. Secondary analysis of loot box data: Are high-spending “whales” wealthy gamers or problem gamblers?. *Addict. Behav.* **117**, 106851 (2021).
8. Zendle, D., Meyer, R. & Over, H. Adolescents and loot boxes: Links with problem gambling and motivations for purchase. *R. Soc. Open Sci.* **6**, 190049 (2019).
9. Brooks, G. A. & Clark, L. The gamblers of the future? Migration from loot boxes to gambling in a longitudinal study of young adults. *Comput. Hum. Behav.* **141**, 107605 (2023).

10. Cruz, E. K. The psychological and virtual siege of loot boxes. *J. Technol. Law Policy* **23**, 215 (2018).
11. Drummond, A. & Sauer, J. D. Video game loot boxes are psychologically akin to gambling. *Nat. Hum. Behav.* **2**, 530–532 (2018).
12. Griffiths, M. D. Is the buying of loot boxes in video games a form of gambling or gaming? *Gaming Law Rev.* **22**, 52–54 (2018).
13. Drummond, A., Sauer, J. D., Hall, L. C., Zendle, D. & Loudon, M. R. Why loot boxes could be regulated as gambling. *Nat. Hum. Behav.* **4**, 986–988 (2020).
14. Zack, M., George, R. S. & Clark, L. Dopaminergic signaling of uncertainty and the aetiology of gambling addiction. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **99**, 109853 (2020).
15. Larche, C. J., Chini, K., Lee, C., Dixon, M. J. & Fernandes, M. Rare loot box rewards trigger larger arousal and reward responses, and greater urge to open more loot boxes. *J. Gambl. Stud.* **37**, 141–163 (2021).
16. BBC. *EA Games: Loot Boxes Aren't Gambling, They're Just Like a Kinder Egg*. bbc.com.
17. Environment and Communications References Committee. *Gaming Micro-transactions for Chance-Based Items* (2018).
18. Garea, S. S., Drummond, A., Sauer, J. D., Hall, L. C. & Williams, M. N. Meta-analysis of the relationship between problem gambling, excessive gaming and loot box spending. *Int. Gambl. Stud.* **21**, 460–479 (2021).
19. Brooks, G. A. & Clark, L. Associations between loot box use, problematic gaming and gambling, and gambling-related cognitions. *Addict. Behav.* **96**, 26–34 (2019).
20. Primi, C., Sanson, F., Vecchiato, M., Serra, E. & Donati, M. A. Loot boxes use, video gaming, and gambling in adolescents: Results from a path analysis before and during COVID-19-pandemic-related lockdown in Italy. *Front. Psychol.* **13**, 1009129 (2022).
21. Schivinski, B., Christodoulides, G. & Dabrowski, D. Measuring consumers' engagement with brand-related social-media content: Development and validation of a scale that identifies levels of social-media engagement with brands. *J. Advert. Res.* **56**, 64–80 (2016).
22. Trunfio, M. & Rossi, S. Conceptualising and measuring social media engagement: A systematic literature review. *Ital. J. Mark.* **2021**, 267–292 (2021).
23. DataReportal, H. Most popular websites worldwide as of June 2021, by total visits (in billions). Graph. *Statista* **13** (2022).
24. Greer, K. N., Jenkinson, R., Boyle, C. M., Jatkar, U., Vandenberg, B. & Sakata. *Consumer Views on Emerging Trends and Future Directions for Gambling in Australia* (2022).
25. Vinberg, M. *Between the Lines: Gambling Among Athletes and Coaches in Four Elite Sports* (Karolinska Institutet, 2020).
26. McGrody, E. *For Many Players, Lootboxes Are a Crisis That's Already Here*. vice.com (2018).
27. Grant, J. E., Potenza, M. N., Weinstein, A. & Gorelick, D. A. Introduction to behavioral addictions. *Am. J. Drug Alcohol Abuse* **36**, 233–241 (2010).
28. Mitchell, J. M., Fields, H. L., D'Esposito, M. & Boettiger, C. A. Impulsive responding in alcoholics. *Alcohol. Clin. Exp. Res.* **29**, 2158–2169 (2005).
29. Lee, R. S., Hoppenbrouwers, S. & Franken, I. A systematic meta-review of impulsivity and compulsivity in addictive behaviors. *Neuropsychol. Rev.* **29**, 14–26 (2019).
30. Zhao, B. Web scraping. In *Encyclopedia of Big Data*, vol. 1 (2017).
31. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
32. Vogels, E. A., Gelles-Watnick, R. & Massarat, N. *Teens, Social Media and Technology 2022* (2022).
33. Delfabbro, P. & King, D. L. Gaming-gambling convergence: Evaluating evidence for the 'gateway' hypothesis. *Int. Gambl. Stud.* **20**, 380–392 (2020).
34. Hayer, T., Kalke, J., Meyer, G. & Brosowski, T. Do simulated gambling activities predict gambling with real money during adolescence? Empirical findings from a longitudinal study. *J. Gambl. Stud.* **34**, 929–947 (2018).
35. Spicer, S. G. et al. Loot boxes and problem gambling: Investigating the "gateway hypothesis". *Addict. Behav.* **131**, 107327 (2022).
36. Boer, M. et al. Adolescents' intense and problematic social media use and their well-being in 29 countries. *J. Adolesc. Health* **66**, S89–S99 (2020).
37. Pujol, J. et al. Video gaming in school children: How much is enough?. *Ann. Neurol.* **80**, 424–433 (2016).
38. TheTruth221. *Do U Enjoy Watching Others Gamble?* (2021).
39. Montiel, I., Basterra-González, A., Machimbarrena, J. M., Ortega-Barón, J. & González-Cabrera, J. Loot box engagement: A scoping review of primary studies on prevalence and association with problematic gaming and gambling. *PLoS ONE* **17**, e0263177 (2022).
40. Liikkanen, L. A. *Three Metrics for Measuring User Engagement with Online Media and a YouTube Case Study*. arXiv preprint <https://arxiv.org/abs/1312.5547> (2013).
41. Cox, C. *Not Acting, Reacting: An Analysis of Reaction Videos, Reactors, Viewers and Authenticity* (Brigham Young University, 2023).
42. Garrett, E. P., Drummond, A., Lowe-Calverley, E. & Sauer, J. D. Current loot box warnings are ineffective for informing consumers. *Comput. Hum. Behav.* **139**, 107534 (2023).
43. Lemenager, T. et al. COVID-19 lockdown restrictions and online media consumption in Germany. *Int. J. Environ. Res. Public Health* **18**, 14 (2021).
44. Klawonn, T. Urheberrechtliche grenzen des web scrapings (web scraping under German copyright law). Available at SSRN 3491192 (2019).
45. Rat für Sozial- und Wirtschaftsdaten. *Big Data in Den Sozial-, Verhaltens-Und Wirtschaftswissenschaften: Datenzugang Und Forschungsdatenmanagement. Mit Gutachten "Web Scraping in Der Unabhängigen Wissenschaftlichen Forschung"* (2019).

Author contributions

Writing—original draft: E. S., Y. P. Writing—review and editing: E. S., Y. P., K. S. Z. S., K. K., J. P. Methodology: E. S., Y. P., K. S. Z. S., J. P. Software: E. S., Y. P. Formal analysis: E. S., Y. P. Visualization: E. S., Y. P. Project administration: E. S., J. P. Supervision: E. S., K. K., J. P. Conceptualization: E. S., Y. P., J. P.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-01482-5>.

Correspondence and requests for materials should be addressed to E.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025