Open Play: A longitudinal dataset of multi-platform video game digital trace data and psychological measures

Nick Ballou[1][*], Tamás Andrei Földes[1], Matti Vuorre[2], Thomas Hakman[1], Kristoffer Magnusson[3,1] & Andrew K Przybylski[1][*]

[1]University of Oxford, Oxford Internet Institute
[2]Tilburg University, Department of Social Psychology
[3]Karolinska Institute

A major limitation to understanding digital technology use, and its potential psychological consequences, is the lack of sufficiently detailed, multidimensional, and accurate data. We present a dataset of 2.0K individuals' video game play telemetry data from Nintendo Switch, Steam, and Xbox, paired with psychological measures across multiple dimensions of mental health, motivations, well-being, and cognitive ability. The data were collected under a preregistered design that included 12 weeks of survey data (thirty daily surveys, six biweekly surveys, three biweekly cognitive tests), and digital trace data for 43 months. Cleaned data include 1.5M hours of video game play across 10,475 titles, 23K responses to 14 survey instruments, and 2.9K attention ability measures to facilitate examining longitudinal associations between play behaviors and psychological functioning. Data and codebook are available under a CC0 (with supplemental reidentification clause) license at https://doi.org/10.5281/zenodo.17536656.

Introduction

Digital trace data—behavioral logs automatically collected by digital devices and online platforms—are necessary to better understand technology use and its psychological and health effects (Burgess et al., 2024; Freelon, 2014; Griffioen et al., 2020). Self-reports of technology use do not accurately reflect objective technology use (Parry et al., 2021; Sewall & Parry, 2021), and are unsuitable for examining many phenomena of interest, such as seasonal patterns over long temporal horizons; high-frequency behavioral analysis at the level of hours, minutes, or even seconds; historical content analysis; and others due to limited accuracy and temporal resolution. Moreover, when technology use's relations to psychological survey instruments are of interest, digital trace data removes the possibility of common methods bias.

To combat these issues, digital trace data is increasingly used in studies on the psychological effects of smartphone and social media (e.g. Sewall et al., 2022; Siebers et al., 2024; Yap et al., 2024). But despite the rapid growth of research using trace data from social media and smartphones, comparable efforts in video games remain rare. This absence is notable given that games constitute one of the most popular and psychologically rich forms of digital media, engaging billions of players worldwide across diverse genres, platforms, and social contexts (Entertainment Software Association, 2024; Kaye, 2019; Ofcom, 2023). As a result, our understanding of video game play is often limited to narrow slices of behavior—single titles, single platforms, or small samples–leaving open questions about how gaming affects and interacts with people's everyday lives.

Existing use of trace data in video games typically takes the form of either game analytics research, often conducted in collaboration with industry at the level of an individual game, typically focused on industry-relevant behavior rather than harms and benefits; or narrow-scope investigations into play and health within a single game or gaming platform (e.g., Xbox). Each of these approaches has limitations that constrain their utility for understanding gaming behavior in the wild.

Game analytics research has made substantial contributions to understanding player behavior, engagement, and game design processes (Elson et al., 2014; Sifa et al., 2021). However, these studies are often conducted within commercial partners using proprietary data that are not publicly shared (Kahn et al., 2014; M. Liu et al., 2024), which limits transparency, reproducibility, and opportunities for independent reanalysis. Moreover, because their objectives often center on optimizing player experience, design, or monetization rather than examining psychological or health outcomes (Canossa et al., 2019; e.g., Rattinger et al., 2016), findings may not speak directly to questions of wellbeing. In other cases, game analytics is conducted on fully anonymized data wherein researchers have access to detailed behavioral logs but little information about who the players are or survey data that could provide indicators of mental health and motivation (Vuorre et al., 2021; Zendle et al., 2023).

Trace data research into games and wellbeing comprises a much smaller part of the literature with its own set of key limitations. Where trace data has been used in video games studies, it has largely been limited to a few games (Johannes et al., 2021; Larrieu et al., 2023; M. Liu et al., 2024; Perry et al., 2018; Vuorre et al., 2022) or single

Table 1. Overview of the dataset and sources.

| Data type	Observable Period	N survey subj.	N users w/ trace data	Events	Hours	Median events / user	Median active days
Intake survey		34,922		35K		1.0	1
Daily surveys	30 days	1,275		16K		8.0	13
Biweekly surveys	10 weeks	1,948		7K		4.0	4
Time use diaries	30 days	1,103		14K		12.0	12
Attention tasks	8 weeks	1,403		3K		2.0	2
Steam	3-12 months	1,503	2805	575K	398.2K	185.0	71
Nintendo	42 months	547	1442	145K	101.5K	94.0	57
Xbox	42 months	174	326	373K	457.3K	1.4K	704
iOS	12 weeks	95	95	4K	3.4K	28.0	27
Android	12 weeks	77	77	3K	2.4K	21.0	21
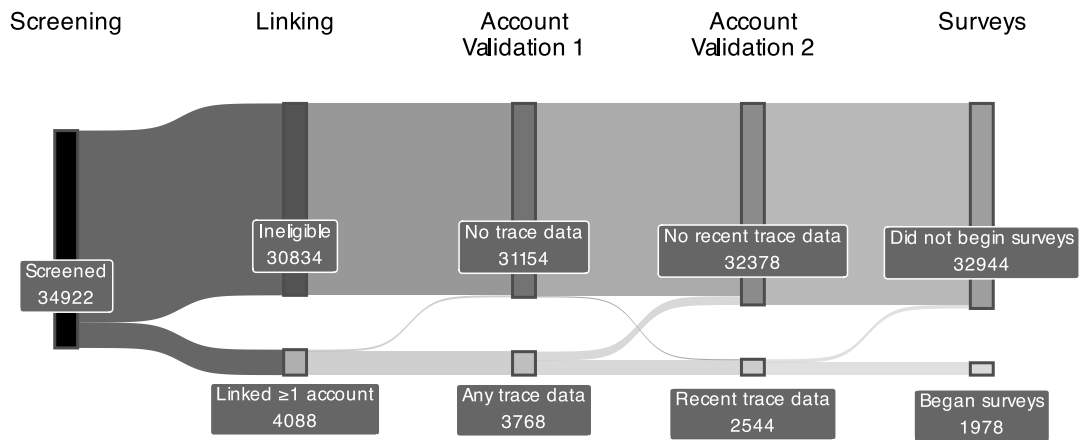
Figure 1. Participant flow during study intake.

platform (Ballou, Sewall, et al., 2024; Ballou, Vuorre, et al., 2025). However, players commonly play many games on multiple platforms. This fragmentation obscures a key part of real-world behavior: players' shifting engagement patterns across consoles, PCs, and mobile devices.

A multi-platform approach enables (1) more diverse samples by allowing researchers to recruit from more than one gaming platform whose players differ in both gaming engagement and wider demographics (Entertainment Software Association, 2024; Vahlo et al., 2017), and (2) more comprehensive and ecologically valid investigations of how and when people play, whether different platforms substitute or complement one another, and how these choices relate to daily routines or broader temporal rhythms (e.g., weekends, seasons, or daylight variation).

Taken together, these considerations underscore the urgent need for comprehensive (i.e., multi-platform) and transparent data to advance the scientific study of video game play. The dataset presented here directly addresses this need by combining longitudinal telemetry from multiple gaming platforms with rich psychological and cognitive measures collected under preregistered, ethically approved conditions. We leverage both industry-facilitated data access from major gaming platforms and open-source, participant-driven tools (King & Persily, 2020; Xiao,

2023) to achieve unprecedented coverage of naturalistic play. By making these data openly available, we aim to provide a foundation for cumulative, methodologically rigorous, and theory-driven research on digital play and its role in everyday life.

Here, we present a longitudinal dataset of digital trace data across multiple gaming platforms consisting of 2.0K participants, 16.2K daily and 6.8K biweekly surveys, 2.9K behavioral attention assessments, and 1.5M gameplay hours distributed across 1.7M sessions. Digital trace data was sourced from five platforms—Xbox, Nintendo, Steam, iOS, and Android—through distinct pipelines detailed below. The data collection methods were preregistered as part of a Stage 1 registered report (https://osf.io/pb5nu).

The dataset is openly available under a CC0 (with supplemental reidentification clause) license at https://doi.org/10.5281/zenodo.17536656 for minimally restricted reuse. Table 1 shows a high-level overview of the dataset.

Method

Design

The study consisted of four stages (Figure 1).

Table 2. Platform Details

Platform	Data Source	Account Linking Process	Type of Data Collected
Nintendo	Data-sharing agreements with Nintendo of America (US) and Nintendo of Europe (UK)	Participants share an identifier contained within a QR code on Nintendo web interface. Nintendo of America/Europe uses this identifier to retrieve gameplay data and share it with the research team.[a]	Session records (what game was played, at what time, for how long) for first-party games only (games published in whole or in part by Nintendo).[b]
Xbox (US only)	Data-sharing agreement with Microsoft	Participants consent to data sharing by opting in to the study on Xbox Insiders with their Xbox account. Microsoft retrieved and shared pseudonymized gameplay data for all consented accounts.[c]	Session records (what game was played, at what time, for how long). Game titles were replaced with a random persistent identifier, but genre(s) and age ratings are shared.
Steam	Custom web app (Gameplay.Science)	Using a web app we developed (https://gameplay.science), participants consented to have their gameplay data monitored for the duration of the study. Authentication uses the official Steam API (OpenID).	Hourly aggregates per game (every hour, the total time spent playing during the previous hour)
iOS	iOS Screen Time Screenshots	Screenshots from the iOS Screen Time app, showing details of up to 3 prior weeks of gaming. Data was extracted using OCR.	Daily aggregates (e.g., 2 total hours of gaming)
Android	Digital Wellbeing Screenshots	Screenshots from the Digital Screen Time interface (if available), showing details of up to 3 prior weeks of gaming. Data was extracted using OCR.	Daily aggregates (e.g., 2 total hours of gaming)

[a] See https://accounts.nintendo.com/qrcode.

[b] Nintendo-published games accounted for 63% of Switch playtime in our sample.

[c] See https://support.xbox.com/en-US/help/account-profile/manage-account/guide-to-insider-program.

Stage 1: Screening

In the first stage, we screened participants in order to find people aged 18–40 who (1) self-report playing video games, (2) self-report that at least 50% of their total video game play takes place on the platforms included in the study, and (3) were willing to link their gaming accounts to provide digital trace data. We screened participants from two panel sources: PureProfile and Prolific.

Participants were recruited under an initial set of ethnicity-based quotas designed to mirror the general population's demographic composition. After we reached approximately 50% of our target sample under quota constraints and found that further quota-eligible recruits were scarce, we suspended the quotas for the remainder of data collection; all subsequent participants were enrolled on a first-come, first-served basis. Final sample characteristic reflect both quota-driven and open-enrollment phases (see below).

Stage 2: Account Linking

Participants who were deemed eligible during screening proceeded directly to an account linking survey wherein they provided details of the gaming platforms they actively use. For UK participants, this includes Nintendo Switch, Steam, Android and iOS. For US participants, this includes the same four alongside Xbox. Details of how participants linked each type of account are shown in Table 2.

Stage 3: Account Validation

After players completed the account linking process, we checked each account for evidence of valid gaming—specifically, records of active gameplay sessions on one or more of Steam, Xbox, and Nintendo within the 2 weeks before survey completion. Participants who did not have recent, valid telemetry on any console platform were excluded from the rest of the study.
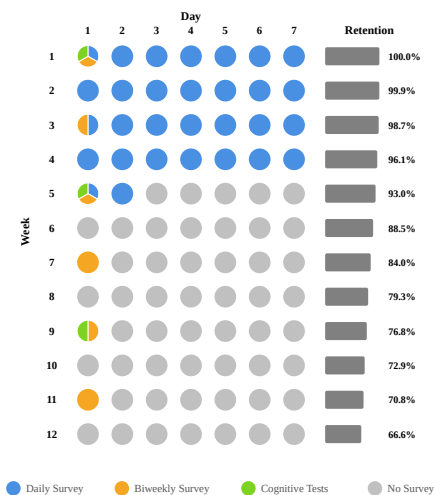
Stage 4: Surveys

Figure 2. Survey administration schedule across the 12-week study period. Participants completed biweekly surveys (orange) every two weeks, with US participants additionally completing daily surveys (blue) for the first 30 days. Cognitive tests (green) were administered during biweekly surveys at weeks 1, 5, and 9. Gray circles indicate days with no scheduled surveys. Retention percentages show the proportion of baseline participants (N=1978) who were still active at each measurement week (defined as having completed either a daily or biweekly survey at any time after that week).

Table 3. Sample demographics for qualified participants with general-population benchmarks (where available).

Variable	Level	US Sample %	US Gen Pop %	UK Sample %	UK Gen Pop %
Gender	Man	63.4	49	70.1	49.6
	Woman	30.1	48.9	25	49.5
	Other gender identity	6.5	2.1	4.9	0.9
Ethnicity	White	63.2	75.5	85	81.7
	Asian	8.1	6.3	7.7	9.3
	Two or More Races	13.9	3	4.4	2.9
	Black	9.9	13.6	2.1	4
	Other	3.9	0	0.9	2.1
	American Indian and Alaska Native	0.8	1.3	0	0
	Native Hawaiian and Other Pacific Islander	0.1	0.3	0	0
Education	Completed secondary or less	63.2	61	49.3	60.1
	Bachelor's degree	30.1	24	35	25.5
	Postgraduate	6.7	15	15.7	14.4

Eligible participants were invited to complete 6 waves of biweekly surveys, one every two weeks (Figure 2). US participants were additionally invited to complete daily surveys for 30 days, concurrently with the first biweekly surveys. During waves 1, 3, and 5, a cognitive task was also administered within the biweekly survey.

Daily survey links were sent every day at 2pm local time for the participant and remained available until 3am. Biweekly survey links were sent every second week from the first day of the study at 12pm and remained available for 96 hours.

Participants

Our final sample consists of 2544 qualified participants, selected from a pool of 34922 screened participants. Of the 2544 with recent telemetry, 1978 also completed at least one survey. On average, participants were 26.0 years old (10th percentile: 21.0, 90th percentile: 35.0). Due to errors in screening, four participants over the age of 40 were included.

Among our participants, 1852 linked digital trace data on one gaming platform, 604 linked two gaming platforms, and 88 linked three or more platforms.

Table 3 shows demographic characteristics of the final sample alongside general population benchmarks. Our participants are more likely to be male, non-binary, and bi- or multiracial than the general population. Although the demographics of the population of people who play video games are less well understood, our sample's demographics are broadly consistent with previously reported data (e.g., Entertainment Software Association, 2024).

Besides those appearing in Table 3, we collected various other demographic variables from eligible participants at intake, including employment status, height and weight, self-identified and diagnosed neurodivergence (e.g., ASD, ADHD, dyslexia), political party affiliation, marital status, caretaking responsibilities, and postal geography (general area only; first three digits of the five-digit US ZIP Code;

UK outward code). Further details of these are available in the online codebook.

Ethics and Compensation

This study received ethical approval from the Social Sciences and Humanities Inter-Divisional Research Ethics Committee at the University of Oxford (OII_CIA_23_107). All participants provided informed consent at the start of the study, including consent to their data being shared openly for reanalysis.

The data released here is pseudonymized for participant privacy. However, we recognize that the possibility of reidentification cannot be ruled out due to the detailed demographics present in the dataset alongside the digital trace data. With careful consideration of the risks (e.g., Shaw et al., 2025), we elect to share the full data, owing to several factors: (1) participants were thoroughly briefed on the procedure and directly consented to their data being pseudonymously shared; (2) for sensitive items (e.g., political affiliation), participants could select "prefer not to share"; (3) evidence suggests even a pared-down version of our dataset with less participant information may be reidentifiable based on trace data alone (e.g., Sekara et al., 2021); (4) data-sharing agreements with Nintendo and Microsoft specifically prohibit the companies themselves from connecting our survey data back to their telemetry; (5) the harms associated with reidentification are moderate, and (6) the research community has a strong interest in open data for reproducibility and cumulative science.

To balance openness with ethical responsibility, the dataset is released under a modified CC0 license that retains full reuse rights but includes an additional No-Reidentification Clause, requiring users to follow standard research ethics and prohibiting any attempt to reidentify participants.

Prolific participants were paid at a rate of £12/hour for all study components, which equates to: £0.20 for a 1-minute screening, £2 for the 10-minute intake survey (plus £5 for linking at least one account with recent data),
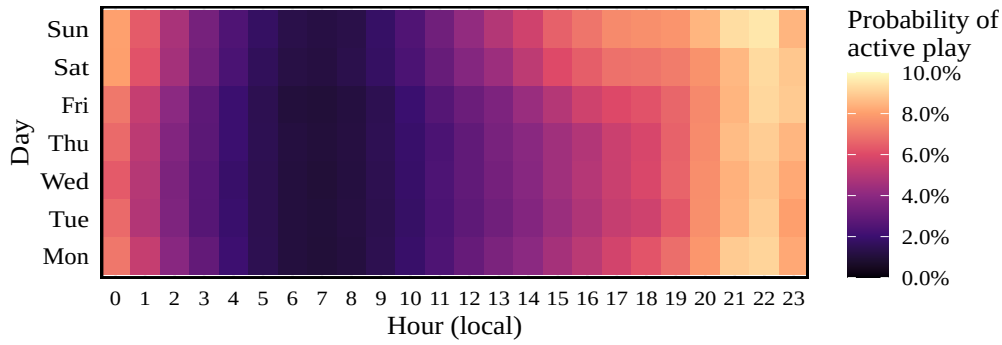
Figure 3. Diurnal play across Xbox, Nintendo, Steam.

£0.80 for each 4-minute daily survey, and £2 for each 10-minute biweekly survey. Participants received £10 bonus payments for completing at least 24 out of 30 daily surveys and/or 5 out of 6 biweekly surveys.

Dataset

Digital Trace Data

As described above, we collected video game play data from five platforms: Xbox, Nintendo Switch, Steam, iOS, and Android (full details in Table 2). To recap, on Xbox and Nintendo, we have *session-level* data, characterized by the following fields: a game ID (Xbox) or title (Nintendo), a start and end time, and genre(s). On Steam we have *hourly game-level aggregates* - every hour, how much time people spent playing each game they played that hour. On iOS and Android, we have *daily aggregates* - every day, how much time people spent playing games as a whole. We describe each platform in more detail below. For concision, we do not repeat the details of Table 2 here, but direct readers to that table or our supplementary materials for the exact variables in each platform's trace data.

To prepare the data, we first merged adjacent sessions of the same game, removed sessions of under 1 minute (reflective of artifacts such as turning on a console that had been previously paused in-game before immediately switching to a new game) and those that took place in non-games (e.g., streaming services or storefronts). These preprocessing steps resulted in the joining or removal of 56.6% of Nintendo rows, and 87.3% of Xbox rows—the vast majority of which were under 5 minutes. Orphaned and non-game sessions are not present on other platforms. We then further filtered suspicious sessions or days (see Data Quality below).

In Figure 3, we visualize the distribution of play across days and times. As expected, we find that the likelihood of play peaks on weekends from 8-11pm local time for each participant, and is lowest in the early morning.

Self-reported Gaming

We also collected self-reported gaming data in each biweekly survey. Participants estimated the time they spent playing games on platforms they had linked during the study (e.g., excluding other platforms such as Playstation) in each of the following periods: last 24 hours, last 7 days, and last 14 days. In addition, participants reported details of at least 1 and up to 3 of their most recent gaming sessions (game, date, and start/end time).

Figure 4 compares the average self-reported distribution of play across platforms to the distribution in our digital trace data capture. It is vital to note that the self-report data not be treated as ground truth: we have good evidence that people's self-reports of media use are inaccurate (Kahn et al., 2014; Parry et al., 2021), with some previous work finding systemic overestimation of video game play (Johannes et al., 2021). Nonetheless, it is likely that some portion of players' true gaming is systematically uncaptured, due to factors such as missing titles (e.g., Nintendo third party), or player privacy settings (e.g., playing in invisible mode; setting certain games to private on Steam). The figure therefore provides a useful overview of the relative coverage of different platforms in our telemetry data.

Survey measures

We collected a variety of self-report measures at different time scales. We briefly describe which constructs we collected here; for further details of the specific measures and example items, see Table A0.2.

To prepare daily and biweekly survey data, we removed duplicate entries from within the same survey wave and those that had errors in the coding of waves (e.g., administration of survey wave 31 due to technical issues), in total comprising 9.9% of daily surveys, 2.2% of biweekly surveys. We further removed rows including failed attention checks (see Data Quality below), comprising 4.7% of daily surveys and 0.5% of biweekly surveys. After cleaning, the final survey dataset includes 16,157 daily surveys and 6,844 biweekly surveys.

Trait / traitlike (baseline). We assessed chronotype, Big 5 personality, player trait typology, and gaming disorder symptoms at baseline. Gaming disorder symptoms were measured twice, at biweekly waves 1 and 6.

Daily. Daily surveys captured: basic psychological need satisfaction and frustration in life in general; basic psychological need satisfaction and frustration in the context of video games; life satisfaction; affective valence; sleep quality; stressors; types of social gaming engaged; and self-reported displacement.

Biweekly. Every two weeks we assessed: general mental wellbeing; depression symptoms; life satisfaction;
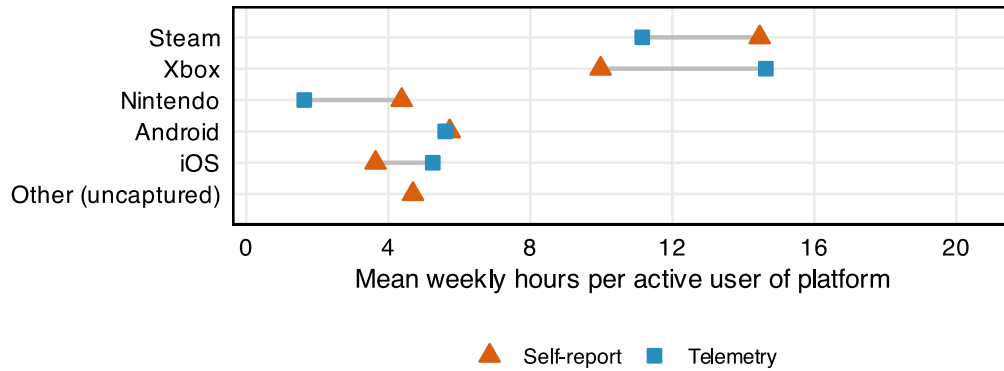
Figure 4. Average weekly playtime (minutes) by platform, self-reported vs telemetry (among users with telemetry). Lines connect estimates for the same platform. Telemetry calculated per user-platform, based only on the period between their first and last recorded session.
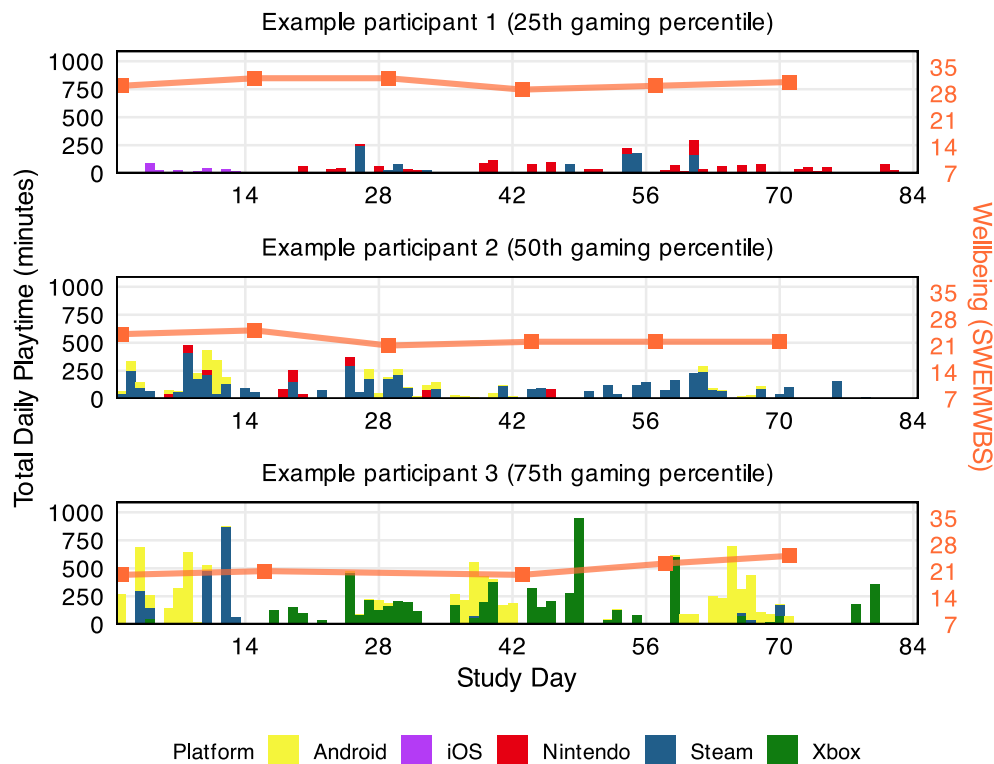

Figure 5. Sample of daily gaming patterns and mental wellbeing for three representative participants. Stacked bars represent total daily playtime across platforms. Orange line shows biweekly mental wellbeing scores (short WEMWBS) measured at six study waves. Participants were selected from those closest to the 25th, 50th, and 75th percentiles of total playtime, prioritizing those with the most varied multi-platform gaming behavior. Participant IDs: p9009984081 (25th percentile), p8809196928 (50th percentile), p7162729307 (75th percentile).

basic psychological need satisfaction and frustration – video games; and subjective displacement.

Monthly (alternating biweekly). On alternating biweekly surveys (i.e., monthly), we assessed: sleep quality; daytime sleepiness; and perceived harms and benefits of gaming.

Figure 5 illustrates the temporal dynamics of gaming behavior and mental wellbeing through three representative case studies.

Attention Control

In study waves 1, 3, and 5 we measured participants' attention control using the Simon Squared task of Burgoyne
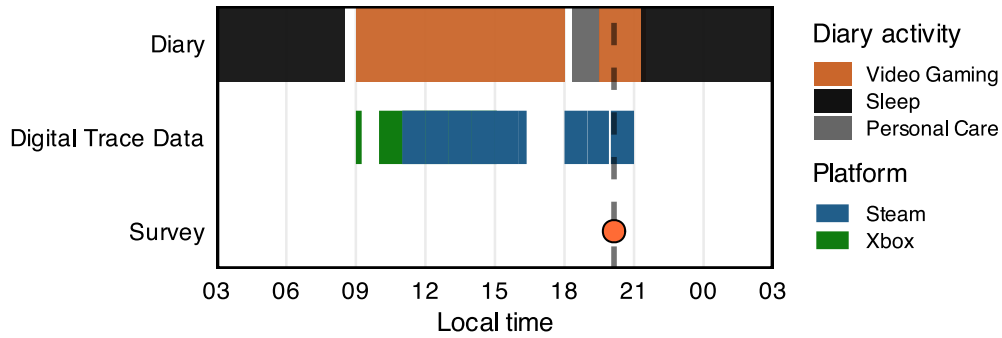
Figure 7. Illustrative overview of one study day (p9730913604, Apr 4 2025), consisting of a participant's time use diary (top row), gaming as recorded through telemetry (middle row), and daily survey completion (bottom row).

et al. (2023), using modified code from Liceralde & Burgoyne (2023). Although the original Squared tasks consist of the Simon, Stroop, and Flanker Squared, due to limited participation time we chose to only use the Simon Squared task as it had the greatest factor loading on attention control in the original study (Burgoyne et al., 2023).

The Simon Squared task is a short and validated measure of attention control that follows the standard Simon task (Simon & Rudell, 1967) but is completed in about three minutes. Participants see a target arrow pointing either left or right, with response labels "LEFT" and "RIGHT" printed underneath. Participants then must select the response option (e.g. "LEFT") that matches the arrow's direction (e.g. ◄). However, the arrow and response options can appear on either side of the screen, and participants must ignore this spatial configuration and attend only to the symbols' meanings.

After reading the instructions, participants practice for 30 seconds with auditory and text feedback for response accuracy. They then see their score from the practice trials, review the instructions again, and are given 90 seconds to gain as many points as possible. Participants gain one point for each correct response, and lose one point for each incorrect response. After the 90 seconds, the number of correct responses minus the number of incorrect responses is the participant's task score. For a complete task description, see Burgoyne et al. (2023) and Figure 6 therein.

Overall, 1149, 991, and 800 participants completed the Simon task at biweekly waves 1, 3, and 5. The average performances were approximately 10 points lower than in the in-person study of Burgoyne et al. (2023) (Figure 16). Participants on mobile devices generally attained lower scores than those not on mobile devices (Figure 6, rows), but performance was generally stable across the three waves (Figure 6, columns).
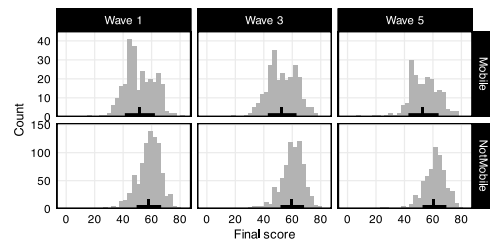

Figure 6. Histograms of participants' Simon Squared task scores across study waves. Points and bars indicate means ±1 standard deviation.

Time Use

In each daily survey (US only), participants completed a light time use diary (Hakman & Foldes, 2025) asking participants to record their daily activities from 4:00am the previous day to 4:00am the morning of survey completion. Participants provided this information by placing activities based on their start and end time using a custom web app (https://github.com/Thomhak/timediary-game). Participants could select from 11 categories: Education, Exercise and Sports, Household and Family Care, Offline Leisure & Social Life, Other Digital Media Use, Personal Care, Sleep, Travel and Transportation, Video Gaming, Work and Employment, and Other. The median number of distinct activities entered per day was 10.

Figure 7 presents an illustrative example of one participant's time use diary for a single day, alongside their gaming sessions recorded through telemetry and survey completion time.

Data Quality

While not exhaustive, the cleaned data we share implements a variety of data quality checks. Specifically, in each daily and biweekly survey, one item from the BANGS (daily) and BPNSFS (biweekly) was duplicated to assess response consistency (Meade & Craig, 2012); participants whose responses to the two identical items differed by more than one scale point were flagged for potential careless responding (daily surveys: 4.7% flagged, biweekly: 0.5%). We flagged 7.1% of time use diary entries as low quality based on excessive unaccounted-for time (>90 minutes) or missing key activities (sleep and/or eating).

In the telemetry, we use several heuristics to identify potential unreliable sessions: (a) sessions beginning or ending in the future (indicative of clock manipulation or other errors), (b) sessions longer than 8 hours (indicative of substantial idle time or background usage), (c) 3 or more games played simultaneously (indicative of inactivity on multiple games or data corruption). In total, we removed 2.2% of Xbox sessions, and 12.4% of Nintendo sessions based on these criteria.

Concurrency was a key quality factor for trace data, and particularly affected Steam: players sometimes ran multiple games simultaneously, resulting in hourly intervals where total recorded playtime exceeded 60 minutes. This occurred because Steam's API reports cumulative playtime for each game independently, and our reconstruction of approximate session times from polling intervals could overlap when multiple games were active. In the raw data, 21.1% of player-hour combinations exceeded 60 minutes. To address this, we implemented proportional scaling at the hourly level. For any hour where the sum of playtime across all games exceeded 60 minutes, we scaled each game's contribution proportionally such that the total equaled 60 minutes, preserving the relative distribution of play across titles while ensuring temporal plausibility. Overall, 30.8% of session segments required scaling, with a median reduction of 2.1 minutes per scaled segment. A further 0.10% of hourly telemetry records indicated concurrent gaming across multiple platforms within the same hour, suggesting potential background usage or idling on at least one platform; these records are retained but flagged for caution.

Missingness

As with most longitudinal studies, attrition and missing data present important challenges for data quality and statistical inference. Table 4 presents a comprehensive overview of missingness patterns across all data sources in our study, broken down by data type (Survey, Cognitive Task, and Telemetry).

The unit of observation differs across data types. For surveys and cognitive tasks, the unit is a completed survey or task. For telemetry, the unit varies by platform: Xbox and Nintendo use binary account linking, with data coverage considered maximal once linked (data provided directly by platform holders); Steam is measured hourly, with observations representing whether the participant's profile was publicly visible in each hour; iOS and Android are measured daily, with observations representing whether a valid screenshot was submitted covering that day's gaming.

Table 4. Missingness patterns across survey, telemetry, and cognitive task data by region. The table shows the number of participants (N), maximum possible observations, median number of missing observations per participant (with maximum in parentheses), and percentage of data completeness for each measure.

Data Type	Measure	N subj.[a]	N survey subj.	N obs.	Max. possible obs.[b]	% Missing
Survey	Daily surveys	1,275	1,275	16,157	38,250	57.8%
	Bi-weekly surveys	1,948	1,948	6,844	11,688	41.4%
Task	Time use diaries	1,103	1,103	13,789	33,090	58.3%
	Attention tasks	1,403	1,403	2,940	5,844	49.7%
Telemetry	Xbox	326	174	174	174	0.0%
	Nintendo	1,442	547	547	547	0.0%
	Steam	2,805	1,503	2,955,796	3,030,048	2.5%
	iOS	95	95	3,608	7,980	54.8%
	Android	77	77	2,518	6,468	61.1%

[a] N subjects can exceed N survey subjects because some participants (1) provided digital trace data that was not recent enough to qualify, or (2) provided recent digital trace data but did not return for subsequent surveys.

[b] Max. possible obs.: Daily survey = N × 30; Biweekly survey = N × 6: Xbox/Nintendo = N participants with ≥1 session; Steam = N × 84 days × 24 hours (visibility hours); iOS/Android = N × 84 days; Simon Task = N × 3 administrations.

Discussion

Spanning 1.5M hours, 3,776 user-platform combinations, and 10,475 games, the dataset presented here represents a major advance in the comprehensiveness and scale at which video game play has been captured. Previous studies on games and health have typically used (1) self-report data with substantial limitations on accuracy, time range, and detail; (2) large-scale data that is either anonymized and not linkable to individuals or survey responses (Vuorre et al., 2021; Zendle et al., 2023) and/or focused on commercially-relevant game analytics topics (Canossa et al., 2019; Rattinger et al., 2016), and thus lacking good measures of ground truth for questions about player health; or (3) small-scale trace data from single platforms or games, providing only a small slice of overall gaming behaviour and thus limiting generalizability (Ballou, Sewall, et al., 2024; Ballou, Vuorre, et al., 2025; Johannes et al., 2021; Larrieu et al., 2023; Perry et al., 2018; Vuorre et al., 2022).

Digital Trace Data Infrastructure

This dataset delivers on both calls for greater collaboration between the tech industry and academics (King & Persily, 2020; Lazer et al., 2020), as well as simultaneous calls for better trace data access without direct industry involvement (Boeschoten et al., 2023; Valkenburg et al., 2024; Xiao, 2023). The data-sharing agreements with Nintendo of America, Nintendo of Europe, and Microsoft are the first of their kind, enabling researchers to access telemetry data from major gaming platforms while ensuring user privacy and data security. By making this dataset openly available under a modified CC0 license, we aim to catalyze further research in this area, allowing other researchers to explore the rich data we have collected.

At the same time, the Steam and smartphone data collection show the value of open trace data to study platforms that are less willing or able to participate actively. Several previous studies have collected smartphone data using screenshots, with some success in automating the extraction of these data (Feiz et al., 2022; Y. Liu et al., 2024; Parry & Sewall, 2021). Open trace data infrastructure remains an arms race, often relying on industry-provided APIs that can change or be deprecated at any time (Ballou et al., 2023; Davidson et al., 2023). Nonetheless, we believe it is vital for researchers to advance both pathways for the best chance to understand the effects of technology.

Empirical Opportunities

We believe this dataset has potential to address a wide variety of pertinent research questions for the field. Some of these questions will be addressed in forthcoming registered reports (Ballou, Földes, et al., 2025): specifically, we have plans to test (1) key hypotheses from the Basic Needs in Games model (Ballou & Deterding, 2024) about how gaming relates to basic psychological needs over time, (2) the relationship between late-night gaming and sleep, and (3) the relationship between playtime in different genres and wellbeing.

The data's richness means that researchers can explore numerous other questions (or, indeed, conduct and compare alternative analysis approaches to the above questions). We particularly encourage work that tests explicit causal models—when assumptions about data generating processes are transparent, observational data can provide meaningful falsifying tests of causal effects. A recent paper presents 13 such potential causal effects of games on wellbeing (Ballou, Hakman, et al., 2025), and several of these could tested using the data here. For example, the relationship between *action game mechanics, executive function, and stress* (Bediou et al., 2018; Hilgard et al., 2019) could be investigated by combining the game genre data, stress data from daily surveys, and digital trace data with genre coding. The relationship between *social experiences in games and wellbeing* (Mandryk et al., 2020) could be investigating by coding games for single- or multiplayer features, using the self-reported social context of play, and subsequent relatedness satisfaction scores.

The dataset need not be limited to questions of wellbeing. Other questions of potential interest include:

How do neurotypical and neurodiverse players differ in their gaming behavior? Using the neurodivergence data we collected (which includes, for example, 362 participants who identify as having autism and 501 who identify as having ADHD), researchers can explore differences in the types of games, gaming experiences, and trajectory of play over time between neurotypical and neurodiverse players. Neurodiversity in games has regularly been studied in the context of specific games and with qualitative methods (Kilmer et al., 2023; Zolyomi & Schmalz, 2017), with some researchers noting that research focuses on therapeutic interventions delivered through games to "cure" neurodivergent players of undesirable traits (Spiel & Gerling, 2021). Research on neurodivergent players' naturalistic gaming behavior across platforms has to date received less attention.

How do seasons and weather impact play? Because we capture time-stamped play sessions alongside participants' geographic locations, researchers can merge in high-resolution weather and daylight data to examine how environmental factors influence gaming behavior. Causal inference techniques such as inverse probability weighting can enable estimates of how, when, and how much people play in response to seasonal and meteorological changes. Quantifying these effects can better distinguish environmental factors driving video game play from other drivers (like work schedules or weekend routines) and improve our ability to predict regional changes to gaming behavior on a day-to-day basis (Palomba, 2019), or the behavioral impacts of climate change in the long term.

Does irregular, extended, or overnight play predict subsequent rises in problematic play, and can telemetry provide early warning signals? Using the data we collected about displacement—the extent to which gaming supports or interferes with other areas of life—as well as about gaming disorder (at Waves 1 and 6), telemetry can provide insights into patterns of gaming that coincide or lead to subsequent changes in the degree of problematic play. Long-term, more accurately identifying the longitudinal behavioral correlates of maladaptive gaming —beyond simple heuristics such as the average amount of weekly play in the last 6 months—can help us create personalized models and early warning systems for players and parents.

We encourage researchers from a wide range of disciplines to explore these or other questions using the data we present.

Limitations

While this dataset represents a substantial step forward in holistic coverage of video game play, it remains imperfect: we did not capture data on PlayStation (~19% of gaming market) or computer games played outside the Steam platform (~11% of gaming market); on Nintendo, we do not have access to third-party titles (42% of Nintendo play, based on Nintendo-provided metrics for each player in our sample); and our coverage of smartphone play is limited by the onerous workflow limiting the response rate, and the difficulties of reliable OCR extraction.

We further are unable to identify idle time (when players have a game open but are not actively playing it) and account sharing (when players let friends or family use their account); some playtime values may therefore be overestimates of the person's true playtime, though we are unable to say by how much.

Finally, while our sample is large, diverse, and reasonably well matched with national averages on certain demographic characteristics (e.g., ethnicity and education), is it unlikely to be representative either of the general population or the population of adults who play games. Selection effects are inherent to our study: not all players will be willing to share their gaming history and identifiable information with researchers, or to participate in an intensive battery of surveys. However, we are largely unable to quantify representativeness, as the (adult) gaming population is poorly understood in its own right. Although we have broad estimates of engagement among key demographic segments showing, for example, that men are more likely to play games (Entertainment Software Association, 2024; Ofcom, 2023), digital trace data describing gaming in terms of amount, seasonality, specific titles and genres, or other factors remains lacking. In other words, we cannot say how representative our sample is of players in the US or UK, because we do not yet know what representative gaming behavior looks like (see e.g., Rehbein et al., 2016) —or whether representativeness is even a meaningful aim (Kaye, 2019).

Future Work

The trace data presented here is broad in scope but limited in granularity: we capture all gaming activity on a given platform, but not what happens within individual games. Prior work and theory make clear that in-game behaviors (e.g., what role a player adopts, whether they compete or cooperate, or how they perform in competitive modes) are critical determinants of player experience and thereby wellbeing (see e.g., Elson et al. (2014) for a review of how in-game contexts shape effects). This highlights a fundamental trade-off in digital trace research between breadth —how comprehensively play can be captured across platforms—and depth—the granularity of in-game behaviors and experiences. At present, our dataset emphasizes breadth, but we see strong potential in future study designs that combine platform-level telemetry with targeted in-game behavioral data to provide a more complete picture.

We also see strong potential in combining digital trace data with experimental designs that enable stronger causal inference—for example, randomizing players to single-player games only, or restricting play to certain times of day, to examine effects on social wellbeing or sleep. Previous researchers have noted a dearth of digital trace data-backed field experiments, while highlighting their potential (Stier et al., 2020): Trace data not only captures naturalistic gaming behavior but also allows researchers to assess *substitution* (what games or platforms participants switch to under intervention) and *adherence* (how closely they follow assigned play patterns).

Data Availability

References

Almeida, D. M., Wethington, E., & Kessler, R. C. (2002). The Daily Inventory of Stressful Events: An Interview-Based Approach for Measuring Daily Stressors. *Assessment*, *9*(1), 41–55. https://doi.org/10.1177/1073191102091006

Ballou, N., Denisova, A., Ryan, R., Rigby, C. S., & Deterding, S. (2024). The basic needs in games scale (BANGS): A new tool for investigating positive and negative video game experiences. *International Journal of Human-Computer Studies*, *188*, 103289. https://doi.org/10.1016/j.ijhcs.2024.103289

Ballou, N., & Deterding, S. (2024). The Basic Needs in Games Model of Video Game Play and Mental Health. *Interacting with Computers*, iwae042. https://doi.org/10.1093/iwc/iwae042

Ballou, N., Földes, T. A., Hakman, T., Vuorre, M., Magnusson, K., & Przybylski, A. K. (2025). *Psychological wellbeing, sleep, and video gaming: Analyses of comprehensive digital traces [stage 1 programmatic registered report]*. PsyArXiv. https://doi.org/10.31234/osf.io/kvqdy_v1

Ballou, N., Hakman, T., Vuorre, M., Magnusson, K., & Przybylski, A. K. (2025). How do video games affect mental health? A narrative review of 13 proposed mechanisms. *Technology, Mind, and Behavior*, *6*(2), 1–21. https://doi.org/10.1037/tmb0000152

Ballou, N., Sewall, C. J. R., Ratcliffe, J., Zendle, D., Tokarchuk, L., & Deterding, S. (2024). Registered report evidence suggests no relationship between objectively-tracked video game playtime and wellbeing over 3 months. *Technology, Mind, and Behavior*, *5*(1), 1–15. https://doi.org/10.1037/tmb0000124

Ballou, N., Vuorre, M., Hakman, T., & Przybylski, A. K. (2025). Perceived value of video games, but not hours played, predicts mental well-being in casual adult nintendo players. *Royal Society Open Science*, *12*, 241174. https://doi.org/https://doi.org/10.1098/rsos.241174

Ballou, N., Zendle, D., Cutting, J., & Gordon-Petrovskaya, E. (2023). Four dilemmas for video game effects scholars: How digital trace data can improve the way we study games. *ICA'23*. https://doi.org/https://osf.io/ua6nr_v1

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, *144*(1), 77–110. https://doi.org/10.1037/bul0000130

Boeschoten, L., De Schipper, N. C., Mendrik, A. M., Van Der Veen, E., Struminskaya, B., Janssen, H., & Araujo, T. (2023). Port: A software tool for digital data dona-

tion. *Journal of Open Source Software*, *8*(90), 5596. https://doi.org/10.21105/joss.05596

Burgess, R., Dolan, E., Poon, N., Jenneson, V., Pontin, F., Sivill, T., Morris, M., & Skatova, A. (2024). *The Potential of Digital Footprint Data for Health & Wellbeing Research*. Pre-published. https://doi.org/10.31234/osf.io/9jgn2

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). Nature and measurement of attention control. *Journal of Experimental Psychology: General*, *152*(8), 2369–2402. https://doi.org/10.1037/xge0001408

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Canossa, A., Azadvar, A., Harteveld, C., Drachen, A., & Deterding, S. (2019). Influencers in multiplayer online shooters: Evidence of social contagion in playtime and social play. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. https://doi.org/10.1145/3290605.3300489

Cantril, H. (1965). *The pattern of human concerns*. Rutgers University Press.

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The consensus sleep diary: Standardizing prospective sleep self-monitoring. *Sleep*, *35*(2), 287–302. https://doi.org/10.5665/sleep.1642

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., Duriez, B., Lens, W., Matos, L., Mouratidis, A., Ryan, R. M., Sheldon, K. M., Soenens, B., Van Petegem, S., & Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, *39*(2), 216–236. https://doi.org/10.1007/s11031-014-9450-1

Davidson, B. I., Wischerath, D., Racek, D., Parry, D. A., Godwin, E., Hinds, J., Van Der Linden, D., Roscoe, J. F., Ayravainen, L., & Cork, A. G. (2023). Platform-controlled social media APIs threaten open science. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01750-2

Elson, M., Breuer, J., Ivory, J. D., & Quandt, T. (2014). More Than Stories With Buttons: Narrative, Mechanics, and Context as Determinants of Player Experience in Digital Games: Narrative, Mechanics, and Context in Digital Games. *Journal of Communication*, *64*(3), 521–542. https://doi.org/10.1111/jcom.12096

Entertainment Software Association. (2024). *2024 essential facts about the u.s. Video game industry*. https://www.theesa.com/wp-content/uploads/2024/05/Essential-Facts-2024-FINAL.pdf

Feiz, S., Wu, J., Zhang, X., Swearngin, A., Barik, T., & Nichols, J. (2022). Understanding screen relationships from screenshots of smartphone applications. *27th International Conference on Intelligent User Interfaces*, 447–458. https://doi.org/10.1145/3490099.3511109

Freelon, D. (2014). On the Interpretation of Digital Trace Data in Communication and Social Computing Research. *Journal of Broadcasting & Electronic Media*, *58*(1), 59–75. https://doi.org/10.1080/08838151.2013.875018

Griffioen, N., Rooij, M. van, Lichtwarck-Aschoff, A., & Granic, I. (2020). Toward improved methods in social media research. *Technology, Mind, and Behavior*, *1*(1). https://doi.org/10.1037/tmb0000005

Hakman, T., & Foldes, T. A. (2025). *LO-TUD - Light Online Time Use Diary* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.17592564

Hilgard, J., Sala, G., Boot, W. R., & Simons, D. J. (2019). Overestimation of action-game training effects: Publication bias and salami slicing. *Collabra: Psychology*, *5*(1), 30. https://doi.org/10.1525/collabra.231

Johannes, N., Vuorre, M., & Przybylski, A. K. (2021). Video game play is positively correlated with well-being. *Royal Society Open Science*, *8*(2), rsos.202049, 202049. https://doi.org/10.1098/rsos.202049

Johns, M. W. (1991). A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep*, *14*(6), 540–545. https://doi.org/10.1093/sleep/14.6.540

Kahn, A. S., Ratan, R., & Williams, D. (2014). Why we distort in self-report: Predictors of self-report errors in video game play. *Journal of Computer-Mediated Communication*, *19*(4), 1010–1023. https://doi.org/10.1111/jcc4.12056

Kahn, A. S., Shen, C., Lu, L., Ratan, R. A., Coary, S., Hou, J., Meng, J., Osborn, J., & Williams, D. (2015). The Trojan Player Typology: A cross-genre, cross-cultural, behaviorally validated scale of video game play motivations. *Computers in Human Behavior*, *49*, 354–361. https://doi.org/10.1016/j.chb.2015.03.018

Kaye, L. K. (2019). Gaming classifications and player demographics. In *The oxford handbook of cyberpsychology* (pp. 608–623). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198812746.013.1

Kilmer, E., Spangler, J., & Kilmer, J. (2023). Therapeutically applied minecraft groups with neurodivergent youth. *F1000Research*, *12*, 216. https://doi.org/10.12688/f1000research.129090.2

King, G., & Persily, N. (2020). A new model for industry–academic partnerships. *PS: Political Science & Politics*, *53*(4), 703–709. https://doi.org/10.1017/S1049096519001021

Larrieu, M., Fombouchet, Y., Billieux, J., & Decamps, G. (2023). How gaming motives affect the reciprocal relationships between video game use and quality of life: A prospective study using objective playtime indicators. *Computers in Human Behavior*, *147*, 107824. https://doi.org/10.1016/j.chb.2023.107824

Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportu-

nities. *Science*, *369*(6507), 1060–1062. https://doi.org/10.1126/science.aaz8170

Liceralde, V. R. T., & Burgoyne, A. P. (2023). *Squared tasks of attention control for jsPsych*. https://doi.org/10.5281/zenodo.8313315

Liu, M., Choi, S., Kim, D. O. (Donna)., & Williams, D. (2024). Connecting in-game performance, need satisfaction, and psychological well-being: A comparison of older and younger players in world of tanks. *New Media & Society*, *26*(2), 692–710. https://doi.org/10.1177/14614448211062545

Liu, Y., Klassen, G., Mee, J., Pointer, J., Baloch, M., Marciano, L., & Osgood, N. (2024). User-donated screenshots analysis: Feasibility of a new approach to collect objective social media app usage in adolescents. In R. Thomson, A. Hariharan, S. Renshaw, S. Alkhateeb, A. Burger, P. Park, & A. Pyke (Eds.), *Social, cultural, and behavioral modeling* (Vol. 14972, pp. 80–89). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72241-7_8

Mandryk, R. L., Frommel, J., Armstrong, A., & Johnson, D. (2020). How passion for playing world of warcraft predicts in-game social capital, loneliness, and well-being. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.02165

Martela, F., & Ryan, R. M. (2024). Assessing Autonomy, Competence, and Relatedness Briefly: Validating Single-Item Scales for Basic Psychological Need Satisfaction. *European Journal of Psychological Assessment*, 1015–5759/a000846. https://doi.org/10.1027/1015-5759/a000846

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Ofcom. (2023). *Online Nation 2023 Report* (p. 106). https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/online-nation/2023/online-nation-2023-report.pdf?v=368355

Palomba, A. (2019). Digital seasons: How time of the year may shift video game play habits. *Entertainment Computing*, *30*, 100296. https://doi.org/10.1016/j.entcom.2019.100296

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*, 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Parry, D. A., & Sewall, C. J. R. (2021). *How consistent are smartphone application preferences? A descriptive study of mobile application repertoires using behavioral data*. https://doi.org/10.31235/osf.io/jzxun

Perry, R., Drachen, A., Kearney, A., Kriglstein, S., Nacke, L. E., Sifa, R., Wallner, G., & Johnson, D. (2018). Online-only friends, real-life friends or strangers? Differential associations with passion and social capital in video game play. *Computers in Human Behavior*, *79*, 202–210. https://doi.org/10.1016/j.chb.2017.10.032

Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., Cella, D., & PROMIS Cooperative Group. (2011). Item Banks for Measuring Emotional Distress From the Patient-Reported Outcomes Measurement Information System (PROMIS®): Depression, Anxiety, and Anger. *Assessment*, *18*(3), 263–283. https://doi.org/10.1177/1073191111411667

Pontes, H. M., Schivinski, B., Sindermann, C., Li, M., Becker, B., Zhou, M., & Montag, C. (2019). Measurement and Conceptualization of Gaming Disorder According to the World Health Organization Framework: the Development of the Gaming Disorder Test. *International Journal of Mental Health and Addiction*, *19*, 508528. https://doi.org/10.1007/s11469-019-00088-z

Rattinger, A., Wallner, G., Drachen, A., Pirker, J., & Sifa, R. (2016). Integrating and inspecting combined behavioral profiling and social network models in destiny. In G. Wallner, S. Kriglstein, H. Hlavacs, R. Malaka, A. Lugmayr, & H.-S. Yang (Eds.), *Entertainment computing - ICEC 2016* (Vol. 9926, pp. 77–89). Springer International Publishing. https://doi.org/10.1007/978-3-319-46100-7_7

Rehbein, F., Staudt, A., Hanslmaier, M., & Kliem, S. (2016). Video game playing in the general adult population of germany: Can higher gaming time of males be explained by gender specific genre preferences? *Computers in Human Behavior*, *55*, 729–735. https://doi.org/10.1016/j.chb.2015.10.016

Roenneberg, T., Wirz-Justice, A., & Merrow, M. (2003). Life between Clocks: Daily Temporal Patterns of Human Chronotypes. *Journal of Biological Rhythms*, *18*(1), 80–90. https://doi.org/10.1177/0748730402239679

Sekara, V., Alessandretti, L., Mones, E., & Jonsson, H. (2021). Temporal and cultural limits of privacy in smartphone app usage. *Scientific Reports*, *11*(1), 3861. https://doi.org/10.1038/s41598-021-82294-1

Sewall, C. J. R., Goldstein, T. R., Wright, A. G. C., & Rosen, D. (2022). Does Objectively Measured Social-Media or Smartphone Use Predict Depression, Anxiety, or Social Isolation Among Young Adults? *Clinical Psychological Science*, *10*(5), 997–1014. https://doi.org/10.1177/21677026221078309

Sewall, C. J. R., & Parry, D. A. (2021). The Role of Depression in the Discrepancy Between Estimated and Actual Smartphone Use: A Cubic Response Surface Analysis. *Technology, Mind, and Behavior*, *2*(2). https://doi.org/10.1037/tmb0000036

Shaw, H., Brown, O., Hinds, J., Nightingale, S. J., Towse, J., & Ellis, D. A. (2025). The DECIDE framework: Describing ethical choices in digital-behavioral-data explorations. *Advances in Methods and Practices in Psychological Science*, *8*(3), 25152459251361013. https://doi.org/10.1177/25152459251361013

Siebers, T., Beyens, I., Baumgartner, S. E., & Valkenburg, P. M. (2024). Adolescents' digital nightlife: The comparative effects of day- and nighttime smartphone use on sleep quality. *Communication Re-*

search, 00936502241276793. https://doi.org/10.1177/00936502241276793

Sifa, R., Drachen, A., & Bauckhage, C. (2021). Large-scale cross-game player behavior analysis on steam. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, *11*(1), 198–204. https://doi.org/10.1609/aiide.v11i1.12804

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304. https://doi.org/10.1037/h0020586

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Spiel, K., & Gerling, K. (2021). The purpose of play: How HCI games research fails neurodivergent populations. *ACM Transactions on Computer-Human Interaction*, *28*(2), 1–40. https://doi.org/10.1145/3432245

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, *38*(5), 503–516. https://doi.org/10.1177/0894439319843669

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes*, *5*(1), 63. https://doi.org/10.1186/1477-7525-5-63

Vahlo, J., Kaakinen, J. K., Holm, S. K., & Koponen, A. (2017). Digital game dynamics preferences and player types. *Journal of Computer-Mediated Communication*, *22*(2), 88–103. https://doi.org/10.1111/jcc4.12181

Valkenburg, P. M., Van Der Wal, A., Siebers, T., Beyens, I., Boeschoten, L., & Araujo, T. (2024). It is time to ensure research access to platform data. *Nature Human Behaviour*, *9*(1), 1–2. https://doi.org/10.1038/s41562-024-02066-5

Vuorre, M., Johannes, N., Magnusson, K., & Przybylski, A. K. (2022). Time spent playing video games is unlikely to impact well-being. *Royal Society Open Science*, *9*(7), 220411. https://doi.org/10.1098/rsos.220411

Vuorre, M., Zendle, D., Petrovskaya, E., Ballou, N., & Przybylski, A. K. (2021). A large-scale study of changes to the quantity, quality, and distribution of video game play during a global health pandemic. *Technology, Mind, and Behavior*, *2*(4). https://doi.org/10.1037/tmb0000048

Xiao, L. Y. (2023). Debate: Academics should collaborate with the technology industry, but not in lieu of noncollaborative research. *Child and Adolescent Mental Health*, camh.12644. https://doi.org/10.1111/camh.12644

Yap, V., Skeggs, A., Ferguson, A. M., Leyland-Craggs, A., Boeschoten, L., Welbers, K., Kurten, S., & Orben, A. (2024). *Digital data donation with adolescents*. PsyArXiv. https://doi.org/10.31234/osf.io/hnvpy

Zendle, D., Flick, C., Halgarth, D., Ballou, N., Demediuk, S., & Drachen, A. (2023). Cross-cultural patterns in mobile playtime: An analysis of 118 billion hours of human data. *Scientific Reports*, *13*(1), 386. https://doi.org/10.1038/s41598-022-26730-w

Zolyomi, A., & Schmalz, M. (2017). Mining for social skills: Minecraft in home and therapy for neurodiverse youth. *Proceedings of the 50th Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2017.411

A1 Appendix

A1.1 Deviations from Preregistration

We made several deviations from our preregistration to ensure we could recruit enough high-quality participants to meet our sample size goals. In our view, none are so severe enough to threaten the validity of the study. Deviations are summarised in Table A1.1.

Table A1.1. Summary of deviations from preregistration

Preregistered	Actual	Justification for Deviation
All participants sourced from PureProfile	Participants sourced from both PureProfile and Prolific	Exhausted PureProfile participant pool before reaching required sample size
Screening sample would be nationally representative by ethnicity and gender	Approximately 50% of screening was done using quotas for national representativeness by ethnicity and gender; all subsequent sampling used convenience sampling with no quotas	Exhausted participant pools of smaller demographic categories on both Prolific and PureProfile before reaching required sample size
Sample consists of participants aged 18–30 in the US and 18–75 in the UK	Sample consists of participants aged 18-40 in both regions	(1) Unable to recruit enough participants in the US aged 18–30; (2) near-zero qualification rates from UK adults over 50; (3) desire for results from both regions to be more easily comparable
To qualify, >=75% of a participant's total gaming must take place on platforms included in the study (Xbox, Steam, Nintendo Switch)	To qualify, >=50% of a participant's total gaming must take place on platforms included in the study (Xbox, Steam, Nintendo Switch)	Low rates of study qualification at 75% threshold, in large part due to substantial uncaptured Playstation play
Qualification contingent upon valid telemetry within last 7 days	Qualification contingent upon valid telemetry within last 14 days	Feedback from participants indicating that play during a 7-day period was subject to too many fluctuations (e.g., a busy workweek)
Daily and biweekly surveys sent at 7pm local time	Daily and biweekly surveys sent at 2pm local time	Feedback from participants indicating that evening plans often interfered with survey completion and thus adversely affected response rate
Session-level Android data captured via the ActivityWatch app	Daily-level Android data captured using screenshots of the Digital Wellbeing interface	Restrictions in PureProfile's privacy policy preventing installation of 3rd party apps; technical challenges in supporting users with the installation and data export

Table A1.2. Summary of survey measures used in the study

Construct	Measure	Example Item	Response format	Frequency
Big 5 Personality	BFI-2-XS (Soto & John, 2017)	I am someone who…is compassionate, has a soft heart.	5-pt Likert scale from 1 (Disagree strongly) to 5 (Agree strongly)	Once (baseline)
Chronotype	Munich Chronotype Questionnaire (Roenneberg et al., 2003)	I go to bed at…	Times and numbers of minutes	Once (baseline)
Player Trait Typology	Trojan Player Typology (Kahn et al., 2015)	It's important to me to play with a tightly knit group.	5-pt Likert scale from 1 (Strongly disagree) to 5 (Strongly agree)	Once (baseline)
Gaming Disorder Symptoms[a]	Gaming Disorder Test (Pontes et al., 2019)	In the past 3 months…I have had difficulties controlling my gaming activity.	5-pt Likert scale from 1 (Never) to 5 (Very often)	Twice (biweekly waves 1 & 6)
Daytime sleepiness	Epworth Sleepiness Scale (Johns, 1991)	How likely are you to doze off or fall asleep in the following situations, in comparison to feeling just tired? […] Watching TV	4-pt Likert scale from 1 (No chance of dozing) to 4 (High chance of dozing)	Monthly (alternating biweekly surveys)
Harms and benefits of gaming	2 free text questions	Do you feel that gaming is sometimes a problem for you? Please describe.	Open text	Monthly (alternating biweekly surveys)
Sleep quality	Pittsburgh Sleep Quality Index (Buysse et al., 1989)	During the past month, what time have you usually gotten up in the morning?	Various	Monthly (alternating biweekly surveys)
Basic psychological need satisfaction and frustration - video games	Basic Needs in Games scale (Ballou, Denisova, et al., 2024), gaming in general version	When playing video games during the last 2 weeks…I could play in the way I wanted.	7-pt Likert scale from 1 (very strongly disagree) to 7 (very strongly agree)	Biweekly (every 2 weeks)
Depression symptoms	PROMIS Short Form 8a Adult Depression Scale (Pilkonis et al., 2011)	In the past 7 days…I felt that I had nothing to look forward to.	5-pt Likert scale from 1 (Never) to 5 (Always)	Biweekly (every 2 weeks)
General Mental Wellbeing	Warwick-Edinburgh Mental Wellbeing Scale (Tennant et al., 2007)	I've been feeling optimistic about the future	5-pt Likert scale from 1 (none of the time) to 5 (all of the time)	Biweekly (every 2 weeks)
Life satisfaction	Cantril Self-anchoring Scale (Cantril, 1965)	On which step of [a ladder from 0 to 10 representing the best possible life] would you say you personally feel you stood over the past two weeks?	10-pt unlabeled scale from 0 to 10	Biweekly (every 2 weeks)
Subjective displacement	ad hoc	Over the last two weeks, to what extent has the time you spend playing video games influenced the following areas of your life? […] Work/school performance	7-pt Likert scale from 1 (greatly interfered) to 7 (greatly supported)	Biweekly (every 2 weeks)
Self-reported Playtime	Time spent playing games on platforms they had linked during the study (e.g., excluding other platforms such as Playstation) in each of the following periods: last 24 hours, last 7 days, and last 14 days.	NA	Time estimates for last 24 hours, last 7 days, last 14 days	Biweekly (every 2 weeks)
Self-reported recent sessions	Details of at least 1 and up to 3 of their most recent gaming sessions (game, date, and start/end time).	NA	Game title, date, and start/end time for 1-3 sessions	Biweekly (every 2 weeks)
Affective valence	ad hoc	How are you feeling right now?	Visual Analogue Scale from 1 (very bad) to 100 (very good)	Daily
Basic psychological need satisfaction and frustration - life in general	Basic Psychological Need Satisfaction and Frustration Scale (Chen et al., 2015), brief version (Martela & Ryan, 2024)	In the last 24 hours…I was able to do things I really want and value in life.	7-pt Likert scale from 1 (very strongly disagree) to 7 (very strongly agree)	Daily
Basic psychological need satisfaction and frustration - video games	Basic Needs in Games scale (Ballou, Denisova, et al., 2024), brief session-level version	In my most recent session of X…I felt disappointed with my performance.	7-pt Likert scale from 1 (very strongly disagree) to 7 (very strongly agree)	Daily
Life satisfaction	Cantril Self-anchoring Scale (Cantril, 1965), daily version	I was satisfied with my life today.	Visual Analogue Scale from 1 (Strongly disagree) to 100 (Strongly agree)	Daily
Self-reported displacement	ad hoc	Think back to your most recent gaming session. If you hadn't played a game, what would you most likely have done instead?	Open response	Daily
Sleep Quality	Sleep quality item (Item 9) from Consensus Sleep Diary (Carney et al., 2012)	How do you rate the quality of your sleep?	5-pt Likert scale from 1 (very poor) to 5 (very good)	Daily
Stressors	Daily Inventory of Stressful Events (Almeida et al., 2002), modified for digital delivery	[In the last 24 hours], what kinds of stressful event(s) occurred? [Participant selects among 7 options, including e.g. argument or disagreement]	Yes/No, followed by a 4-pt Likert scale from 1 (Not at all stressful) to 4 (Very stressful)	Daily
Social context of play	Types of social play engaged during the last 24 hours (single-player games only, multiplayer with real-world friends, multiplayer with online-only friends, multiplayer with strangers). Participants could select more than one option.	NA	Multiple selection from listed options	Daily

[a] Measured twice, at biweekly waves 1 and 6.

A1.2 Codebook

Interactive codebook available in the HTML version of this document: https://digital-wellbeing.github.io/open-play/#codebook.