Toward a Twitch Research Toolkit: A Systematic Review of Approaches to Research on Game Streaming

Erik Harpstead  
Carnegie Mellon University  
Pittsburgh, PA, USA  
harpstead@cmu.edu

Juan Sebastian Rios  
New Jersey Institute of Technology  
Newark, NJ, USA  
jsr44@njit.edu

Joseph Seering  
Carnegie Mellon University  
Pittsburgh, PA, USA  
jseering@andrew.cmu.edu

Jessica Hammer  
Carnegie Mellon University  
Pittsburgh, PA, USA  
hammerj@andrew.cmu.edu

ABSTRACT
The rise of game streaming services has driven a complementary increase in research on such platforms. As this new area takes shape, there is a need to understand the approaches being used in the space, and how common practices can be shared and replicated between researchers with different disciplinary backgrounds. In this paper, we describe a formal literature review of game streaming research. Papers were coded for their research focus, primary method, and type of data collected. Across the prior work we found three common themes: (1) work that is readily supported by existing technical infrastructure, (2) work that does not require explicit technical support, (3) and work that would benefit from further technical development. By identifying these needs in the literature, we take the first step toward developing a research toolkit for game streaming platforms that can unify the breadth of methods being applied in the space.

CCS CONCEPTS
• General and reference–Surveys and overviews  
• Information systems–Multimedia streaming  
• Applied computing–Computer games

Author Keywords
Game streaming; twitch.tv; research method; literature review.

INTRODUCTION
Game streaming platforms, such as Twitch.tv and Youtube.live, have become popular in recent years. In 2018, there were 45 billion minutes watched per month on Twitch alone [54]. Beyond their popularity, these game streaming platforms afford new forms of interaction and engagement including professional esports [24] and audience participation games [46].

With the rise of game streaming platforms as services, there has been a corresponding rise in the academic study of these platforms. Key examples include ethnographic explorations of Twitch as a third place [16], big data analytics investigations of esports viewership trends [24], and surveys of viewer motivations for watching game streams [49]. Each of these investigations contributes to a growing body of research on these new environments and raises different methodological concerns and constraints.

One exciting aspect of the rise of game streaming as a medium is that it makes new forms of data related to gameplay available for study. Streamers make their gaming activity publicly visible, and hence available for analysis. Meanwhile, viewers provide a rich base for studying how people come to understand games from an external perspective. While this is a potential boon to researchers, there are several challenges that can make working with game streaming platforms difficult. For example, the sheer volume of data throughput from game video, streamer video, and chat can lead to difficulties in determining what to capture [59]. Further, overlaps in game and streamer audio can lead to challenges in transcription for data analysis [39].

To better understand this space and its unique challenges our team has begun to explore the development a toolkit for research on game streaming platforms, such as Twitch, that other researchers can use to expand their investigations and address pain points in the current research process. In service of that goal, this paper takes stock of the landscape of research on game streaming. We identify common themes in current approaches and potential gaps in the literature. After performing a formal literature review of game streaming research, we categorized papers in terms of their research focus, primary methodology, and type of data.
used to produce a picture of the space of research approaches to date.

We found that most papers either rely on public APIs that produce log data or use methods that do not require technical support such as ethnographies. On the other hand, few papers explore the relationship between viewer and streamer data, users’ system interactions beyond chat, or conduct experimental manipulations in game streams. We observe that these studies are poorly supported by existing tools for working with Twitch, and we discuss potential avenues for addressing these issues in the development of a research toolkit.

GAME STREAMING

Before describing our literature review procedure, we first describe the context of game streaming platforms in more detail and illustrate their value as research platforms. Game streaming platforms such as Twitch [52] and YouTube Live [57] allow streamers to share their gameplay in real time with live audiences [11,37]. A viewer typically sees a stream of the game being played, a camera view of the streamer themselves, and a chat window where they can discuss the game with other viewers (see Figure 1).

![Figure 1. An example of the Twitch.tv game streaming interface.](image)

Normally, during a streaming session, viewers can participate by entering text in a public chat [16]. Alternately, viewers can turn to external services to support interaction [1]. For example, the multi-channel chat services Discord is often used by streamers and viewers to communicate behind the scenes [12]. In response, streaming platforms are beginning to introduce new types of interactions, such as interactive overlays that a streamer can place over the video feed [53]. Some alternate interfaces incorporate gameplay data. For example, Helpstone provides contextual information about the game Hearthstone for stream viewers and lets audience participants suggest hints for the streamer [26]. Streamote allows audience members to place bets on the game using virtual currency [38]. Researchers have also begun to explore the space of designing new interactions for Twitch. TwitchViz helps connect game data with chat, focusing on post-game visualizations [34]. Other work beyond games is also exploring augmenting live streaming activities, such as Conversational Circles to help manage conversation in large streams [28] and Rivulet to make multi-stream viewing more participatory [17].

Game streaming platforms provide a space for hosting and searching for individual game streams. As the nexus of game streaming activity, they make ideal candidates to serve as platforms for study. Current streaming platforms include YouTube Live, Hitbox, Beam, Mixer, and others. While many streaming platforms are functionally similar, Twitch.tv is both the largest and has the highest proportion of game content [37,51].

Several features of the game streaming ecosystem make it interesting as a space for research. As noted above, streamers make their gameplay activities publicly available for analysis. The narration that typically accompanies gameplay can serve as a think-aloud, allowing researchers access to streamers’ thoughts and reactions as well as their actions in the game [3]. Meanwhile, viewers offer their own reactions to gameplay in chat. Streamers and their moderation teams must manage their communities of viewers using both live and automated moderation practices [44], and different channels develop unique subcultures [3]. Activities such as Twitch Plays Pokémon demonstrate a space in which to better understand the coordination of large live crowds [25]. All these features represent aspects of the streaming ecosystem that would be relevant and valuable to study.

METHODS

In performing our review we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [27]. Following the example of other systematic reviews in the broader game user research community [5,31,65] we identified 11 literature databases: ACM Digital Library, IEEE Xplore, DiGRA Digital Library, EBSCO Host, Infosci, ProQuest, ScienceDirect, Scopus, SpringerLink, Web of Science, and Sagepub. Additionally, we directly searched the proceedings and archives of relevant game studies venues that are often missed in common archiving processes. These included the proceedings of the Games+Learning+Society and Meaningful Play conferences, and the online archive of the International Journal of Computer Games Research at www.gamestudies.org.

To be included within our search a paper must be:

1. A peer-reviewed conference or journal article
2. Published between 2011 (the launch of Twitch) and May 2018 (the publication of the CHI 2018 proceedings)
3. Written in English
4. A study of some aspect of behavior in, on, or of a game streaming service
5. An empirical investigation in which data of some kind was collected
To reflect our focus on understanding research approaches to game streaming, we developed keyword list consisting of the names of well-known game streaming services as well as the phrase “game streaming” itself. Our final keyword list was: “twitch”, “game streaming”, “twitch.tv”, “twitch streaming”, “youtube live”. We did explore adding the names of some less well-known streaming services but found this did not yield any additional results not already covered by the existing keywords.

For each literature source, we searched using our keywords in any metadata field associated with a paper where applicable. One of the authors performed the queries and read titles and abstracts to decide whether to include a paper in the initial list. This initial pass focused on ensuring our first three inclusion criteria were met.

Search queries were performed between June 2018 and July 2018. A total of 922 papers were returned by the initial search, 91 abstracts were collected after inspecting the results for obvious mismatches from title or publication venue (the “twitch” keyword tended to return medical results related to twitch in muscles), which yielded a list of 55 candidate papers. All authors then had the chance to read through the papers in detail to determine whether a candidate paper met the rest of our inclusion criteria or was a duplicate of another paper. This detailed pass process eliminated 9 candidate papers resulting in a final set of 46.

**Coding**

Once we had a corpus of papers, we used an iterative coding process where members of our team independently coded a subset of papers and then discussed disagreements in codes before iterating on a final coding manual. The final corpus of 46 was small enough that all researchers were able to review all papers. Coding was iterated until we agreed on all categories.

Each paper was coded along three independently-developed dimensions: Focus, Primary Method, and Type of Data.

**Focus** - In coding for the research focus of the paper we were interested in whether the work was primarily interested in the behavior/experience of game streamers, viewers, both viewers and streamers, or broader platform behaviors (i.e., population level trends). In cases where this focus was not explicitly stated, we judged which of the options was most closely addressed by the work. Further, when work was related to a subset of one of the populations, we considered the focus to be for the broader set (e.g., work looking at the experience of moderators was considered a viewer focus).

**Primary Method** - The codes for Primary Method described the high-level methodological approach taken by the work. To reduce bias induced by our own methodological background in having to define what was research and what were valid research methods, we decided to ground ourselves using a specific set of methods accepted within HCI. We based our coding of methods on the book *Ways of Knowing in HCI* [33], in which each chapter explores a different methodological tradition in the field. The approach was to decide which chapter best described the work presented in each paper. While not every chapter of the book was relevant to our corpus, this coding scheme proved sufficient to be applied to every paper in our corpus. The final set of Primary Method codes we employed were: Ethnography, Experiment, Eye Tracking, Grounded Theory, Log Data Analysis, Research through Design, Sensor Research, Social Network Analysis, Survey Research, and Technical HCI.

**Type of Data** - The codes for Type of Data described the kinds of data being collected and analyzed in each paper. These codes were collaboratively developed among our team using an iterative process. Further, unlike the Focus and Primary Methods codes papers could be coded with multiple types of data. The codes for this dimension where:

- **Field Notes**: Any data where the source is notes taken by the researcher in a field experience setting. While these notes often described aspects of a stream or chat, the stream and/or chat data were described rather than scraped or otherwise recorded programmatically.
- **Interview**: Any data arising from a direct interview with a user or participant. We made no distinction for whether the interview was structured or free form.
- **Questionnaire**: Any form of questionnaire presented to a user. Commonly this represented survey methods, but it would also include psychometric instruments used in the context of an experimental manipulation. No distinction was made for whether these questionnaires were administered online or on paper.
- **Stream Data**: This code was used for any case of data arising from the stream itself. This would include any analysis of video content (either through computer vision or manual coding) or of data from the game being broadcast (e.g., an experiment where researchers record actions from a game that is also being broadcast).
- **Chat Data**: Any use of data from the chat stream. In the majority of cases this involved textual analysis of the content of chat messages but would also include metadata like participation rates. Often this data was collected through some kind of IRC client.
- **Interaction Data**: Used to describe any kind of interaction with the interface that was neither the video stream nor chat. We also used this code to encompass auxiliary sensor data like eye tracking because such cases were too rare to warrant their own code.
- **Platform Data**: Any data derived from the broader game streaming platform that is not directly related to user interaction. Examples would include channel listings, network load data, or broader user demographic information. Often this data was collected through an official API.
Our final corpus (see Table 1) contained 46 papers from 125 unique authors. There was a single paper in the corpus [25] that described multiple investigations using different primary methods, which we consider as two separate studies in our analyses, resulting in a final set of 47.

Looking at the trends of publication over time (Figure 2), we can see the recent increase in interest on the viewer experience, and a relatively recent sustained interest in streamers. While much of the early work on game streaming focused on the broader behavior of platforms, this work has tapered off in recent years in favor of a user orientation.

For the Primary Method dimension (Figure 3) there is a steady presence of log data analysis over time (two per year on average). Further, ethnography has also had a one or two cases per year. More recently, survey methods have grown in popularity with four instances in each of the past two years, and the overall set of methods being applied to game streaming has become more diverse with 2017 containing one instance of each method except Social Network Analysis and Sensor Research.

The trends in data type usage (Figure 4) show that the use of platform data has become less common in since 2016. Further, there is notably little use of interaction data (3 total papers, all in 2017). Finally, there is a rise in the use of questionnaires (12 papers in the past 2 years) to go along with the corresponding rise in the use of survey methods.

Table 1. Our corpus with final codes. *Note that [25] described multiple investigations using different primary methods.

<table>
<thead>
<tr>
<th>Year</th>
<th>Reference Number</th>
<th>Focus</th>
<th>Primary Method(s)</th>
<th>Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>24</td>
<td>P</td>
<td>LDA</td>
<td>PD</td>
</tr>
<tr>
<td>2014</td>
<td>16</td>
<td>S</td>
<td>ETH</td>
<td>IN; FN</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>B</td>
<td>LDA</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>P</td>
<td>LDA</td>
<td>PD</td>
</tr>
<tr>
<td>2015</td>
<td>7</td>
<td>V</td>
<td>SEN</td>
<td>CD; SD</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>P</td>
<td>LDA</td>
<td>PD</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>P</td>
<td>LDA</td>
<td>PD</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>V</td>
<td>LDA</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>P</td>
<td>LDA</td>
<td>PD</td>
</tr>
<tr>
<td></td>
<td>59</td>
<td>P</td>
<td>LDA</td>
<td>PD; SD</td>
</tr>
<tr>
<td>2016</td>
<td>4</td>
<td>S</td>
<td>SUR</td>
<td>QN; PD</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>P</td>
<td>SNA</td>
<td>PD</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>P</td>
<td>LDA</td>
<td>PD; SD</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>V</td>
<td>LDA</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>S</td>
<td>TH</td>
<td>CD; SD</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>S</td>
<td>ETH</td>
<td>FN</td>
</tr>
<tr>
<td></td>
<td>61</td>
<td>P</td>
<td>TH</td>
<td>PD</td>
</tr>
<tr>
<td>2017</td>
<td>2</td>
<td>B</td>
<td>GT</td>
<td>FN</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>S</td>
<td>GT</td>
<td>FN; SD</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>V</td>
<td>EYE</td>
<td>SD; ID</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>V</td>
<td>LDA</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>V</td>
<td>SUR</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>V</td>
<td>SUR</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>S</td>
<td>GT</td>
<td>IN; FN</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>V</td>
<td>TH</td>
<td>QN; IN; ID</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>V</td>
<td>EXP / GT*</td>
<td>CD; SD</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>V</td>
<td>EXP</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>B</td>
<td>ETH</td>
<td>CD; SD</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>V</td>
<td>EXP</td>
<td>QN; IN; CD</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>S</td>
<td>EXP</td>
<td>IN; ID</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>V</td>
<td>RTD</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>V</td>
<td>LDA</td>
<td>CD; PD</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>V</td>
<td>SUR</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>V</td>
<td>SUR</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>P</td>
<td>TH</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>S</td>
<td>ETH</td>
<td>FN; IN</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>P</td>
<td>TH</td>
<td>PD</td>
</tr>
<tr>
<td>2018</td>
<td>6</td>
<td>S</td>
<td>SUR</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>S</td>
<td>ETH</td>
<td>IN; FN</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>V</td>
<td>SUR</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>S</td>
<td>ETH</td>
<td>IN; FN</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>B</td>
<td>LDA</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>V</td>
<td>SUR</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>B</td>
<td>LDA</td>
<td>PD</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>S</td>
<td>SUR</td>
<td>QN</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>S</td>
<td>SNA</td>
<td>PD</td>
</tr>
</tbody>
</table>

Table 1. Our corpus with final codes. *Note that [25] described multiple investigations using different primary methods.

RESULTS

Our final corpus (see Table 1) contained 46 papers from 125 unique authors. There was a single paper in the corpus [25] that described multiple investigations using different primary methods, which we consider as two separate studies in our analyses, resulting in a final set of 47.

Looking at the trends of publication over time (Figure 2), with the exception of 2013, there has been a steady increase in game streaming research over the years. Notably, the number of game streaming papers more than doubled (from 7 to 20) between 2016 and 2017. These trends suggest a healthy field that is continuing to grow.

Looking more closely at trends in research Focus (Figure 2), we can see the recent increase in interest on the viewer experience, and a relatively recent sustained interest in streamers. While much of the early work on game streaming focused on the broader behavior of platforms, this work has tapered off in recent years in favor of a user orientation.

Figure 2. Distribution of Research Foci over time.

For the Primary Method dimension (Figure 3) there is a steady presence of log data analysis over time (two per year on average). Further, ethnography has also had a one or two cases per year. More recently, survey methods have grown in popularity with four instances in each of the past two years, and the overall set of methods being applied to game streaming has become more diverse with 2017 containing one instance of each method except Social Network Analysis and Sensor Research.

Figure 3. Distribution of Primary Methods over time.

The trends in data type usage (Figure 4) show that the use of platform data has become less common in since 2016. Further, there is notably little use of interaction data (3 total papers, all in 2017). Finally, there is a rise in the use of questionnaires (12 papers in the past 2 years) to go along with the corresponding rise in the use of survey methods.

Figure 4. Distribution of Data Types over time.
Table 1 shows co-occurrence patterns that appeared in at least two different research patterns from our review that have support further research on game streaming, we identify capture a more holistic picture of research approaches.

In addition to individual trends, we also looked at the co-occurrence of codes across the different dimensions to capture a more holistic picture of research approaches. Table 1 shows co-occurrence patterns that appeared in at least two different studies within the corpus. Some common themes are evident from these patterns. For example, surveys, primarily of viewers but also of streamers, are a common approach. There were several studies coded as Log Data Analysis, which leveraged platform data, either alone or in conjunction with other sources, to explore broader platform trends. Finally, several ethnographies have been performing studies of the streamer experience, while none have yet looked at the experience of viewers.

<table>
<thead>
<tr>
<th>Focus</th>
<th>Primary Method</th>
<th>Data Type</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewers</td>
<td>Survey Methods</td>
<td>Questionnaire</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>Platform</td>
<td>Log Data Analysis</td>
<td>Platform Data</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Streamers</td>
<td>Ethnography</td>
<td>Field Notes + Interview</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Viewers</td>
<td>Log Data Analysis</td>
<td>Chat Data</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Both</td>
<td>Log Data Analysis</td>
<td>Chat Data</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Platform</td>
<td>Log Data Analysis</td>
<td>Platform Data + Stream Data</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Streamers</td>
<td>Survey Methods</td>
<td>Questionnaire</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Platform</td>
<td>Technical HCI</td>
<td>Platform Data</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>26</td>
<td>55</td>
</tr>
</tbody>
</table>

To support our long-term goal of developing a toolkit to support further research on game streaming, we identify three main research patterns from our review that have varying needs for support through additional research tools. First, what we call technically supported studies are studies that use methods and data that are generally available through the use of web scrapers or calls to public APIs (e.g. [24,30,59]). Second, technically agnostic studies use a game streaming platform to communicate, but do not necessarily need additional technical support (e.g., [10,16,23]). These studies primarily employ ethnographic or survey techniques that are adaptable to many technical contexts without an explicit need for software.

The final category is technically challenging studies that are poorly supported by the existing technical landscape of game streaming research. These studies include explorations of the relationships between streamers and viewers, investigations with user interaction data beyond chat, and experimental interventions involving game streaming systems. This category of under-supported methods would benefit most from the development of a research toolkit.

If we hope to foster new research using these under-supported methods, we must understand the challenges they currently face. For example, given that viewers and streams are free to come and go, their interactions are often irregular and serendipitous making them difficult to track from a single side of the interaction. Multimodality also poses a challenge to studying streams and viewer interactions. Streamers in particular possess many avenues for interacting with their audience including replying in textual chat, speaking over the video stream, or even altering their in-game behaviors. Finally, the relationships between viewers and streamers take place within complex subcultures that can be unique to different streaming channels and may be hard to account for in current large-scale analyses.

Research involving interaction data, on the other hand, faces challenges around the instrumentation of interaction data in the game streaming context. While web scrapers and calls to APIs can provide a significant amount of data, the interfaces that users use to interact with game streams and that streamers use to interact with viewers are often proprietary and federated. For example, the actions taken by Twitch’s “AutoMod” feature, which streamers may set to assist with moderation, cannot be scraped using public methods, but their role in engaging with viewers and managing a community is important to understand. These barriers can make it difficult to instrument systems to record interaction data and preserve users’ context as they transition between applications. While these challenges may be somewhat less difficult in studies with streamers, as researchers can foster relationships on an individual basis, they are extremely difficult when trying to study the interaction patterns of viewers.

In addition to technical challenges there are several complex ethical issues in capturing interaction data from large anonymous and pseudonymous crowds. For example,
if a study is being run on a live broadcast channel, how should new audience members be introduced to the study? What if a new viewer is a minor or a member of some other protected class? How, if at all, should users be made aware when messages are being collected at-scale, as in [44]? How should participants be handled when the treatment of their data may be subject to different laws depending on their country of origin?

Controlled experimental studies on game streams pose further challenges over and above those related to instrumentation. One such challenge is in assigning individual participants to conditions. If the causal impact of an experimental intervention is to be determined, users must be assigned to conditions and users within conditions must experience the same intervention. However, this carries a trade-off with ecological validity - on Twitch, users do not naturally stay within “conditions”; they enter and leave at inconsistent intervals, likely meaning that many participants will join midway through a study. Moreover, users make new accounts to play with different features, which could plausibly lead to users being in multiple conditions at once in a way that could not consistently be tracked. Realistically measuring impact of an intervention requires reconsidering the restrictions of standard experimental procedures.

Further, there are several challenges to presenting stimuli to viewers conditionally with varying degrees of complexity. For example, providing different viewers with different chat or interface experiences could be implemented with viewer overlays, while providing them with completely different video streams would require significantly more technical investment. In both cases, user engagement with stimuli depends on their ability and willingness to overcome the minor or moderate technical hurdles of installing, approving, or navigating to an intervention.

Existing features of the game streaming context could be adapted to address some of these issues. For example, recently introduced stream extensions [53] could ease much of the instrumentation burden and enable conditional presentation of stimuli in experimental contexts. Providing tools to build relationships with moderators [47] could help mitigate subcultural concerns in understanding streaming communities. Finally, leveraging chatbot systems and other automation techniques [43] could facilitate the research introduction process for viewers, helping to address some of the ethical concerns with doing research in a public virtual space.

Understanding existing research methods that are under-supported can help us identify potential features for a future research toolkit; however, it does not provide us with a clear picture of how these features should be built to integrate with existing research workflows. Future work could build upon these findings working with researchers and developers to better understand their specific needs, both for how to improve the process of data collection on game streaming platforms and also how to help make it usable for analysis. Indeed, such work may surface further challenges within some disciplines that are unlikely to be addressed by a technical solution alone.

Another limitation to our current work is the heavy emphasis on Twitch as the main platform of game streaming and by extension, game streaming research. While other game streaming platforms do exist in the market, the relative scarcity of research done on those platforms makes it hard to judge how generalizable our findings would be to the broader space of game streaming at large. Future work building on our findings would do well to consider the broader scope of game streaming and non-game streaming in identifying researchers’ needs for tools and best practices.

CONCLUSION
In this paper we have undertaken a review of game streaming research to better understand the challenges the community faces, with the long-term goal of developing a toolkit to facilitate such research in the future. We found several kinds of research that are poorly supported by existing infrastructure, particularly research that studies the relationship between game streamers and their audiences, investigations that make use of broader interaction data, and studies that employ game streaming in the context of an experimental intervention. We believe there is a rich space of possibility in addressing some of the methodological issues in game streaming research and look forward to collaborating with the community to support the future of research on Twitch.

ACKNOWLEDGMENTS
We would like to thank the reviewers for their valuable comments on earlier versions of this work. We are grateful to Amazon for their generous support of our work.

REFERENCES
https://doi.org/10.1177/1354856517736974


