

VIBES: Exploring Viewer Spatial Interactions as Direct Input for Livestreamed Content

Michael Yin

University of British Columbia
Vancouver, BC, Canada
jiyin@cs.ubc.ca

Robert Xiao

University of British Columbia
Vancouver, BC, Canada
brx@cs.ubc.ca



Figure 1: VIBES allows audience members of a livestream to spatially interact directly with the streamed application through mouse events on the video player. In this example, viewers can spawn items into a game by first clicking on the item on the Twitch extension (a) and then clicking on the desired spawn location on the streamed game (b).

ABSTRACT

Livestreaming has rapidly become a popular online pastime, with real-time interaction between streamer and viewer being a key motivating feature. However, viewers have traditionally had limited opportunity to directly influence the streamed content; even when such interactions are possible, it has been reliant on text-based chat. We investigate the potential of spatial interaction on the livestreamed video content as a form of direct, real-time input for livestreamed applications. We developed VIBES, a flexible digital system that registers viewers' mouse interactions on the streamed video, i.e., clicks or movements, and transmits it directly into the streamed application. We used VIBES as a technology probe; first designing possible demonstrative interactions and using these interactions to explore streamers' perception of viewer influence and possible challenges and opportunities. We then deployed applications built using VIBES in two livestreams to explore its effects on audience engagement and investigate their relationships with

the stream, the streamer, and fellow audience members. The use of spatial interactions enhances engagement and participation and opens up new avenues for both streamer-viewer and viewer-viewer participation. We contextualize our findings around a broader understanding of motivations and engagement in livestreaming, and we propose design guidelines and extensions for future research.

CCS CONCEPTS

- Human-centered computing → Empirical studies in HCI.

KEYWORDS

livestreaming; interaction; video; mouse events

ACM Reference Format:

Michael Yin and Robert Xiao. 2025. VIBES: Exploring Viewer Spatial Interactions as Direct Input for Livestreamed Content. In *ACM International Conference on Interactive Media Experiences (IMX '25), June 3–6, 2025, Niterói, Brazil*. ACM, New York, NY, USA, 21 pages. <https://doi.org/10.1145/3706370.3727867>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IMX '25, June 3–6, 2025, Niterói, Brazil

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1391-0/2025/06
<https://doi.org/10.1145/3706370.3727867>

1 INTRODUCTION

Livestreaming (or simply **streaming**) is a massive media phenomenon in which streamers broadcast their activities in real-time to audience members around the world [74, 79]. Streaming platforms, such as Twitch and YouTube Live, provide the means of distribution for a streamer to share content [5, 88]. The appeal of livestreaming has been described as transforming “private play into public entertainment”, in which the streamer, the audience,

and their interactions all become part of the spectacle [79]. This idiosyncratic element of mixing real-time interactivity with performative elements has had a profound influence on academic research [50, 52, 74, 89]. Research into motivations behind livestreaming, from both viewer and streamer perspectives, highlights the importance of social interactions [29, 52, 55, 76, 92] as a unique form of motivation for engagement [79].

However, viewer-streamer interaction methods remain rather limited in practice, with the primary facilitating medium being text-based chat [14, 52, 68, 75, 78]. To address these constraints, streaming platforms have introduced additional interaction mechanisms such as subscriptions, gifting, or donation systems [9, 51]. Explorations have also looked into expanding the degree of influence of chat messages, for example, having direct effects on the streamed content itself as in *Twitch Plays Pokemon* [46, 63] or *Marmes on Stream* [62]. Recently, researchers have extended these ideas by investigating how non-textual forms of viewer interaction may affect the content of the stream, further broadening the scope of viewer participation and engagement [14, 68].

Building on this prior research, our exploratory work investigates the broad input method of **spatial interaction events**. To probe upon this concept, we scope our research specifically on mouse events (clicks and motion) on the streamed video as a direct input mechanism, leveraging existing elements within browser-based livestreams. Unlike text-based chat, mouse events can encode **direct spatial input** that maps onto the visual content of the livestream, enabling users to interact *through* the stream rather than *about* it. Our research aims to explore the following two research questions:

- **RQ1** – How does direct spatial input affect the streamer’s experience, and what challenges and opportunities emerge for streamers?
- **RQ2** – How do viewers perceive and engage with direct spatial input, and how does it influence their interactions within the stream?

To address these, we developed **VIBES** (Video-based Interactions for Broadcasted Events in Streaming), a lightweight, flexible prototype that collects and deciphers users’ mouse-based interaction events and transmits them to livestreamed applications. VIBES was a technology probe [36] – a simple, adaptable system that allowed us to gain insights into user and streamer experiences using such events. An initial formative exploration illuminated streamer perspectives on how audience spatial input could integrate with their motivations to stream and affect their streaming experiences. A subsequent user study involving the deployment of VIBES in livestreams highlighted practical insights and the effect of direct spatial input on the viewer and streamer engagement (addressing **RQ2**). Our findings extend prior research on livestreaming engagement and motivations, and we highlight challenges and present suggestions for future exploration.

2 RELATED WORKS

We consider prior related work to motivate the research gap, inform the development of our technology probe, and contextualize potential motivational and engagement impacts on streamers and viewers.

2.1 Motivations and Relationships – Viewer and Streamer Perspectives

We considered the foundational motivations for streaming and viewership. From a streamer perspective, Taylor argued that motivations for streaming stem from social connections – the transformation of play experiences into public performances [79]; Li et al. found that streamer motivations comprise reasons relating to social integration, personal integration, and affection [52]. Similar motivations have been found in studies regarding educational [12] and cultural streaming [56] – social engagement through connecting with viewers forms a major part of why streamers stream.

For viewers, streams satisfy desires for social contact and entertainment. The real-time aspect allows for personal, intimate connections with the streamer [53, 87] and other viewers [31, 56] – being alongside others in a shared community forms this social motivator [59, 79]. Li et al. [52] found that viewer identification and streamer relationship were core facets underlying the motivation for viewership; Hamilton et al. found that direct recognition by the streamer can be a rewarding experience for viewers [30]. Sjöblom and Hamari considered viewer motivations from a uses and gratifications perspective, finding that viewership was correlated to contextually translated human needs, such as acquiring information, enhancing connections, and escape [74]. From a viewer’s perspective, social engagement with the streamer and the other viewers extends the underlying entertainment factor as a major reason to watch streams.

Common in both streamer and viewer motivations is the high degree of focus on bi-directional sociality – as livestreaming is a real-time activity, it promotes a high degree of synchronous engagement [52, 99]. In the act of streaming, the audience and their interactions become integrated into the entire experience and become “part of the show” [79]. Smith et al. stated that the interaction between these two agents can be “entertaining and reciprocal” [75], and that viewers can help co-create content. Hamilton et al. expanded on how viewers want to interact with the stream, and streamers often give opportunities for viewers to do so beyond simply chatting, such as having them play games with the streamers [30]. Our work evaluates how providing users with a way to directly impact the streamed application using their visuospatial mouse interaction events affects this co-creation of entertainment through the lens of streamer and viewer engagement. We consider both how this affordance affects the present level of communication and engagement in the streamer-viewer relationships and how it might encourage new forms of social interaction.

2.2 Performative Audience Participation and Collaboration

Past research has investigated methods of audience participation within streaming. Interactive interactions with video-based media have long existed, such as collaborative annotation of online videos [15, 44]. Outside the digital realm, Webb et al. coined the term “distributed liveness”, which comprises the real-time relationships between performers and audiences [85]. The idea of having audience members directly influence real-time performance and artistry has been explored in many fields, such as theatre, [8], orchestra [67], VR experiences [33], and video meetings [80].

Livestreaming opens up new opportunities regarding distributed liveness, especially due to the asymmetric nature of audience participation at scale. One area in which this has been considered is in audience-participation games (APGs), which allow audience members to impact gameplay in significant ways [73]. One example of an APG is Twitch Plays Pokemon (TPP), in which audience members were given decentralized control over the button input of a livestreamed game [46]; this later spawned the formation of a community and surrounding narrative [63]. The impact of TPP has been well explored in academic literature, and findings suggest that it creates unique forms of social agency [73]. More generally, Striner et al. presented a theme map of audience participation in livestreaming based on student designers, where concepts such as streamer-viewer relationships, interaction, and agency were key themes [78].

APGs allow audience members to encapsulate individual and collective roles within the game itself, opening up new methods of interactions [73]. User integration into the streamed game affords engagement, but the medium of chat limits the ability for viewer expressiveness to flourish [46]. This indicates a possible research gap — what are ways to directly engage with the content in more expressive manners than text-based chat? We hark towards research comparing GUIs and CLIs. For GUIs, input to a device is passed through interaction events such as click or touch events, which encode spatial expression; certain implementations of such systems have had advantages over text-based CLIs in terms of user satisfiability and effectiveness [2, 77]. An analogous implementation for interactive livestreaming is to use similar events as input into the streamed content, which we investigate in this work. This can relate to theories drawn from psychophysiological foundations, understanding how different sensorimotor signals (such as vision and movement) can have different responses and outcomes, such as affective state [27, 64].

A related variant to APGs are danmaku-participation games (DPGs) [84], in which viewers can use danmaku as a form of control in a streamed game; these games have often been studied without the presence of an actual streamer [11, 84]. Danmaku is a popular audience interaction paradigm on Asian video sites in which people's textual comments float across the screen, often used for expression, entertainment, and social communication [54, 61, 98]. DPGs, like APGs, provide audience members with control over the streamed application through their danmaku input, as well as offer similar motivations, such as engagement and connection, with similar challenges, such as latency and instability [84].

Persistent obstacles for APGs include latency, screen real-estate, and player agency [26]. These are all aspects we consider in our probe. Regarding agency — unlike other systems such as StreamSketch [58], which allow user input on the video to be visualized for the streamer but puts the onus of decision-making with the streamer; we study the fundamental transfer of agency towards the audience, as their inputs can influence the application as much as the streamer's inputs. Regarding latency, we directly address this issue of latency in development, and regarding screen real-estate, we re-purpose components that already exist in livestreams. Ultimately, in our work, we consider how visual input through direct spatial input capturing interactive events modulates existing learnings on audience participation and interaction within livestreaming.

2.3 Beyond Just Chatting – Livestreaming Input Methods

Researchers and developers have built systems that extend the prevalent paradigm of chat interaction on livestreams and video content. Some of these systems are integrated directly into the platform, such as polling or gifting systems [9, 51]. Others are more exploratory, such as using audience physiological markers [35, 68], voice messages [4], or haptic feedback [41] to affect content. For instance, Huang et al. studied the use of viewer frissons to affect video content in increasing the social affective dimension [35]; Robinson et al. studied how using audience heartbeat to influence gameplay resulted in a positive boon to engagement [68] — our study follows a similar evaluation as the latter to assess the impact of VIBES. Furthermore, Fanzo et al. built a game that asymmetrically allows users to view different camera channels within a game [24], and Lessel et al. performed case studies on an audience communication extension that allows polling and tracking in Hearthstone [48] and aggregation of move-selection for chess [47].

Hammad et al. developed MARS, which augments the stream's video player with a context-aware visualization of game metadata to support novel game designs and provide added information to the viewers [32]; however, it does not necessarily support input from the viewers in affecting the stream itself. Yang et al. developed Snapstream, which allows users to create snapshots of the stream and share these annotated snapshots [93]. The use of interactive visual elements based on the streamed content was found to increase engagement and communicative expression. The importance of visual modalities in creating new affordances was also a key finding in Lu et al.'s StreamSketch, which allows viewers to offer their drawing suggestions through interaction on the stream's video player [58]. Chung et al. developed VisPoll, a polling system using visual communication that can be aggregated by the streamer and summarized for the viewers, and found that the visual input led to increased expressiveness for viewers [14]. From all these works, we learn that visual input extends the limited expressiveness in text-based APGs [46], something that our work also investigates and designs for. This is further corroborated through research that has shown that interactive visual-based inputs can produce spontaneous and playful interactions [40] and increased reflection and understanding [57].

We also consider some commercial alternatives that have also been developed to help integrate novel interactions into specific games. Software systems like CrowdControl¹, Streamer.bot², and Dixper³ offer unique integrations into games. Whereas these applications do offer direct spatial input through audience interaction, their impact on audience-streamer motivations and relationships has not been heavily explored in academic research.

All in all, many of these prior works, e.g., StreamSketch [58], Vispoll [14], and Lessel's Hearthstone work [48] take mouse events primarily to *visualize* the aggregated data — inputs are used to communicate *about* the stream, but it remains the onus of the streamer to decide how to use this information. We are the first to investigate using such inputs directly *into* streamer application,

¹<https://crowdcontrol.live/> [Last accessed: March 12, 2025]

²<https://streamer.bot/> [Last accessed: March 12, 2025]

³<https://dixper.gg/landing> [Last accessed: March 12, 2025]

providing the audience with direct influence over the streamed content and giving rise to some level of uncontrollability for the streamer.

3 DEVELOPING THE TECHNOLOGY PROBE – BUILDING VIBES

This section summarizes the implementation and features of VIBES. Although VIBES was used primarily as a means of exploring our research questions, there were several key design challenges that we came across, outlined here.

3.1 Design Decisions, Goals, and Objectives

We motivate our decisions of platform and medium, which helped in focusing our exploratory probe. We chose Twitch as our platform for its readily-accessible and extensible API and its prominence in prior livestreaming literature. We chose to develop for desktop-based systems as a medium because of technical restrictions on extending the Twitch experience to mobile devices. While it is relatively easy to integrate 3rd-party features into web browser applications, it is much more difficult on a mobile application; at the time, the mobile version of Twitch also lacked several key features necessary for our desired interactive system. In developing a robust system to translate spatial interaction events, we focused on three primary goals:

- **Functionality** — developing a system that would be able to transmit real-time mouse events on the video player to the backend without error. This was crucial for ensuring reliable interaction in livestreaming environments, where synchronization is expected.
- **Adaptability** — building a system that could be extensible to a variety of applications to support the broad use cases in interactive livestreams, given its use as a technology probe [36].
- **Expressiveness** — extending simple mouse events into complex actions, enhancing the depth and richness of interaction that viewers could use to directly impact the streamed content.

3.2 Architecture and Implementation

The system architecture of VIBES consists of 4 main components (Figure 2).

- A **browser extension** running on the viewer’s computer captures the type (either a click or a gesture) and coordinates of the mouse event, as well as the viewer’s username and the stream latency.
- An optional **Twitch extension** provides additional data packaged with the mouse event (e.g., a clicked shape) along with the associated username. This is hosted on the Twitch servers and can be activated by the streamer, varying depending on the streamed application.
- Both these components send data to the **websocket server**, which is hosted in the cloud.
- The server then passes the data to the **streamer application** running on the streamer’s computer, which decodes the data

to perform an action in the application (e.g., spawn an item at a selected coordinate).

The streamer’s application can also send data back to the viewer’s client; this can be used to communicate changes to the Twitch extension. VIBES is our implementation of a system that translates spatial mouse events into input for livestreaming applications. It is a lightweight, flexible system that can adapt based on the specific streamed content and the desired use of inputs. We discuss more detailed information for each component in the following sections.

3.2.1 Collecting Mouse Events from Users. VIBES collects mouse events from a viewer using a **browser extension** that activates when viewing a Twitch stream. Spatial interaction events on the browser can be dimensionalized through clicking and movement, leading us to develop two main events as the vocabulary for our exploratory study — (1) mouse clicks — an instantaneous click on the video stream, and (2) mouse gestures — a click-and-drag motion on the video stream. This extension retrieves data regarding 1) the type of mouse event performed and 2) its spatial information relative to the video player. To get the spatial location of the mouse click and motion, we localized and normalized the mouse position relative to the video player element, obtained through query selections on the DOM. To differentiate a gesture from a singular click, we set a timeout and motion threshold starting from when the user presses the mouse down. DOM manipulation was used to automatically retrieve the broadcast latency of the stream from the Twitch video stats panel (i.e., pipeline latency associated with streaming real-time content [1]). Once these values were obtained, they were used to generate a JSON object, which was then sent to the **websocket** hosted on a cloud server.

3.2.2 Adding Flavour to Events through Twitch Extensions. Having only mouse events as input is rather limiting; we used **Twitch extensions** to generate flexibility and expressiveness by allowing users to augment their mouse events with contextual data. These extensions are inline frames (**iframes**) that can be added to a stream and are hosted on Twitch servers. Viewers can pass data through forms, button clicks, etc., following the usual suite of web interactions. This data, along with the username of the viewer, is collected and sent to the **websocket** and then to the **streamer’s application**.

The streamer application then stores this extension information using the viewer’s username as a key for a hashed key-value pair. Once a mouse event by the same user is read by the application, it retrieves the hashed value, and that user’s data can be used in the application logic to generate interesting outcomes beyond simply spatial coordinate data. For example, a viewer could first click on the overlay iframe to select an item and then spawn the item into the application at a specified location by then clicking on the video player.

3.2.3 Passing Data back from Applications. To adapt VIBES to diverse stream applications, we provided the ability for the streamer application to pass data back to the viewer (the Twitch extension) so that there could be bidirectional interaction — this allows the extensions described in the prior section to adapt towards the streamed application. For example, there may be times in which the streamer applications need a change in the level of expressiveness, to lock out viewers from providing spatial interactions, and so forth. Thus, data

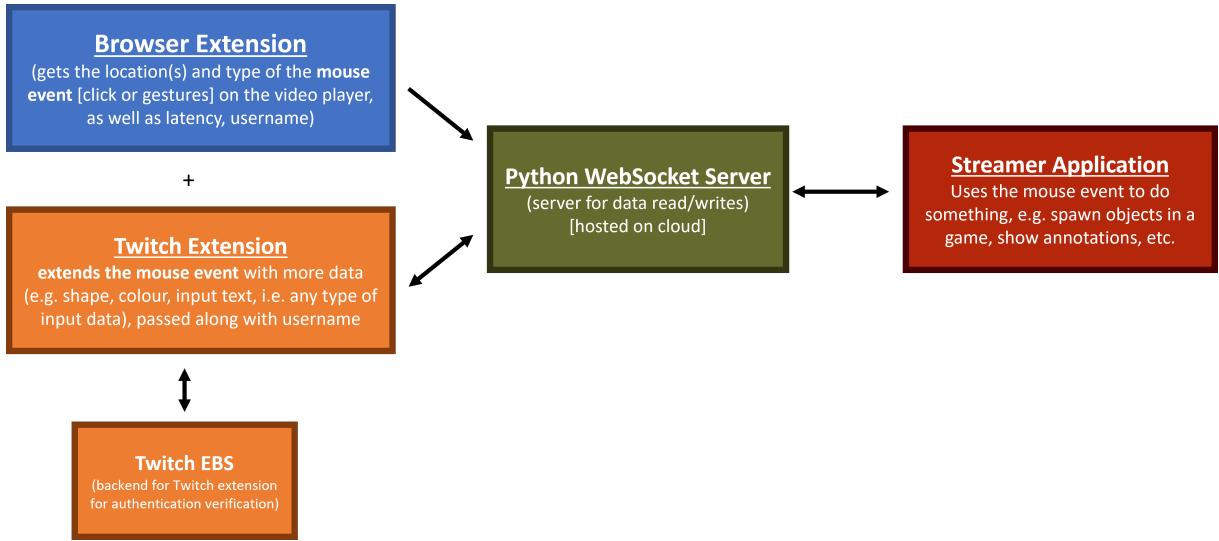


Figure 2: Simplified architecture diagram of VIBES. The browser extension captures a viewer’s mouse input; the Twitch extension provides added expressiveness to the spatial data. This data is passed to the websocket server, which transmits it to the streamer’s application. The bi-directional nature of the architecture allows the application to send data back to influence the Twitch extension.

can be passed back from the streamer application to the websocket and later propagated to the client Twitch extensions.

3.2.4 Dealing with Broadcast Latency. One major challenge that immediately became apparent was that of broadcast latency, the inevitable delay in the stream as it is presented to viewers [95]. Even at ultra-low latency settings, there exists a lag between 0.2 and 2 seconds between the streamer’s and viewers’ perspective [1]; this exceeds the delay for interaction instantaneity that users may expect [18]. During this loading period, the viewer receives feedback in the form of a loading circle whose countdown animation time is equal to the broadcast latency obtained from the Twitch video stats panel. Through informal testing over a 5-minute stream, we found that this latency variable differs from the screen-to-screen latency by a mean of 233 ms (s.d. = 66 ms), which is adequate for our usage. The major advantage of using this scraped latency variable is that it can be automatically retrieved and updated for each viewer, eliminating the need for manual clock calibration of timestamps as in prior systems, e.g., Helpstone [48]. Although there were other sources of delay as well (e.g., web requests), this was largely negligible in comparison.

Another issue with broadcast delay involves matching user action with intent. When a user performs a mouse event, they perform it on the video player that is showing the application at a time in the past relative to the streamer. However, the input is received by the application near-instantaneously. If the application’s appearance changes within the broadcast delay period (e.g., if the application’s camera moves), this will cause a dissonance in intent — should the user mouse event be registered in regards to the current camera or the past camera of the application? In our system, we chose the latter, as we believed that the latter would more closely align with viewer intent, i.e. the viewer is interacting with the stream *as they*

presently see it rather than the streamer’s live state. As the stream from the viewer’s perspective happens in the past from the present stream, this implies we must keep a record of what happened in the past to reconcile this latency. In our applications, we considered that the most common aspect that may change in a single streamed application is the “camera” (for games, but more specifically the “viewport”).

To address this, we implemented a buffer - a time-shifted record of the application’s visible interface (e.g. game camera position, scroll location, etc.). In our demo systems, we retained the past 10 seconds of data (given the assumption that latency would not surpass this period) and updated every 0.1 seconds. Thus, when the viewer triggers a mouse event, the system 1) notes the broadcast latency for that user (which is already logged from the browser extension), 2) retrieves the historical camera state from the buffer, and 3) applies the input relative to that camera state. Currently, the implementation of this optional camera buffer is the responsibility of the application developer; however, in the future, it could be abstracted directly into the system to simplify integration for applications with dynamic, moving interfaces. Ultimately, broadcast latency was a major identified issue in development — we acknowledge that viewers interact with a time-shifted perspective of the stream. However, we aimed to design for it rather than overlook the issue.

4 DEMO INTERACTIONS

With VIBES built, we brainstormed potential ways in which direct spatial input could be incorporated into stream applications, taking the perspective of the researchers as viewers. We built out these samples using VIBES, forming lightweight, illustrative interactions. One limitation of these demo interactions is that they were created and tested by the researchers in-house, with researchers acting as

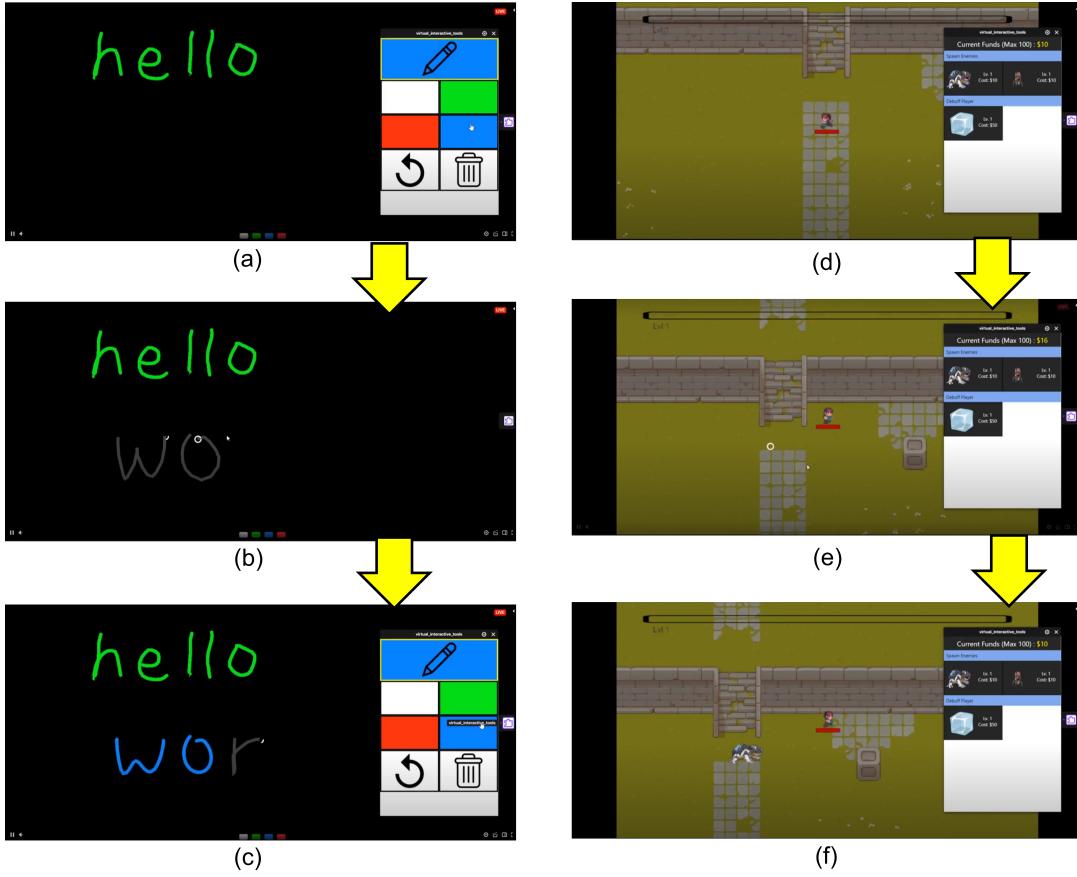


Figure 3: The left is an application for drawing on a shared canvas. In (a), the viewers use the Twitch extension to select colour, undo, or delete. In (b), the user uses their mouse to draw the letters “w” and “o”. Although small, note the spinner circles used to denote latency. In (c), the user’s drawing has been passed to the Unity application, where it renders the line with the correct colour onto the application. The right is an interaction for spawning an enemy into a game. In (d) viewers can select which enemy or debuff to apply. In (e), viewers click into the game where they want the enemy to spawn and in (f), the viewers see the enemy spawned.

both the individual streamer and the viewers (i.e., using multiple accounts). This entails a very small number of viewers; scaling to a larger audience is important for future research.

4.1 Sharing an Experience Together – Participating With and Against the Streamer

Viewers and streamers can participate in a single application using VIBES (Figure 3). In the first interaction, users can collaboratively draw together with the streamer on a shared canvas. Users can use the associated Twitch extension to select a colour before making the drawing on the screen. The coordinates are sent to the Unity backend, which processes them and renders the drawn lines with the appropriate colour. Viewers can furthermore use the Twitch extension to undo their prior input or to clear all of their inputs.

In the second interaction, viewers can play against the streamer in a top-down 2D shooter game. In this game, viewers can spawn enemies by using the extension to click the desired enemy and then the video player at the desired spawn location. This game also

illustrates a way in which the application data can be passed back – when the streamer defeats a certain number of enemies, they level up, and this information is communicated to the Twitch extension.

4.2 Polling the Audience, But with Spatial Information

Streamers can use VIBES to survey their viewers through spatially-dependent polling, which we define as polls where responses are dependent on position or actions within the visual space, contrasting text-based polls (Figure 4). The first interaction polls for spatial clicks in a game of tic-tac-toe. On the viewers’ turn, they can vote (click) on which square they would like to place their symbol in during the duration of a countdown. Each square’s value increments upwards from zero as users click to vote. Under the hood, the spatial location of the click is projected as a coordinate onto the game plane (corresponding with a square if the user clicks correctly) and the user is registered to have voted. When the timer reaches zero,

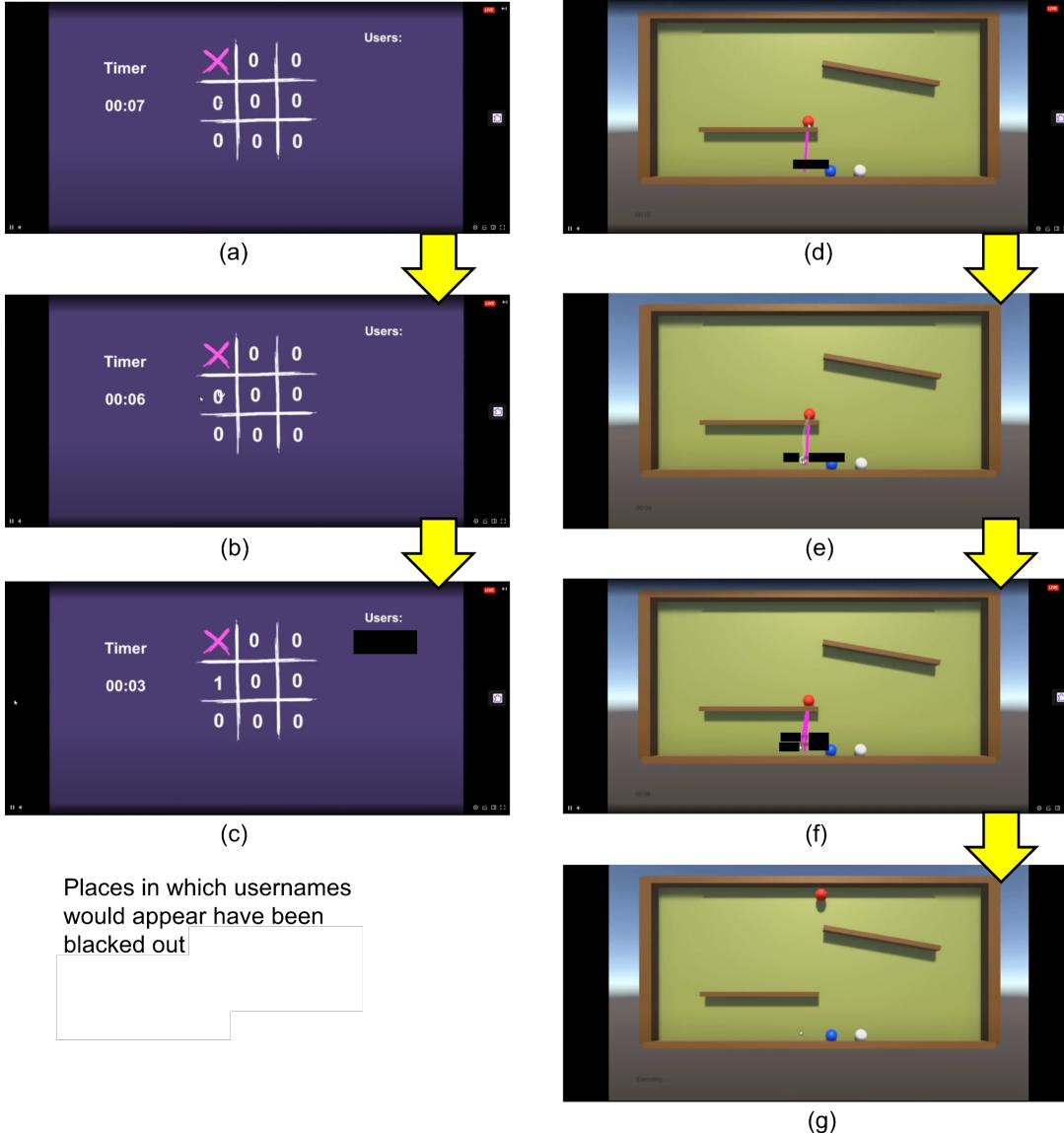


Figure 4: The left interaction surveys clicks. (a) illustrates the initial scene, where viewers are prompted to vote on a move. In (b), the viewer clicks to vote for their selection (the middle-left square); in (c), the number at this square increments and the username of the viewer is added to a list on the right of the stream. The right interaction surveys gestures. (d) illustrates an initial scene, where a different viewer has already “voted” for a force. In (e), the viewer drags from the ball to add another force, which is then represented as a line in (f). Once the timer has completed, the average of the forces is applied, as seen in (g).

the application places the appropriate symbol onto the most-voted square.

The second interaction polls for spatial gestures in a ball-flicking game. Viewers can draw a line on the video screen by clicking on a ball on the screen and dragging their mouse to control the direction and force of the flick during a countdown timer. This becomes a “primer” for a force, which is illustrated as a line on the application screen. When the timer reaches zero, these forces can be aggregated and applied. On the backend, the application keeps track of the

various mouse gestures and uses that information to determine which ball is affected and how the force should be applied, showing the result on the screen.

4.3 Interactive Controls and Annotations to Extend Stream Content

VIBES allows viewers to interact with streamer software directly through gestural controls or annotations. For instance, we demonstrate how viewers can use mouse gestures to control the user’s

music application. A user's mouse movement is read by a backend Python application on the streamer's computer and then translated through a simple gesture recognition system (Figure 5). The application then makes a call to the Spotify API, which then changes to the next or previous song, depending on the gesture. Ultimately, this integration pattern could potentially apply to any number of applications that the streamer uses, as long as they have a suitable API.



Figure 5: VIBES can be used to input mouse gestures as control. For example, we developed a simple arrow gesture decoder that can be used to control the streamer's Spotify application.

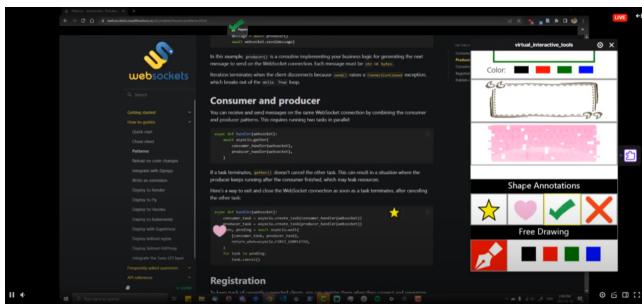


Figure 6: VIBES can be used to annotate on the streamer's screen. On the stream, we see a heart, checkmark, and star icon that has been spawned in. This tool can also be used to spawn textboxes with viewer-specified text or to freehand draw by the viewers.

Drawing inspiration from functions in online meeting software like Zoom but at a much broader audience scale, VIBES can facilitate annotations on the streamer's screen (Figure 6). We showcase a suite of possible annotations for viewers — text boxes, shapes, and freehand drawing. Users can select the desired annotation using the Twitch extension and click on the stream where they want the annotation to appear. An Electron web application renders the annotation onto the streamer's screen, which is seen by all viewers.

5 FORMATIVE EXPLORATION WITH STREAMERS

With a prototype of VIBES and some illustrative interactions, we conducted an initial formative exploration with streamers to probe into how this technology might affect viewer interactions and motivations, how it could be implemented into practical streaming contexts, and what the possible future improvements could be, addressing **RQ1**.

5.1 Participant Recruitment

We recruited participants who had a history of active streaming. Although VIBES was built as a Twitch integration, we were open to recruiting participants from a variety of streaming platforms, such as YouTube Live, Facebook Live, or even Zoom, as we were primarily focused on the technology-agnostic overarching concept of extending livestreams with mouse events rather than any specific technical implementation. We used a variety of recruitment channels, including personal connections, the institute's paid studies listings, and posting on the /r/Twitch subreddit.

We were able to recruit 6 streamers for this study. The ages of our participants leaned young, ranging from 19 to 28 (mean: 23.5), with a gender distribution of 2 male, 3 female, and 1 non-binary. Although our sample size was rather small, we felt that it was sufficient for a formative exploration in terms of capturing meaning-richness and that it furthermore represented a diverse range of streaming practices — most participants indicated that they streamed video games, but some also brought perspectives from educational streaming (e.g., online teaching and collaboration) or music-related streaming. A more detailed representation of the participants can be found in Table 1.

5.2 Study Protocol

Before the study, participants were asked to review and sign a consent form, outlining data collection and ethics approval; audio recording was collected with consent. The study was done through remote audio calls over Zoom and was split into two sections. The introductory part was a semi-structured interview focused on the streamer's stream activities and motivations, from which we explored what, how, and why participants stream, as well as their interactions with their viewers. Examples of questions asked in this section were *"What are ways that you currently engage with the audience?"* and *"What are ways in which the audience can communicate with you?"*. These questions aimed to illuminate the streamers' present motivations for streaming and help highlight possible ways VIBES can impact such motivations.

The second part of the study focused on addressing **RQ1** through our technology probe. To get participants familiar with VIBES, we presented demonstrative videos from the applications described prior, plus an additional one to show the effects of latency. Although videos do not fully capture the impression of the system compared to a full demo, they were convenient for a formative exploration — setting up the actual interactions takes significantly more time and effort on the participant's end. While watching the videos, we asked the streamers to freely talk about their thoughts before later delving into a discussion focusing on how direct spatial input might be integrated for unique streamer interactions, as well as the

Table 1: Summary of Interview Participants

ID	Age	Gender	Avg. Viewers	Hours Streamed / Week	Streaming Platforms	Primary Streaming Content
P1	22	F	10-50	5-10	Twitch	Games
P2	21	NB	0-10	10+	Twitch, Tiktok	Games
P3	28	F	0-10	0-2	Zoom, Twitch, Discord, Teams	Games, Education
P4	25	M	0-10	0-2	Twitch, YouTube Live	Games, Music
P5	26	F	200+	0-2	YouTube Live, Facebook Live, Zoom	Music, Performing Arts
P6	19	M	0-10	0-2	YouTube Live, Zoom, Google Meet	Education

limitations and challenges presented by the idea. Studies generally took around an hour, and participants were compensated with \$16 CAD for their participation.

5.3 Data Analysis

The collected qualitative data was examined and analyzed through a content analysis approach [23, 34]. After familiarizing ourselves with the data, the qualitative texts were then coded by the primary researcher [71], with feedback and discussion from the other researchers to attenuate single-coder bias; nonetheless, we note that coder bias could be a possible limitation of our study. These initial codes represented the base, semantic content of the textual data. Afterward, the codes were iteratively refined and related codes were grouped into broader categories. The final codebook consisted of 24 codes and 6 categories; some examples of the final codes were “Motivations for Interactivity” and “Potential Applications”; some of the final categories were “Engagement and Interactivity” and “Experience and Intentions”. The process of content analysis helped us garner a holistic understanding of how streamers perceive and might integrate direct spatial input into their practices, helping us answer **RQ1**.

6 FINDINGS FROM THE FORMATIVE EXPLORATION

6.1 Streamer Perspective

RQ1 – How does direct spatial input affect the streamer’s experience, and what challenges and opportunities emerge for streamers?

6.1.1 Extending Audience Engagement, Which is Important for Streamers. All streamers were overall positive and receptive to VIBES as a way of extending audience interaction. Although they indicated that the level of engagement might differ depending on the application they were streaming, all participants emphasized the importance of audience engagement and interaction as a reason and motivation for streaming, agreeing with prior research [52, 56, 79, 87].

Although engagement was usually indicated to be done through chat, streamers pondered on potential applications of direct spatial input in their livestreamed content. Through their responses, we identified two main distinguishing features of such input that allow for potentially novel interactions — **visual spatial encoding** — the ability to encode visual information, and **continuous inputs** — the ability to provide gestural or movement-based input. Regarding the former, P2 mentioned that this tool can allow the audience to help them in games by providing visual information in a game — “*I got lost a lot... I would love this. People have been pointing where I have to go*”. This is expanded by P1 — “*you can actually draw the direction that you want the player to go to*”, which P3 expands as “*360-degree movement; there’s so many possibilities of where someone could draw the gesture*”.

Altogether, the streamers indicated that VIBES could promote a more collaborative and interactive environment between streamer and audience, e.g., “*I think this will just add more options for viewers and streamers [to control] how much interactivity they want*” (P4). P1 hypothesizes a collaborative drawing tool that could provide “*more freedom to the viewer*”. VIBES could also create a more continuous and seamless stream experience. The streamers indicated that they had to actively keep up with chat, but “*if you’re able to kind of simulate that with gestures instead on a screen... [it] actually reduces how often you, as a streamer, need to look at a message*” (P3), and, regarding the polling interactions, “*this one is actually pretty cool for like choice-based games ... because sending up a poll, it just takes too long*” (P2).

6.1.2 Impacting the Balance between Streamer and Viewer Intention and Control. Streamers noted that direct spatial input could complicate their intentional control over the streamed content. The content of a livestream is typically under the full authority of the streamer, yet providing users with the possibility for direct input alters the balance of agency. As such, the stream is then influenced by viewer intention, and streamers discussed the challenge of matching viewer intention and desires to the streamer’s desired experience. For example, P1 mentioned that if viewers continually

use the system to spawn enemies behind a streamer, “[the streamer] won’t even know it... I think, after a bit of the streamer getting killed every two seconds spawning in and dying, they’re probably gonna get tired of it”, highlighting a possible mismatch between viewer and streamer desire. P3 stressed the importance of matching viewer intention and streamer expectations, suggesting increased streamer control to help manage this dynamic, e.g., “You can add a setting from the streamer because then it’s based on the streamer’s comfort level” (P3).

Streamers expressed concerns over matching intentions given the inevitable latency in livestreaming. As the streamer content is slightly ahead of what the viewer sees, scenarios in which the intentions might mismatch may be common, especially if the broadcast camera shifts during that time. Preliminary ideas were suggested to address this issue, e.g., “There should be some kind of notification [if the viewer makes an annotation]” (P3) and “I think different sound effects, different colours” (P4), recognizing that “any application where the world moves around is going to be difficult” (P4).

6.1.3 Raising Possible Issues Related to Usage and Moderation at Scale. More speculatively, streamers noted that practical scaling to a large number of participants could pose a challenge – if everyone could pass direct spatial input at all times, this would likely inevitably lead to “a really big mess” (P2). One proposed solution to deal with a massive amount of input data was filtering by who can provide inputs, e.g., P2 indicated that you can make the tool available for “only subscribers, or only VIPs, or only mods”. Tying interactions to points and rewards on the livestreaming platform was also a popular suggestion, e.g., “you can leave a note on my screen for points or bits” (P2), as was adding a timeout or cooldown for interactions, e.g., “it goes into a global cooldown” (P2).

Moderation at scale was another speculated challenge. With more viewers comes more trolls – unwanted agents that disrupt the stream or harass the streamer, and direct spatial input adds another way of accomplishing this. One way of dealing with these unwanted agents, in addition to existing methods like timeout or banning⁴, may be through input moderation at scale. P3 brought up a scenario in which a streamer allows free drawing on their application, and a viewer chooses to draw something inappropriate. P3, P4, and P6 indicated that detection algorithms could be used to filter out these inputs, “train something to detect an image of like a swastika and then remove that” (P3). Reactive mechanisms for moderation (such as individual bans or restricting the tool to specific individuals) could also be made available by tying interactions to usernames. We highlight that issues regarding scale were more hypothetical in this context, as our work focussed on smaller, controlled audiences. Still, these considerations inform the importance of robustness and required safeguards for a real-world deployment.

7 PRELIMINARY APPLICATION INVESTIGATION

The prior formative exploration focused on the abstract potential of direct spatial input on livestreamed content; to attain more practical insights, we investigated its deployment under controlled livestreaming environments. We recruited two participants from

⁴https://help.twitch.tv/s/article/how-to-manage-harassment-in-chat?language=en_US [Last accessed: March 12, 2025]

the first exploration (P1 and P4), and based on their input and feedback, we developed an individualized application for each of them using VIBES. We then ran these applications in 2 livestreams with recruited viewers. Our main goal of this investigation was to address **RQ2** – looking to investigate how extending actual livestreams with mouse events might affect the *viewers* in terms of metrics of motivation, including participation, engagement, and enjoyment; but we also extended our findings on **RQ1** through assessing the experience of the involved streamers.

7.1 Application Development

A brief initial meeting (around 15-20 minutes) was conducted to collect information on the type of application they envisioned for facilitating spatial input. The streamers were kept informed regarding progress and provided feedback regarding potential changes. P1 described the development of a mod to an existing game (*Terraria*) in which viewers could collaboratively interact with the streamer; P4 proposed a tower-defence-based game in which viewers could help the streamer through spawning traps and picking upgrades. We developed these applications, respectively known as *Terraria Interaction Mod* and *Storm the Village*, and describe them in brief below (detailed information can be found in the supplemental material). Streamers were compensated \$24 CAD for their help in designing the application.

7.1.1 Terraria Interaction Mod. *Terraria* is a bestselling sandbox-based adventure game in which a player can explore, build, interact with NPCs, etc., in a procedurally-generated world [65]. Our mod (Figure 1) does not change the overarching mechanics of the game but provides viewers with three points of interaction:

- **Spawning Items and Enemies** – Audience members can select an item or enemy from the Twitch extension and spawn this selection at a cursor location within the game (there is a minimum distance at which enemies must be spawned). To limit the number of spawned objects, there is a funding system that provides funds at a rate inversely proportional to the number of viewers.
- **Writing a Message** – Audience members can write a message in the Twitch extension in an input field and then click into the game to spawn that message into the game for a few seconds at the clicked location.
- **Voting on NPC appearance** – During the night phases of the game, audience members can vote for which NPC they want to arrive the next day by clicking on the voting icons on the screen.

7.1.2 Storm the Village. *Storm the Village* is a roguelike tower-defence game in which the player must repel waves of enemies from a castle by shooting them with their cursor, inspired by a similar game “Storm the House” [69]. *Storm the Village* was developed in Unity by the research team. The game operates on a day system – during the day, the viewer shoots enemies that gradually become stronger. After each day, there is a vote for an upgrade – these can provide new guns, increased gun damage, etc. To win, the player must repel 8 days’ worth of enemies before the enemies manage to deplete the health of the castle. There are three points of interaction for the audience members (Figure 7):

- **Spawning Traps** — Viewers can select a trap from the Twitch extension and click onto the video player to spawn it at the clicked location. Traps deal damage to enemies and apply status effects, and more traps are unlocked as the game progresses. To limit the traps spawned, there is a funding system — users must spend in-game currency (that increments at a constant rate) to build a trap.
- **Voting on Upgrades** — At the end of each day, audience members can vote for which upgrade they want the streamer to have by clicking the option directly on the stream.
- **Freehand Drawing** — There is a sequence in the post-game in which the audience members can draw something on the video screen.

7.2 Participant Recruitment

We had initially intended for the streamer involved in the development to stream the corresponding application in the live session deployment. However, P1, who was involved in the development process of the *Terraria* mod, was ultimately unable to attend the livestream session. Thus, we recruited a new streamer (P7 — gender: female, age: 20) who had prior experience with both livestreaming and the *Terraria* game; as such, we found them suitable for streaming the *Terraria Interaction Mod*. For the other stream (*Storm the Village*), P4 was able to attend and stream the application.

For P4's stream (Session 1), we were able to recruit 16 viewers (average age: 25.3, std: 3.91, ranging from 19 to 33; 13M, 3F) and for P7's stream (Session 2) we were able to recruit 18 viewers (average age: 23.7, std: 4.35, ranging from 19 to 38; 14M, 3F, 1 Preferred not to report). For standardization, the participating viewers were not recruited as prior viewers of the assigned streamer; furthermore, there was no overlap between recruited viewers across the two streams. The eligibility criteria for recruitment were rather general in terms of previous livestreaming experience, aiming to develop broader results regarding the use of direct spatial input for those at any point on the spectrum of livestreaming familiarity. Participants were recruited through a mix of voluntary responses from our institute's paid studies listings and convenience sampling.

Before the study, streamers and viewers alike were asked to read and sign a consent form relating to data collection and usage. Each of the recruited viewers was asked to install the browser extension and to provide their Twitch username for access to the Twitch extension. In addition, we provided each of the viewers with a document with information regarding the stream that they would be watching, outlining the game and the various interactions they could engage in. Viewers will henceforth be denoted by combining the participant ID of the streamer they watched with a viewer ID, e.g., P4V3 will refer to viewer ID 3 that watched P4's stream.

7.3 Study Protocol

The livestreams took place on the Twitch platform. Streamers were asked to stream on a researcher-created account, providing us with control over the distribution of the Twitch extension and the viewership. During the livestream session, the researchers were largely hands-off, letting the streamer take control over their style of streaming, their interactions with viewers, and the overall atmosphere. The streamers played the customized games, and the

viewers were able to interact with the streamer using direct spatial input as well as normally through chat. We recorded the stream locally and saved a private video-on-demand (VOD) of the stream and a transcription of the messages sent in chat. Overall, the stream sessions took about 45 minutes.

After the stream, viewers were asked to answer a survey regarding their experience. This survey combined two scales — we first employed the questions from Chung et al.'s study to evaluate implementation factors such as the usefulness and helpfulness of the tool in achieving interaction with the streamer and other viewers (including open-ended questions for expanding on their rationales) [14], then asked participants to fill out the *Audience Experience Questionnaire (AEQ)* (a questionnaire developed to assess audience members in social video gaming sessions) to evaluate on the subscales of enjoyment, mood, game engagement, social engagement, and participation within the study [21]. We also performed a brief exit study with the streamer. Streamers were first asked to fill out the *Game-Specific Attribution Questionnaire (GSAQ)* to gauge their attribution of the events of the game; we only used internality and controllability subscales to assess how in control the streamer felt over their experience [19]. We briefly interviewed the two streamers about their experience using VIBES, what they thought could be improved, and how it affected their perceived engagement with viewers. Viewers and streamers were compensated \$24 for their participation in the livestreaming session.

7.4 Data Analysis

Our data consisted of both quantitative and qualitative data, which we analyzed using a mixed-methods content analysis approach [23]. The quantitative data (Likert-scale questions) was investigated through exploratory data analysis processes. Due to the exploratory nature of the study and the uncontrolled nature of Twitch streams, we refrain from presenting formal statistical tests [82] and use them more as illustrative values. The graphs are presented separately for each study session due to the uniqueness of each streamed game and its associated interactions.

For the qualitative data, we first performed initial open coding on the questionnaire responses to understand the impact of direct spatial input on viewers [71]. We then refined, iterated, and grouped these codes into broader categories. Although the qualitative data was coded by the primary researcher, the analysis underwent discussion and feedback from the other researchers. For the exit interviews with streamers, we also underwent a similar coding process. In both cases, we took a largely deductive approach, contextualized by our formative exploration findings and background research. Overall, we developed 9 total codes (e.g., “Latency”, “Agency in the Game”) spanning 4 broader categories (“Challenges”, “Engagement and Agency”, “Enjoyment and Helpfulness”, “Interaction with Agents”) that tie towards both streamer and viewer experience, we developed into themes.

We integrated the exploratory analysis of the quantitative data with the qualitative data to highlight similarities and potential disparities in the interpreted findings. Since the data sources were generally aligned in terms of their conclusions, quantitative evidence largely supports our qualitative findings. Our results primarily focus on the experience of the viewers due to the larger sample size,

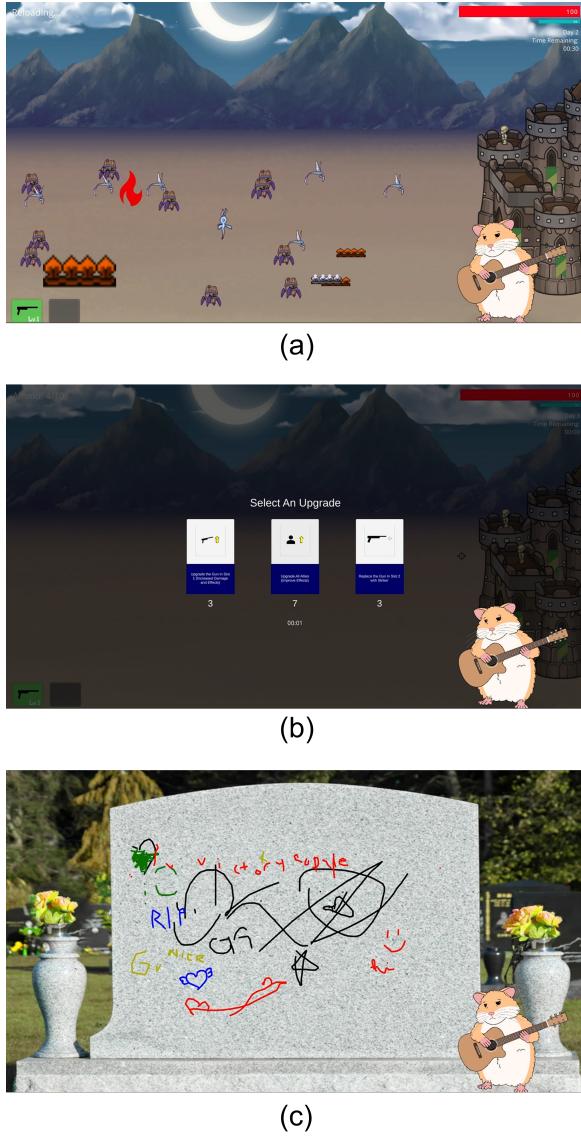


Figure 7: Images from the livestream of P4's study (Session 1) involving the game “Storm the Village”. In (a), we see that viewers have spawned in traps to help the streamer repel enemies. In (b), viewers vote on the upgrade that they want. In (c), viewers have a chance to sign off or write something on a memorial after they have successfully beaten the game.

but we do provide brief deductive findings based on the streamers' experiences as well, extending our formative exploration findings in practice.

8 FINDINGS FROM THE APPLICATION INVESTIGATION

8.1 Audience Perspective

RQ2 — How do viewers perceive and engage with direct spatial input, and how does it influence their interactions within the stream?

8.1.1 Generating Positive Feelings of Enjoyment and Engagement. Direct spatial input was generally well-received by viewers from both an implementation and conceptual standpoint, evidenced by the responses to the questionnaire questions (Figures 9 and 10) and viewers' open-ended responses. From the AEQ (Figure 8), participants scored highly on aspects of enjoyment (Session 1: mean = 5.5, Session 2: mean = 6.1), and mood (Session 1: mean = 5.5, Session 2: mean = 6.0). These metrics were supported through participants' responses; for instance, *“I felt like I was playing a game with friends which made my experience of the stream very fun”* (P4V5) and *“it is a fun way to have creative interactions between the viewers and streamers”* (P7V11). Furthermore, almost all participants agreed that they would like to see other streamers use such tools within livestreaming (Q10 for Figures 9 and 10), *“it would be amazing to see it integrated into a variety of games to change the way streaming is conducted and bring the viewer closer to the streamer”* (P4V14).

The use of VIBES to pass input positively affected viewer engagement and involvement for all participants, with every participant agreeing or strongly agreeing with the statement that they felt more engaged with the stream using the tool (Q3 for Figures 9 and 10). The responses to the AEQ support this sentiment, as the metrics of Social Engagement (Session 1: mean = 5.3, Session 2: mean = 5.8) and Participation (Session 1: mean = 6.1, Session 2: mean = 6.3) both scored highly across both study sessions. The active participation of the viewers within this stream was a contributing factor towards their enjoyment and engagement, e.g., *“it helped me participate and actually play the game as well. So like I was able to have [a] direct effect on the game”* (P4V10) as was the real-time visual nature of the tool — *“I could see first hand the implications of my actions on the game ... it was really really interesting to see in real-time how my actions were affecting the game”* (P7V12). In particular, the idea of essentially co-playing a game in a significant way led people to feel that their actions were impactful — *“It felt like I was actually impacting the gameplay”* (P4V13). This level of engagement went beyond the traditional chat interaction — *“the interactive tools helped me engage more in the overall experience than Twitch chat”* (P4V5) — perhaps due to the visual nature of seeing the results of one's interactions, e.g., *“being able to see the result of my interaction made it more interesting”* (P4V1). Overall, VIBES helped provide an increased degree of participatory agency for viewers by allowing them to influence the game.

The metric of Game Engagement scored relatively lower compared to the others (Session 1: mean = 3.9, Session 2: mean = 4.3). A hypothesis regarding the lower-scored metric may be related to the nature of streaming as a passive entertainment source in general, where many may have it open “in the background” as passive lurkers [22, 91]. This relatively lower game engagement metric from the AEQ has been observed in other studies regarding livestreaming interactions [68], and could suggest that the engagement felt by

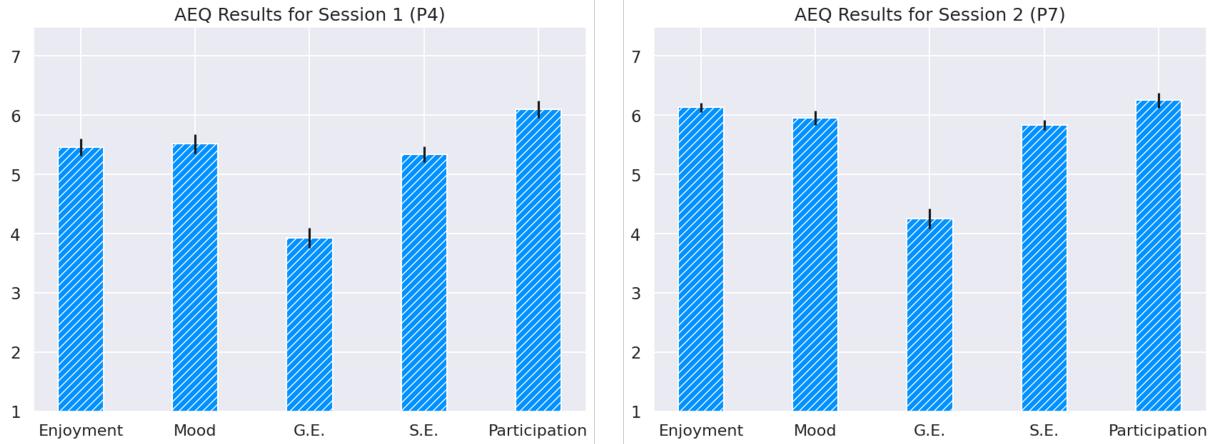


Figure 8: Mean AEQ values with 95% confidence error bars separated by stream session, from 1 (strongly disagree) to 7 (strongly agree). G.E. and S.E. are abbreviations of game engagement and social engagement, respectively.

the audience is more centred around the presence of others rather than immersion within the game itself [21].

8.1.2 Creating Novel Forms of Audience-Streamer and Audience-Audience Interaction. Interaction between streamer and viewer and between viewers themselves form important components of the livestreaming experience [52]. Participants indicated that engagement brought forth from interaction induced increased forms of connection with the streamer, e.g., “*it is a 2-way interaction unlike conventional streams which are 1-way*” (P4V12). P7V18 described it as a “*more intimate interaction than just watching*”. Participants felt that the streamer interacted with the chat more because of the direct input because the direct impact of the viewers incentivized communication – “*as we were affecting the gameplay they would comment and interact with us more based on that*” (P4V14).

Viewers also interacted with each other in interesting ways. The expressiveness of spatial input allowed for avenues of visual collaboration, e.g., “*I could see where other users spawned traps and vote on the upgrade we needed*” (P4V11), although the intent was somewhat murky at times – “*harder to understand intent since I didn't know when they would place traps compared to me*” (P4V12). An interesting development was the use of chat for planning and coordination; which also led to an increase in feelings of camaraderie – “*being able to coordinate with and chat live with other participants. We could strategize together on how to attack (enemies) and also how to help*” (P7V12). Viewer coordination was observed on both streams. In Session 1, viewers collaborated within chat to vote on augments and spawn upgraded traps; in Session 2, viewers expressed intention in chat to save up funds to give the streamer enough gold coins to get a helmet. These provide examples of the symbiotic relationship that direct spatial input could have with chat, with chat providing a place for audience discussion revolving around interactions afforded by spatial input.

8.1.3 Challenges with Latency, Scalability, and Feelings of Individual Agency. Viewers also presented challenges and suggested improvements for the system. One problem that was brought up was that of latency, e.g., “*it was interesting though slightly limited by delay*.”

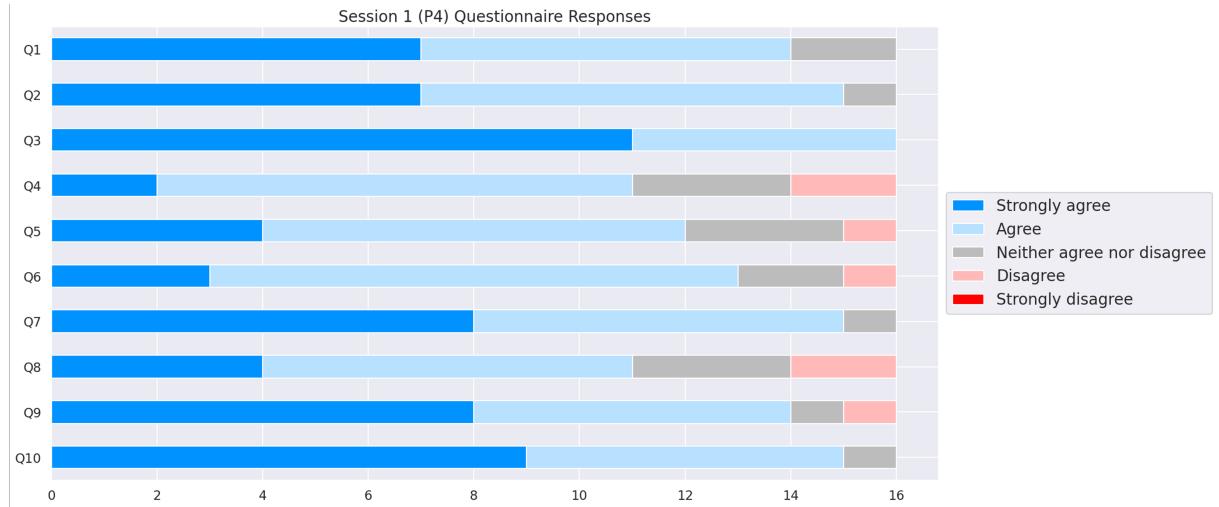
(P4V12). Even though our system tries to design for delay by using visual feedback to show that inputs are being processed and camera buffers to align the input with the viewer's tune-shifted perspective of the stream, we highlight delay as still a salient problem. We try to circumvent it through design, but its existence still negatively impacts the experience. Delay is inevitable given present livestreaming technology; however, viewers stated that a way to compensate for delay would be to have increased feedback on their interactions. This could provide increased recognition of interaction – “*sometimes I had trouble telling if my inputs were the ones registering*” (P4V8) – and understanding of interaction effects – “*I think if the feedback is more obvious... [right now] I am not quite sure whether my interaction is that helpful*” (P4V7). Overall, this suggests that increased transparency and feedback for viewers regarding their inputs could improve the overall experience in terms of individual agency.

Some viewers also raised scalability concerns. Viewers noted that the significance of an interaction might be relatively reduced with large numbers of viewers, e.g., “*I think with more people, the less impact each person has with the tools*” (P4V8) and that a large audience involves the large degree of potential interactions – “*It might be overwhelming if too many people can spawn things at once*” (P7V16). Given the limited and controlled number of audience members in our study, this is an aspect that still requires future research.

8.2 Streamer Experience

RQ1 – How does direct spatial input affect the streamer's experience, and what challenges and opportunities emerge for streamers?

8.2.1 Generating Enjoyment and Positive Engagement. VIBES was similarly well-received by the streamers, e.g., “*I think the game mechanic is cool and I got to see it in action*” (P4), “*I think it was fun*” (P7). The streamers also noted that VIBES changed interactions between them and the viewers – “*because we're playing the game together it does change how I interact. I probably was asking what the chat wants to do and such*” (P4). P4 also noticed how VIBES encouraged



Q1: The feature to interact with the stream was helpful.

Q2: The feature to observe the effects of my interactions on the stream was helpful.

Q3: I was more engaged with the stream with the tools, compared to the experience of not using the tools.

Q4: I felt like my message was delivered better to the streamer with the tools, compared to the experience of not using the tools.

Q5: I felt like I could understand how other users interacted with the tool.

Q6: I could express my intention more effectively with the tools, compared to the experience of not using the tools.

Q7: I felt like I had richer interaction with the streamer using the tools, compared to the experience of not using the tools.

Q8: It was easy to create visual input with the tools.

Q9: It was easy to know the range of inputs I could provide.

Q10: I would like to see other streamers use such tools in live streaming.

Figure 9: Raw distribution of questionnaire responses for P4’s stream (Session 1)

viewers to chat as well, as they would discuss which upgrade to get. Similarly, P7 indicated that during her stream, VIBES encouraged users to formulate a goal together — “*they had a goal that they wanted to accomplish which was to give me a bald cap [author note: bald cap = gold helmet]*”. Overall, much of the streamers’ sentiments regarding engagement and interaction echoed those of the viewers, as well as those raised in the initial formative exploration.

8.2.2 Impacting Streamer Agency and Control. We briefly explored understanding streamer agency and attribution while playing the games using VIBES, relating to the point of intentionality from the formative exploration. Both streamers scored rather highly on the metrics of internality (mean = 6.3 for P4 and mean = 6.0 for P7) and controllability (mean = 6.0 for P4 and mean = 5.7 for P7) on the GSAQ (Figure 11), implying that they believed themselves to have a high degree of internal control over their in-game performance. However, there were instances in which agency seemed to be mismatched in both games. P4 described that they thought “*the player was too OP [overpowered]*” and that “*for this game, I think it would be cooler if the power was shifted to the viewers*” (P4). In P7’s stream session, there was a moment where viewers collectively spawned many monsters at the same time, leaving the streamer unable to respond suitably and defend. Although this latter experience was chaotic and fun for the viewers, it shows that balancing streamer agency against viewer agency within the context of overall game

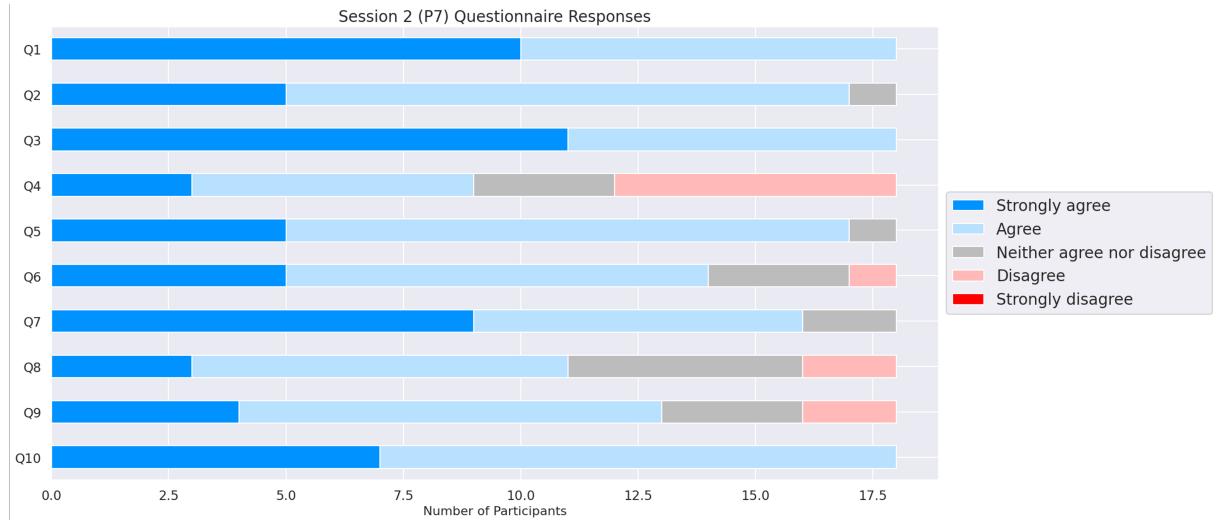
balance becomes a key consideration, given the asymmetric nature of games involving audience participation.

9 DISCUSSION

We discuss the implications and broader takeaways of our findings, contextualizing them around prior research into livestreaming, its motivations, and the overall experience for audiences and streamers. Furthermore, given that both studied applications were games, we discuss our findings through the lens of audience participation in game design.

9.1 Livestreaming Motivations and Experiences for Streamers and Audience

Direct spatial input enhanced livestream engagement and participation, fostering closer bonds between the viewers and streamers, as well as among viewers themselves. The connections between viewers and the streamer tie towards the social motivator for livestream viewership [52, 53]. Beyond just chat, spatial input generated a topic for communication with other audience members, which we interpret as strengthening the motivation of being alongside others in a common community [56, 79], and offering a sense of reward through direct impact [30]. By extending the range of interactions between streamer and viewer, spatial input fundamentally reinforces the motivations to watch and participate in streams [52, 59]. Considering this from previously discussed theoretical frameworks



- Q1: The feature to interact with the stream was helpful.
 Q2: The feature to observe the effects of my interactions on the stream was helpful.
 Q3: I was more engaged with the stream with the tools, compared to the experience of not using the tools.
 Q4: I felt like my message was delivered better to the streamer with the tools, compared to the experience of not using the tools.
 Q5: I felt like I could understand how other users interacted with the tool.
 Q6: I could express my intention more effectively with the tools, compared to the experience of not using the tools.
 Q7: I felt like I had richer interaction with the streamer using the tools, compared to the experience of not using the tools.
 Q8: It was easy to create visual input with the tools.
 Q9: It was easy to know the range of inputs I could provide.
 Q10: I would like to see other streamers use such tools in live streaming.

Figure 10: Raw distribution of questionnaire responses for P7's stream (Session 2)

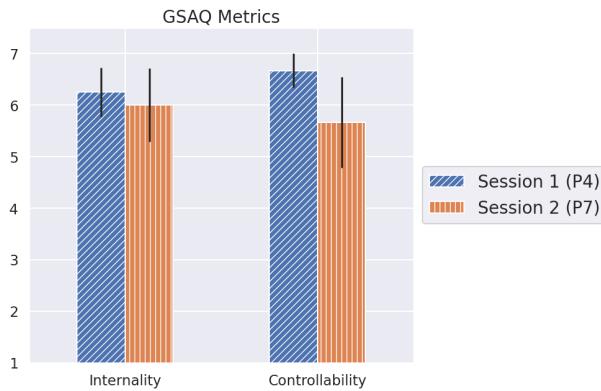


Figure 11: Mean GSAQ values with 95% confidence error bars for the subscales of internality and controllability, from 1 (strongly disagree) to 7 (strongly agree).

for motivation, VIBES improved social engagement and interaction from a uses and gratifications perspective [74], and ties towards the constructs of social benefits from a self-determination viewpoint [97].

Both background research [12, 52, 79] and our formative evaluation revealed the importance of interaction and engagement for

streamers. In our study, streamers generally noted how applications with VIBES encouraged inter-viewer discussion and increased social engagement relating to novel interactions. Tying into the theoretical framework, Young and Wiedenfeld indicated that this social engagement is important when relating to both self-determination and uses and gratification approaches regarding motivations for livestreaming [94]. Thus, our findings suggest that providing viewers with direct spatial input into the game can also positively affect the social motivators for streamers. Future work can consider extending the range of possibilities of such interactions, e.g. perhaps to leave persistent elements in the streamed content based on viewer input similar to messages in Dark Souls [20].

Taylor stated that, in livestreams, the audience and their interactions are integrated into the entire experience [79]. This is particularly evident in our work, as the streamer and audience interact together. We observed a reciprocal push-and-pull feedback loop in the streams — the streamer would interact with the audience, who would then use VIBES to interact with the streamer. This concept of reciprocity [75] allows for the co-creation of content [75] together with the streamer in a fundamentally different manner; direct spatial input allows the audience to more literally play together with the streamer. Spatially-mapped semantic information through mouse events provides an additional interaction medium for streamers and viewers to bi-directionally engage, co-creating a

unique livestreaming experience (as P7V11 states, “*the tool gave an experience that the game nor stream was able to provide*”).

However, from a self-determination perspective [70], direct spatial input from the audience also transfers autonomy from the streamer to the audience collective, reflected in the balance of agency within the streamed content. Increased audience influence may inhibit the streamer’s ability to maintain leadership or demonstrate competence [94]. The loss of agency could also affect how streamers attribute their performance in a games’ context [19], tying it towards external factors rather than their competence. The removal of autonomy from the streamer can create a level of uncertainty and unpredictability. Although uncertainty can be engaging for digital experiences (being a way to keep people interested and support serendipitous experiences [42, 45]), a lack of control can also result in frustration for the streamer. When we take the perspective of streamers and viewers co-creating, then co-destruction can occur when there is a lack of trust, a lack of clear expectations, misbehaviour, and so forth [39], which can in turn ruin the relationship between the streamer and viewers. The balance between positive effects towards social relatedness and the negative impact of loss of streamer autonomy is important to consider when using a system like VIBES or most general APGs. We consider how the maintenance of a trusting co-creation environment can be supported – facilitating the balance of autonomy in a nuanced and controlled way to ultimately be mindful of the streamer, the viewers, and their respective needs.

9.2 Perspectives on Audience-Participation Games and Beyond

Both studied applications were games, therefore, we relate our findings to audience participation games in research. Agreeing with past research [73], we found that audience members were able to take on both individual as well as collectivist roles within the studies. However, we run into similar challenges as presented by prior studies on APGs [26], namely, the issues of latency and the balance of player agency. One key finding in prior research was that text-based chat was limited in viewer expressiveness for APGs [46]. Our work did improve on this aspect – we highlighted the possibility for audience members to interact visuospatially on a visual medium (the stream), and we found that this aspect of direct spatial feedback and input was well-received. Our findings suggest that this led to higher degrees of participatory gameplay as players can directly see the outcome of their action; agreeing with prior work that visual interactions enable richer and more extensive interactions [14, 93].

VIBES contrasts existing tools such as Vispoll [14] and StreamSketch [58], which all allow a user to interact on a stream but do not directly impact the streamed content – in these systems, the mediating layer of the streamers themselves decides how they want to use player input. As VIBES passes the audience input directly into the games themselves, we revisit the aspect of *autonomy* within the audience collective (as it relates to self-determination theory [70]). While collective audience input enhances engagement, individual agency decreases as viewer numbers grow [11]. Feedback is important, and viewers want to feel like their actions mean something. Yet, with a large viewer base and increased individual agency, the

stream may lack the social feeling of having a ‘crowd’ [11]. As such, there exist challenges for individual motivators when the audience is viewed as a collective. Future work can look into building games and systems that explore individual audience identity and autonomy within a collective, perhaps being inspired by popular danmaku-participation games, which give viewers individualized control [11, 84].

Beyond just motivators of enjoyment and sociality, games have also been well-studied to have potential prosocial outcomes [49, 86], which can derive from the analogue between decisions in-game and behaviour outside of it [37, 38]. We hypothesize that audience-participation games using direct spatial input can encourage prosocial outcomes for both the audience and the streamer, extending upon the work of Apostolellis and Bowman [6]. Compared to usual livestreaming, which can be more passive [91], our work encodes an active visual component to encourage co-playing and co-presence. This could potentially positively affect prosocial and serious games as well, as the audience becomes more involved and feels like they have some aspect of control of that their decisions have meaning or choice. Although livestreaming has often been associated with a gaming context, its practice has been applied to many different domains, such as education [12] or cultural exchange [56]. This aligns well with serious games [43], which are games that have been used for educational purposes. Although our work has studied the context of spatial input in games contexts, future work could potentially extend these findings by leaning into studying its value in learning [6, 7], especially taking advantage of visual learners.

10 FUTURE WORK

Our findings revealed challenges brought up in our explorations by both streamers and viewers, we expand upon challenges for improvement and possible future extensions. These recommendations aim to balance the discussed positives of creative audience interaction while being mindful of streamer agency, safety, control, and enjoyment.

10.1 Scalability and Streamer-Viewer Agency

Both streamers and viewers brought up scalability as a point of concern. Streamers wondered how agency might be affected by thousands of inputs all at once, and viewers were concerned about their agency of their interactions within a large collective. To preface, the vast majority of streams on Twitch do not have particularly high viewership⁵), so designing for scale entails designing for the exception (with a distinct set of needs and requirements) rather than the norm. Scalability, and its associated effects on agency, was a dimension that we considered in design and to explore in the future, but was not as relevant towards our specific exploration. However, we speculate on the possible challenges associated with scale and propose future explorations and designs based on our findings.

One method of controlling the effect of scale on the balance of streamer-viewer agency, which we attempted in *Terraria Interaction Mod*, was by lowering the possible number of interactions of a single individual when there is a higher total number of participants. One participant commented on this – “*I liked that the more people there*

⁵<https://sullygnome.com/channels/14/metadata> [Last accessed: March 12, 2025]

were, the fewer points we get so it's not just bullying the streamer but balanced" (P7V18). In such a case, each viewer still retains the impact of their actions, just at a less frequent rate. On the other hand, another method of reducing individual viewer agency is through aggregation of inputs, for example, in Figure 4, so that no one individual interaction can have a significant impact. Although this may dilute individual impact, it can be done more frequently, pointing towards the hypothetical trade-off between the impact of viewer action and their frequency.

These previous suggestions presume viewers as equal, yet this is not true in existing streams — viewers can be categorized in a hierarchical system of VIPs, mods, subscribers, etc.; leaderboards may also exist to reinforce this hierarchy. Thus, one way of balancing agency following this existing hierarchy may be to shift individual agency towards higher-tiered viewers. Furthermore, to incentivize increased engagement and interaction for viewers, some suggestions were made to allow for more individual interactions if the viewer is already engaging — serving as a positive reward feedback system [83] akin to this hierarchy. Overall, several potential mechanisms to tune VIBES at scale could be explored in future deployments; our design recommendations include dynamic interaction scaling and toggleable options for streamers to control the interaction intensity (and thus the balance of agency) in their stream.

10.2 Interaction Privacy and Anti-Harassment

We found that many streamers were concerned with issues regarding trolling and harassment given a new medium of audience-streamer interaction. The rise of livestreaming has inevitably led to issues with toxic behaviour in the medium [25, 28, 66, 96]. Toxicity by viewers toward a streamer using live chat can take the form of spamming or harassment [66], and can be challenging to deal with due to overwhelming flow [25]. Equivalent occurrences can take place using direct spatial input, e.g., spamming or generating unwanted inputs. To address this issue, we consider future design possibilities that could incorporate proactive and reactive measures as options for safety and anti-harassment [72]. For example, VIBES could incorporate proactive detection of spam-like inputs and prevent them from affecting the application or display interactions publicly with an attached username to deter anonymous trolling; on the other hand, reactive measures could involve banning or timing out a troublesome user after the infraction has occurred [28].

10.3 Practical Deployment and Present Overhead

In our work, VIBES is a prototype that we use as a technology probe; a more formalized deployment of such a system in practice still needs several steps that were not performed in our present implementation. For example, our Twitch extensions were deployed in the testing stages, meaning that each viewer and streamer had to be individually added to the extension as a tester rather than it being publicly available to all; moving to a public deployment requires review from Twitch themselves. This finalized deployment of the

Twitch extension would likely provide several usability improvements, e.g. increased responsiveness through caching, standardized distribution channels, etc.

At present, VIBES currently has an inevitable installation overhead at the moment due to the browser extension, which needs to be installed by the viewers to collect and pass mouse events. Because of the nature of the present architecture, there is some set-up for both the streamer and the viewer. In the future, we recommend working toward collaborating with streaming platforms for native mouse-event tracking, allowing for more streamlined public access and integration. This may also pave the way for cross-platform accessibility — we envision that direct spatial input can be developed for and assessed for other mediums and interaction events, e.g. swipe and touch on mobile devices.

11 LIMITATIONS

We reiterate that our research is largely exploratory, focusing on probing into the research questions regarding viewer and streamer experiences. Still, we identify and discuss several limitations in our study. Despite the open-ended nature of our research questions, our technology probe focused on a specific platform (Twitch) on a specific medium (browser-based computers). Future work could investigate mobile integration as well, considering the popularity of mobile viewership [13, 16] and the differences in viewership behaviour [17]. Although our mouse-based interactions would map naturally to touch inputs in mobile implementation from a technical perspective, we emphasize the need to extend the study of engagement and participation, especially given differences in screen size, haptic feeling, etc [3, 90]. Furthermore, while the research question was application-agnostic, our application investigation focused primarily on two specific gaming contexts. Although many livestreams do involve gameplay elements, future research could consider direct spatial input in other contexts, such as education [12] or cultural exchange [56]. We highlight limitations regarding participant demographics. Firstly, the streamer participants in the initial formative exploration were all low-viewer, non-full-time streamers. Although the viewership numbers are expected — the vast majority of streamers on Twitch do have a relatively low number of viewers (stream viewership with the same order of magnitude as our deployment study is already quite rare⁶). Still, it would be important to see how direct spatial input might affect streamers who stream on a more regular basis for prolonged periods to higher viewership, especially as we highlighted scalability as a key issue to address and explore in the future. Streamer experience would be an additional factor to consider in understanding how streamers build and engage their community. In addition, although the number of participants (especially streamers) was deemed adequate for initial exploration, future work could expand our work with a more diverse and larger participant sample for increased generalizability and information power [60].

In the livestream study sessions, we recruited viewer participants who had varied levels of experience with livestream viewership; participants likely individually brought some preconceived experience towards livestreaming when answering the research questions, which ask to compare the stream with VIBES to a hypothetical

⁶<https://sullygnome.com/channels/14/metadata> [Last accessed: March 12, 2025]

stream without it. We acknowledge that in the study, we primarily relied on a comparison of a stream with VIBES to a generic recollection of the experience of watching streams, which can differ with experience. Furthermore, we also highlight limitations regarding our choice of recruiting assigned audiences rather than using existing audience members. Although this recruiting approach has been applied in prior relevant research [10, 14, 31] and offers practical methodological benefits such as controlling the level of viewership and managing scheduling, we acknowledge it may affect the nature of interactions and relationship-building between the viewers and streamers; future works could investigate deploying towards a streamers' existing audience, who also may offer a better baseline for comparison. Lastly, to extend the qualitative data captured in this work, future work could look into capturing and comparing the more quantitative metrics in psychophysiological work, such as heart rate or electrodermal activity [64] to understand affective state; offering an alternative but important way to assess quality of experience [81].

12 CONCLUSION

In this work, we investigated the potential effects of providing audience members with real-time direct spatial input to livestreamed content, using interactive events such as click and motion on the streamed video. To facilitate this investigation, we conducted a technology probe using VIBES, our developed system that allows viewers to pass mouse events as input to affect livestreamed content on Twitch. We demonstrated the system to streamers in an initial formative exploration to understand the potential of the concept, where and how it may be useful, and what challenges might exist. We then developed two complete applications that we deployed in livestreams with viewers. We found that VIBES was well-received – viewers and streamers alike were pleased with its ability to extend present interactions, promote streamer-audience engagement, and form collaborative participatory spaces. We reflected on VIBES within the sphere of livestreaming research and outlined present limitations and potential future work for improvement.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Science and Engineering Research Council of Canada (NSERC) under Discovery Grant RGPIN-2019-05624.

REFERENCES

- [1] Amazon Web Services 2023. *Video Latency in Live Streaming*. Amazon Web Services. <https://aws.amazon.com/media/tech/video-latency-in-live-streaming/>
- [2] A. Feizi and Chui Yin Wong. 2012. Usability of user interface styles for learning a graphical software application. In *2012 International Conference on Computer & Information Science (ICCIS)*, Vol. 2. 1089–1094. <https://doi.org/10.1109/ICCISci.2012.6297188> Journal Abbreviation: 2012 International Conference on Computer & Information Science (ICCIS).
- [3] Sushma Adepu and Rachel F. Adler. 2016. A Comparison of Performance and Preference on Mobile Devices vs. Desktop Computers. In *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. 1–7. <https://doi.org/10.1109/UEMCON.2016.7777808>
- [4] Chee Eun Ahn, Woohun Lee, Hunmin Park, and Jiwoo Hong. 2022. MeowPlayLive: Enhancing Animal Live Streaming Experience Through Voice Message-Based Real-Time Viewer-Animal Interaction. In *Designing Interactive Systems Conference (Virtual Event, Australia) (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 849–864. <https://doi.org/10.1145/3532106.3533553>
- [5] Nolan Alexander, David Brennan, John Eshirow, Joshua Rosenblatt, Justin Wolter, William Scherer, and James Valeiras. 2021. The Future of Livestreaming: Strategy and Predictive Analysis for Future Operations of Facebook Live. In *2021 Systems and Information Engineering Design Symposium (SIEDS)*. 1–6. <https://doi.org/10.1109/SIEDS52267.2021.9483785>
- [6] Panagiotis Apostolellis and Doug A. Bowman. 2016. Audience Involvement and Agency in Digital Games: Effects on Learning, Game Experience, and Social Presence. In *Proceedings of the The 15th International Conference on Interaction Design and Children (Manchester, United Kingdom) (IDC '16)*. Association for Computing Machinery, New York, NY, USA, 299–310. <https://doi.org/10.1145/2930674.2930700>
- [7] Panagiotis Apostolellis and Thanasis Daradoumis. 2010. Exploring the Value of Audience Collaboration and Game Design in Immersive Virtual Learning Environments. In *Proceedings of the 9th International Conference on Interaction Design and Children (IDC '10)*. Association for Computing Machinery, New York, NY, USA, 326–330. <https://doi.org/10.1145/1810543.1810605>
- [8] Teresa Cerrato-Pargman, Chiara Rossitto, and Louise Barkhuus. 2014. Understanding Audience Participation in an Interactive Theater Performance. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (Helsinki, Finland) (NordiCHI '14)*. Association for Computing Machinery, New York, NY, USA, 608–617. <https://doi.org/10.1145/2639189.2641213>
- [9] Chih-Ping Chen. 2021. Digital gifting in personal brand communities of live-streaming: fostering viewer–streamer–viewer parasocial relationships. *Journal of Marketing Communications* 27, 8 (2021), 865–880. <https://doi.org/10.1080/13527266.2021.1910327>
- [10] Di (Laura) Chen, Dustin Freeman, and Ravin Balakrishnan. 2019. Integrating Multimedia Tools to Enrich Interactions in Live Streaming for Language Learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300668>
- [11] Qijia Chen, Jingwen Gan, Andrea Bellucci, and Giulio Jacucci. 2024. "I Felt Everyone Was a Streamer": An Empirical Study on What Makes Avatar Collective Streaming Engaging. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 67:1–67:25. <https://doi.org/10.1145/3637344>
- [12] Xinyue Chen, Si Chen, Xu Wang, and Yun Huang. 2021. "I Was Afraid, but Now I Enjoy Being a Streamer!": Understanding the Challenges and Prospects of Using Live Streaming for Online Education. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 237 (jan 2021), 32 pages. <https://doi.org/10.1145/3432936>
- [13] Urszula Chrȧżek-Barczyk et al. 2018. Live streaming and its reception among the recipients. *European Journal of Service Management* 26, 2 (2018), 29–35.
- [14] John Joon Young Chung, Hijung Valentina Shin, Haijun Xia, Li-yi Wei, and Rubaiat Habib Kazi. 2021. Beyond Show of Hands: Engaging Viewers via Expressive and Scalable Visual Communication in Live Streaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 109, 14 pages. <https://doi.org/10.1145/3411764.3445419>
- [15] Andrew Cross, Mydhili Bayyapunedil, Dilip Ravindran, Edward Cutrell, and William Thies. 2014. VidWiki: Enabling the Crowd to Improve the Legibility of Online Educational Videos. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (Baltimore, Maryland, USA) (CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1167–1175. <https://doi.org/10.1145/2531602.2531670>
- [16] Stuart Cunningham, David Craig, and Junyi Lv. 2019. China's livestreaming industry: platforms, politics, and precarity. *International Journal of Cultural Studies* 22, 6 (2019), 719–736. <https://doi.org/10.1177/1367877919834942> arXiv:<https://doi.org/10.1177/1367877919834942>
- [17] Daniel V. C. da Silva, Antonio A. de A. Rocha, and Pedro B. Velloso. 2021. Mobile vs. Non-Mobile Live-Streaming: A Comparative Analysis of Users Engagement and Interruption Using Big Data from a Large CDN Perspective. *Sensors* 21, 16 (2021). <https://doi.org/10.3390/s21165616>
- [18] James R. Dabrowski and Ethan V. Munson. 2001. Is 100 Milliseconds Too Fast?. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems (Seattle, Washington) (CHI EA '01)*. Association for Computing Machinery, New York, NY, USA, 317–318. <https://doi.org/10.1145/634067.634255>
- [19] Ansgar E. Depping and Regan L. Mandryk. 2017. Why is This Happening to Me? How Player Attribution Can Broaden Our Understanding of Player Experience. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (Denver, Colorado, USA) (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1040–1052. <https://doi.org/10.1145/3025453.3025648>
- [20] Daniel M. Dooghan. 2025. Fantasies of Adequacy: Mythologies of Capital in Dark Souls. *Games and Culture* 20, 2 (March 2025), 151–170. <https://doi.org/10.1177/15554120231192080>
- [21] John Downs, Frank Vetere, Steve Howard, and Steve Loughnan. 2013. Measuring Audience Experience in Social Videogaming. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (Adelaide, Australia) (OzCHI '13)*. Association for Computing Machinery, New York, NY, USA, 217–220. <https://doi.org/10.1145/2541016.2541054>
- [22] Noella Edelmann. 2017. Lurkers' Contribution to E-Participation. In *Conference for E-Democracy and Open Government*. 139.

- [23] Satu Elo and Helvi Kyngäs. 2008. The Qualitative Content Analysis Process. *Journal of Advanced Nursing* 62, 1 (2008), 107–115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>
- [24] Justin Fanzo, Rachel Gu, Jinchao Han, Ketul Majmudar, Tony Deng, Flora Cheng, Chaoya Li, and Jessica Hammer. 2017. What Lurks in the Dark: An Audience Participation Horror Game. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play* (Amsterdam, The Netherlands) (*CHI PLAY '17 Extended Abstracts*). Association for Computing Machinery, New York, NY, USA, 621–624. <https://doi.org/10.1145/3130859.3130865>
- [25] Enrico Gandolfi and Richard E Ferdig. 2022. Sharing dark sides on game service platforms: Disruptive behaviors and toxicity in DOTA2 through a platform lens. *Convergence* 28, 2 (2022), 468–487. <https://doi.org/10.1177/13548565211028809> arXiv:<https://doi.org/10.1177/13548565211028809>
- [26] Seth Glickman, Nathan McKenzie, Joseph Seering, Rachel Moeller, and Jessica Hammer. 2018. Design Challenges for Livestreamed Audience Participation Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) (*CHI PLAY '18*). Association for Computing Machinery, New York, NY, USA, 187–199. <https://doi.org/10.1145/3242671.3242708>
- [27] Monica Gori, Sara Price, Fiona N. Newell, Nadia Berthouze, and Gualtiero Volpe. 2022. Multisensory Perception and Learning: Linking Pedagogy, Psychophysics, and Human–Computer Interaction. (April 2022). <https://doi.org/10.1163/22134808-bja10072>
- [28] Maike Groen. 2020. *Digital Governmentality: Toxicity in Gaming Streams*. Springer Fachmedien Wiesbaden, Wiesbaden, 97–111. https://doi.org/10.1007/978-3-658-28175-5_7
- [29] Daniel Gros, Brigitte Wanner, Anna Hackenholt, Piotr Zawadzki, and Kathrin Knautz. 2017. World of Streaming. Motivation and Gratification on Twitch. In *Social Computing and Social Media. Human Behavior*, Gabriele Meiselwitz (Ed.). Springer International Publishing, Cham, 44–57.
- [30] William A. Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on Twitch: Fostering Participatory Communities of Play within Live Mixed Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (*CHI '14*). Association for Computing Machinery, New York, NY, USA, 1315–1324. <https://doi.org/10.1145/2556288.2557048>
- [31] William A. Hamilton, John Tang, Gina Venolia, Kori Inkpen, Jakob Zillner, and Dereh Huang. 2016. Rivulet: Exploring Participation in Live Events through Multi-Stream Experiences. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (Chicago, Illinois, USA) (*TVX '16*). Association for Computing Machinery, New York, NY, USA, 31–42. <https://doi.org/10.1145/293206.2932211>
- [32] Noor Hammad, Erik Harpstead, and Jessica Hammer. 2023. The View from MARS: Empowering Game Stream Viewers with Metadata Augmented Real-Time Streaming. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (*UIST '23*). Association for Computing Machinery, New York, NY, USA, Article 62, 13 pages. <https://doi.org/10.1145/3586183.3606753>
- [33] Sebastian Herscher, Connor Defanti, Nicholas Gregory Vitovitch, Corinne Brenner, Hajin Xia, Kris Layng, and Ken Perlin. 2019. CAVRN: An Exploration and Evaluation of a Collective Audience Virtual Reality Nexus Experience. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) (*UIST '19*). Association for Computing Machinery, New York, NY, USA, 1137–1150. <https://doi.org/10.1145/3332165.3347929>
- [34] Hsiu-Fang Hsieh and Sarah E. Shannon. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15, 9 (Nov. 2005), 1277–1288. <https://doi.org/10.1177/104973205276687>
- [35] Zeyu Huang, Xinyi Cao, Yuanhao Zhang, and Xiaojuan Ma. 2024. Sharing Frissons among Online Video Viewers: Exploring the Design of Affective Communication for Aesthetic Chills. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (*CHI '24*). Association for Computing Machinery, New York, NY, USA, 1–19. <https://doi.org/10.1145/3613904.3642818>
- [36] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) (*CHI '03*). Association for Computing Machinery, New York, NY, USA, 17–24. <https://doi.org/10.1145/642611.642616>
- [37] Glenna H. Iten, Julia Ayumi Bopp, Clemens Steiner, Klaus Opwis, and Elisa D. Mekler. 2018. Does a Prosocial Decision in Video Games Lead to Increased Prosocial Real-Life Behavior? The Impact of Reward and Reasoning. *Computers in Human Behavior* 89 (Dec. 2018), 163–172. <https://doi.org/10.1016/j.chb.2018.07.031>
- [38] Glenna H. Iten, Sharon T. Steinemann, and Klaus Opwis. 2018. Choosing to Help Monsters: A Mixed-Method Examination of Meaningful Choices in Narrative-Rich Games and Interactive Narratives. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (*CHI '18*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3173915>
- [39] Henna Järvi, Anni-Kaisa Kähkönen, and Hannu Torvinen. 2018. When Value Co-Creation Fails: Reasons That Lead to Value Co-Destruction. *Scandinavian Journal of Management* 34, 1 (March 2018), 63–77. <https://doi.org/10.1016/j.scaman.2018.01.002>
- [40] Rubaiat Habib Kazi, Fanny Chevalier, Tovi Grossman, and George Fitzmaurice. 2014. Kitty: Sketching Dynamic and Interactive Illustrations. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) (*UIST '14*). Association for Computing Machinery, New York, NY, USA, 395–405. <https://doi.org/10.1145/2642918.2647375>
- [41] Seokyel Kim and Jinah Park. 2018. Touchable Video Streams: Towards Multi-Sensory and Multi-Contact Experiences. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video* (SEOUL, Republic of Korea) (*TVX '18*). Association for Computing Machinery, New York, NY, USA, 155–160. <https://doi.org/10.1145/3210825.3213555>
- [42] Shringi Kumari, Sebastian Deterding, and Jonathan Freeman. 2019. The Role of Uncertainty in Moment-to-Moment Player Motivation: A Grounded Theory. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (*CHI PLAY '19*). Association for Computing Machinery, New York, NY, USA, 351–363. <https://doi.org/10.1145/3311350.3347148>
- [43] Fedwa Laamarti, Mohamad Eid, and Abdulmaleb El Saddik. 2014. An Overview of Serious Games. *International Journal of Computer Games Technology* 2014, 1 (2014), 358152. <https://doi.org/10.1155/2014/358152>
- [44] Scott Lee-Tierman and Jonathan Grudin. 2001. Fostering Engagement in Asynchronous Learning through Collaborative Multimedia Annotation.. In *INTERACT*. Citeseer, 472–479.
- [45] Tuck Wah Leong. 2006. Designing for Experiences: Randomness as a Resource. In *Proceedings of the 6th Conference on Designing Interactive Systems (DIS '06)*. Association for Computing Machinery, New York, NY, USA, 346–347. <https://doi.org/10.1145/1142405.1142458>
- [46] Pascal Lessel, Michael Mauderer, Christian Wolff, and Antonio Krüger. 2017. Let's Play My Way: Investigating Audience Influence in User-Generated Gaming Live-Streams. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video* (Hilversum, The Netherlands) (*TVX '17*). Association for Computing Machinery, New York, NY, USA, 51–63. <https://doi.org/10.1145/3077548.3077556>
- [47] Pascal Lessel, Alexander Vielhauer, and Antonio Krüger. 2017. CrowdChess: A System to Investigate Shared Game Control in Live-Streams. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Amsterdam, The Netherlands) (*CHI PLAY '17*). Association for Computing Machinery, New York, NY, USA, 389–400. <https://doi.org/10.1145/3116595.3116597>
- [48] Pascal Lessel, Alexander Vielhauer, and Antonio Krüger. 2017. Expanding Video Game Live-Streams with Enhanced Communication Channels: A Case Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (*CHI '17*). Association for Computing Machinery, New York, NY, USA, 1571–1576. <https://doi.org/10.1145/3025453.3025708>
- [49] Hua Li and Qian Zhang. 2023. Effects of Prosocial Video Games on Prosocial Thoughts and Prosocial Behaviors. *Social Science Computer Review* 41, 3 (June 2023), 1063–1080. <https://doi.org/10.1177/08944393211069599>
- [50] Jie Li, Xinning Gui, Yubo Kou, and Yukun Li. 2019. Live Streaming as Co-Performance: Dynamics between Center and Periphery in Theatrical Engagement. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 64 (nov 2019), 22 pages. <https://doi.org/10.1145/3359166>
- [51] Ran Li, Yaobin Lu, Jifeng Ma, and Weiquan Wang. 2021. Examining gifting behavior on live streaming platforms: An identity-based motivation model. *Information & Management* 58, 6 (2021), 103406. <https://doi.org/10.1016/j.im.2020.103406> PACIS2019 Special Issue: Emerging technology, business, and applications in digital economy.
- [52] Yi Li, Chongli Wang, and Jing Liu. 2020. A Systematic Review of Literature on User Behavior in Video Game Live Streaming. *International Journal of Environmental Research and Public Health* 17, 9 (2020). <https://doi.org/10.3390/ijerph17093328>
- [53] Jinglan Lin and Zhicong Lu. 2017. The Rise and Proliferation of Live-Streaming in China: Insights and Lessons. In *HCI International 2017 – Posters' Extended Abstracts*, Constantine Stephanidis (Ed.). Springer International Publishing, Cham, 632–637.
- [54] Lili Liu, Ayoung Suh, and Christian Wagner. 2017. Who Is with You? Integrating a Play Experience into Online Video Watching via Danmaku Technology. In *Human-Computer Interaction. Interaction Contexts*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 63–73. https://doi.org/10.1007/978-3-319-58077-7_6
- [55] Danielle Lottridge, Frank Bentley, Matt Wheeler, Jason Lee, Janet Cheung, Katherine Ong, and Cristy Rowley. 2017. Third-Wave Livestreaming: Teens' Long Form Selfie. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) (*MobileHCI '17*). Association for Computing Machinery, New York, NY, USA, Article 20, 12 pages. <https://doi.org/10.1145/3098279.3098540>
- [56] Zhicong Lu, Michelle Annett, Mingming Fan, and Daniel Wigdor. 2019. "I Feel It Is My Responsibility to Stream": Streaming and Engaging with Intangible Cultural Heritage through Livestreaming. In *Proceedings of the 2019 CHI*

- Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300459>
- [57] Zhicong Lu, Seongkook Heo, and Daniel J. Wigdor. 2018. StreamWiki: Enabling Viewers of Knowledge Sharing Live Streams to Collaboratively Generate Archival Documentation for Effective In-Stream and Post Hoc Learning. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 112 (nov 2018), 26 pages. <https://doi.org/10.1145/3274381>
- [58] Zhicong Lu, Rubaiat Habib Kazi, Li-yi Wei, Mira Dontcheva, and Karrie Karahalios. 2021. StreamSketch: Exploring Multi-Modal Interactions in Creative Live Streams. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 58 (apr 2021), 26 pages. <https://doi.org/10.1145/3449132>
- [59] Zhicong Lu, Haijun Xia, Seongkook Heo, and Daniel Wigdor. 2018. You Watch, You Give, and You Engage: A Study of Live Streaming Practices in China. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI '18*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3174040>
- [60] Kirsti Malterud, Volkert Dirk Siersma, and Ann Dorrit Guassora. 2016. Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qualitative Health Research* 26, 13 (Nov. 2016), 1753–1760. <https://doi.org/10.1177/1049732315617444>
- [61] Xin Pan. 2023. Motivations for Game Stream Spectatorship: A Content Analysis of Danmaku on Bilibili. *Global Media and China* 8, 2 (June 2023), 190–212. <https://doi.org/10.1177/20594364231179750>
- [62] Pixel by Pixel Studios Inc. 2018. *Marbles on Stream*. Game [Windows]. Pixel by Pixel Studios Inc..
- [63] Dennis Ramirez, Jenny Saucerman, and Jeremy Dietmeier. 2014. Twitch plays pokemon: a case study in big g games. In *Proceedings of DiGRA*. 3–6.
- [64] Niklas Ravaja. 2004. Contributions of Psychophysiology to Media Research: Review and Recommendations. *Media Psychology* 6, 2 (May 2004), 193–235. https://doi.org/10.1207/s1527285xmp0602_4
- [65] Re-Logic. 2011. *Terraria*. Game [Windows, PlayStation 3, Xbox 360, PlayStation Vita, iOS, Android, Windows Phone, PlayStation 4, PlayStation 5, Xbox One, Xbox Series X/S, macOS, Linux, Nintendo 3DS, Wii U, Nintendo Switch, Google Stadia]. 505 Games.
- [66] Olivia Rines. 2021. *Conceptualizing Toxicity in Women Twitch Streamers' Communities*. Ph. D. Dissertation. Arizona State University.
- [67] Charles Roberts and Tobias Hollerer. 2011. Composition for Conductor and Audience: New Uses for Mobile Devices in the Concert Hall. In *Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology* (Santa Barbara, California, USA) (*UIST '11 Adjunct*). Association for Computing Machinery, New York, NY, USA, 65–66. <https://doi.org/10.1145/2046396.2046425>
- [68] Raquel Breejon Robinson, Ricardo Rheeder, Madison Klarkowski, and Regan L Mandryk. 2022. "Chat Has No Chill": A Novel Physiological Interaction For Engaging Live Streaming Audiences. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 562, 18 pages. <https://doi.org/10.1145/3491102.3501934>
- [69] Ruffle. 2015. *Storm the House*. Game [Web Browser]. <https://www.crazygames.com/game/storm-the-house>.
- [70] Richard M Ryan and Edward L Deci. 2022. Self-determination theory. In *Encyclopedia of quality of life and well-being research*. Springer, 1–7.
- [71] Johnny Saldana. 2021. The Coding Manual for Qualitative Researchers. (2021), 1–440. <https://www.torrossa.com/en/resources/an/5018667> Publisher: SAGE Publications Ltd.
- [72] Joseph Seering, Robert Kraut, and Laura Dabbish. 2017. Shaping Pro and Anti-Social Behavior on Twitch Through Moderation and Example-Setting. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) (*CSCW '17*). Association for Computing Machinery, New York, NY, USA, 111–125. <https://doi.org/10.1145/2998181.2998277>
- [73] Joseph Seering, Saiph Savage, Michael Eagle, Joshua Churchin, Rachel Moeller, Jeffrey P. Bigham, and Jessica Hammer. 2017. Audience Participation Games: Blurring the Line Between Player and Spectator. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) (*DIS '17*). Association for Computing Machinery, New York, NY, USA, 429–440. <https://doi.org/10.1145/3064663.3064732>
- [74] Max Sjöblom and Juho Hamari. 2017. Why do people watch others play video games? An empirical study on the motivations of Twitch users. *Computers in Human Behavior* 75 (2017), 985–996. <https://doi.org/10.1016/j.chb.2016.10.019>
- [75] Thomas Smith, Marianna Obrist, and Peter Wright. 2013. Live-Streaming Changes the (Video) Game. In *Proceedings of the 11th European Conference on Interactive TV and Video* (Como, Italy) (*EuroTV '13*). Association for Computing Machinery, New York, NY, USA, 131–138. <https://doi.org/10.1145/2465958.2465971>
- [76] Abbie Speed, Alycia Burnett, and Tom Robinson II. 2023. Beyond the Game: Understanding why people enjoy viewing Twitch. *Entertainment Computing* 45 (2023), 100545. <https://doi.org/10.1016/j.entcom.2022.100545>
- [77] Nancy Staggers and David Kobus. 2000. Comparing Response Time, Errors, and Satisfaction Between Text-based and Graphical User Interfaces During Nursing Order Tasks. *Journal of the American Medical Informatics Association* 7, 2 (March 2000), 164–176. <https://doi.org/10.1136/jamia.2000.0070164>
- [78] Alina Striner, Andrew M. Webb, Jessica Hammer, and Amy Cook. 2021. Mapping Design Spaces for Audience Participation in Game Live Streaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 329, 15 pages. <https://doi.org/10.1145/3411764.3445511>
- [79] T. L. Taylor. 2019. *Watch Me Play*. Princeton University Press, Princeton. <https://doi.org/doi:10.1515/9780691184975>
- [80] Kimberly Tee, Saul Greenberg, and Carl Gutwin. 2006. Providing Artifact Awareness to a Distributed Group through Screen Sharing. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (Banff, Alberta, Canada) (*CSCW '06*). Association for Computing Machinery, New York, NY, USA, 99–108. <https://doi.org/10.1145/1180875.1180891>
- [81] Sara Vlahovic, Mirko Suznjevic, and Lea Skorin-Kapov. 2022. A Survey of Challenges and Methods for Quality of Experience Assessment of Interactive VR Applications. *Journal on Multimodal User Interfaces* 16, 3 (Sept. 2022), 257–291. <https://doi.org/10.1007/s12193-022-00388-0>
- [82] Jan B. Vornhagen, April Tyack, and Elisa D. Mekler. 2020. Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Canada) (*CHI PLAY '20*). Association for Computing Machinery, New York, NY, USA, 4–18. <https://doi.org/10.1145/3410404.3414229>
- [83] Hao Wang and Chuen-Tsai Sun. 2011. Game reward systems: Gaming experiences and social meanings.. In *DiGRA conference*, Vol. 114.
- [84] PiaoHong Wang and Zhicong Lu. 2023. Let's Play Together through Channels: Understanding the Practices and Experience of Danmaku Participation Game Players in China. *Proc. ACM Hum.-Comput. Interact.* 7, CHI PLAY (Oct. 2023), 413:1025–413:1043. <https://doi.org/10.1145/3611059>
- [85] Andrew M. Webb, Chen Wang, Andriud Kerne, and Pablo Cesar. 2016. Distributed Liveness: Understanding How New Technologies Transform Performance Experiences. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) (*CSCW '16*). Association for Computing Machinery, New York, NY, USA, 432–437. <https://doi.org/10.1145/2818048.2819974>
- [86] Jodi L. Whitaker and Brad J. Bushman. 2012. "Remain Calm. Be Kind." Effects of Relaxing Video Games on Aggressive and Prosocial Behavior. *Social Psychological and Personality Science* 3, 1 (Jan. 2012), 88–92. <https://doi.org/10.1177/1948550611409760>
- [87] Donghee Yvette Wohr and Guo Freeman. 2020. Audience Management Practices of Live Streamers on Twitch. In *ACM International Conference on Interactive Media Experiences* (Cornella, Barcelona, Spain) (*IMX '20*). Association for Computing Machinery, New York, NY, USA, 106–116. <https://doi.org/10.1145/3391614.3393653>
- [88] Donghee Yvette Wohr and Guo Freeman. 2020. Live streaming, playing, and money spending behaviors in eSports. *Games and Culture* 15, 1 (2020), 73–88.
- [89] Donghee Yvette Wohr, Guo Freeman, and Caitlin McLaughlin. 2018. Explaining Viewers' Emotional, Instrumental, and Financial Support Provision for Live Streamers. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI '18*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3174048>
- [90] Jifei Wu, Yimin Zhu, Xiang Fang, and Pramit Banerjee. 2024. Touch or Click? The Effect of Direct and Indirect Human-Computer Interaction on Consumer Responses. *Journal of Marketing Theory and Practice* 32, 2 (April 2024), 158–173. <https://doi.org/10.1080/10696679.2022.2158872>
- [91] Jackson Wyndow. 2022. *The subculture of the Lurker: participatory non-participation in digital subculture*. Master's thesis.
- [92] Yi Xu and Yixin Ye. 2020. Who watches live streaming in China? Examining viewers' behaviors, personality traits, and motivations. *Frontiers in Psychology* 11 (2020), 1607.
- [93] Saelyn Yang, Changyoon Lee, Hijung Valentina Shin, and Juho Kim. 2020. Snapshot-stream: Snapshot-Based Interaction in Live Streaming for Visual Art. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376390>
- [94] Andrew Young and Grant Wiedenfeld. 2022. A Motivation Analysis of Video Game Microstreamers: "Finding My People and Myself" on YouTube and Twitch. *Journal of Broadcasting & Electronic Media* 66, 2 (March 2022), 381–399. <https://doi.org/10.1080/08838151.2022.2086549>
- [95] Cong Zhang and Jiangchuan Liu. 2015. On Crowdsourced Interactive Live Streaming: A Twitch.TV-Based Measurement Study. In *Proceedings of the 25th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video* (Portland, Oregon) (*NOSSDAV '15*). Association for Computing Machinery, New York, NY, USA, 55–60. <https://doi.org/10.1145/2736084.2736091>
- [96] Ge Zhang and Larissa Hjorth. 2019. Live-streaming, games and politics of gender performance: The case of Nüzhubo in China. *Convergence* 25, 5–6 (2019), 807–825. <https://doi.org/10.1177/1354856517738160> arXiv:<https://doi.org/10.1177/1354856517738160>

- [97] Qun Zhao, Chun-Der Chen, Ho-Wen Cheng, and Jin-Long Wang. 2018. Determinants of Live Streamers' Continuance Broadcasting Intentions on Twitch: A Self-Determination Theory Perspective. *Telematics and Informatics* 35, 2 (May 2018), 406–420. <https://doi.org/10.1016/j.tele.2017.12.018>
- [98] Yuxiang Zhao and Jian Tang. 2016. Exploring the Motivational Affordances of Danmaku Video Sharing Websites: Evidence from Gamification Design. In *Human-Computer Interaction. Novel User Experiences*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 467–479. https://doi.org/10.1007/978-3-319-39513-5_44
- [99] Franziska Zimmer and Katrin Scheibe. 2019. What Drives Streamers? Users' Characteristics and Motivations on Social Live Streaming Services.