CATALYST

Toxicity v. toxicity

How Ethnography Can Inform Scalable Technical Solutions

JAMIE SHERMAN, PhD, *Intel Corporation*
ANNE PAGE MCCLARD, PhD, *McClard LLC*

While a number of scholars have studied online communities, research on games has been mostly focused on the business, experience, and content of gameplay. Interactions between players within games has received less attention, and toxic behavior is a newer area of investigation in academia. Inquiry into toxicity in gaming is part of a larger body of literature and public interest emerging around disruptive and malicious social interactions online, cyberbullying, child-grooming, and extremist recruiting). Through our research we reaffirmed that toxicity in gaming is a problem at a global scale, but we also discovered that on a micro scale, what behavior gamers perceive as toxic, or how toxicity is enacted in gaming is different depending on cultural context amongst other things. The generalized problem at scale, and its particular manifestations on the micro level raise philosophical and technology design questions, which we address through examples from our own research and its applications in the industrial settings.

Keywords: Ethnographic research, Culture, Gender, Technology, Toxicity, Scale, Community, Global, Te

INTRODUCTION

Over the last year and a half, we—two anthropologists, one working for Intel, and the other a consultant—engaged on a variety of research topics related to digital game playing. We executed numerous multi-cultural ethnographic studies, each with a similar structure, and each building upon the previous studies. The basic design for each study included: 1) an online open-ended survey screening interview, 2) selection of participants based on stakeholder-defined criteria, 3) an introductory interview, 4) a week-long diary documenting daily life, gaming activities, frustrations encountered, social interactions, etc.; and 5) a follow-up interview. A majority of interviews were conducted in people's homes, but due to the COVID_19 pandemic, some were moved to online. In all, we conducted interviews with 49 participants across all studies.

The nominal focus for these studies was broad: to understand motivations, practices, needs, and "pain points" of different kinds of gamers in different places. We documented cultural themes around gender, social dynamics of teams, daily practices, consumption and engagement with game-related media, and attitudes and understandings of gaming as an activity and, for some, an identity. In parallel, and leveraging the "thickness" of ethnographic data, we asked questions and paid close attention to issues of "toxicity" or disruptive, unpleasant, and harassing behavior between players in gaming and game-related activities. This line of inquiry was motivated by a parallel project within the gaming team at Intel focused on how to reduce toxicity in gaming through AI moderated voice chat. That project—design of a real-time, scalable technical solution for managing toxicity in a game

voice chat system—is in progress as of this writing, fed in large part by the findings of our broader studies, with deeper investigation of voice chat interactions in games to come.

While a number of scholars, notably anthropologists, have looked at online communities (Boellstorff 2008; Nardi 2010), the study of games has, until relatively recently, been largely focused on the business, experience, and content of gameplay (Dyer-Witherford and de Peuter 2009). Interactions between players within games has received less attention, and toxic behavior is a relatively new area of investigation in academia, largely due to events like "Gamergate"(see Dewey 2014; Gray et. al. 2017). Inquiry into toxicity in gaming is part of a larger body of literature and public interest emerging around disruptive and malicious social interactions online, cyberbullying, child-grooming, and extremist recruiting (Adinolf and Turkay 2018, Fredman 2018).

Through our research we reaffirmed that toxicity in gaming is a problem at a global scale, but we also discovered that on a micro scale, what behavior gamers perceive as toxic, or how toxicity is enacted in gaming is different depending on cultural context amongst other things. We found that toxicity characterizations reflect the tensions, cultural beliefs, and attitudes within the local communities, and that these characterizations can even differ within the same geographic location between sub-groups. What gamers consider acceptable behavior in online voice chat varies dramatically by locale, by the game community of a particular game (e.g. PUBG, League of Legends, Overwatch, etc.), by type of player (e.g., competitive team player, recreational player), and even by ones gender, ethnicity, language, or race.

We raise three key questions in this paper, for which we do not promise definitive answers. However, we discuss each of them in the context of how our ethnographic work intersects and illuminates our technological pursuits:

1. What does scale mean? What is the relationship between the objectives of technological development at scale and the ethnographic project as an investigation into patterns of culture?
2. To what extent can ethnography's focus on the relationship between the local and the global facilitate generalizable recommendations for developing technical solutions? Is it possible to create technologies that are sensitive to local distinctions, and yet scale globally?
3. How does one go about identifying which aspects of a problem are important at a micro level, but not so important at a macro level? Is there a way to get to the core underlying problem and generalize it?

Throughout, we use examples from our own research, other's research on related topics, and refer to an ample body of anthropological theory that pertains to these questions.

WHAT DOES SCALE MEAN?

"Scale" has many meanings, so we begin our discussion with some of the different ways we think of it. One common use of scale from a linguistic perspective refers to the relative size of something as measured against some standard unit of measure—inches, feet, centimeters, number of people, and so forth. To scale a drawing, for example, means to keep the proportions the same when one is reducing a very large something to a smaller

something. All of the ways we present scale here pivot in some way on this general definition.

In the tech industry, one meaning ascribed to scalability in a technological solution is that it can serve the few or the many, technology that can grow with the needs of a business, for example. Another way of thinking about scalable technology is in how flexible it is, how easy it is to change, to add on to, or modify. And then, from a business perspective, scale is almost always about how many people will purchase or benefit from a technology—market size is of highest importance. Scalability of technology is important to the business because it increases market size ultimately.

Likewise, scale in research means different things. It can literally refer to the number of people included in a study, so a small-scale study might have as few as three people, and a large-scale study might include thousands to hundreds of thousands of people. Scale in research might also refer to the focus of study—micro or macro, local or global. All of these definitions of scale have relevance and implications in our work, and at times, the goals of scalability in each of these domains can be at odds with one another.

Scale in Research

Our job as ethnographic researchers in the tech industry is to generate new ideas, inform the design, and guide marketing of technologies that will appeal to the greatest number of people possible. Therefore, our research serves both notions of business scaling and technological scaling. Scale also comes into play for research itself, and how research is conducted, at what scale; some of the various methods that ethnographers use scale up better than others.

Participant observation is arguably the linchpin methodology of ethnographic inquiry, and is always present in ethnographic research; it requires deep dives into the daily lives of individuals, groups, and communities. Ethnographers, however, use a plethora of methodologies in addition to participant-observation, including, but not limited to, surveys, diaries, experimental frameworks, photo-elicitation, telemetry, and essentially any method that will help answer the questions that are in scope. Each method employed in ethnographic inquiry has different scaling characteristics.

By way of example, intensive participant observation research is not easily scaled to include as many participants as survey research. Not only are large-scale participant-observation studies onerous for the amount of data they generate, but they are prohibitively labor-intensive and too costly for most companies to justify executing them. This doesn't mean that they are not valuable; it just means that this key aspect of ethnographic research is usually done at a small scale. One question every practicing ethnographer gets is how does one determine whether ones findings are spurious given the typically small sample sizes of participant-observation research? There is not an easy answer to that question. The quality of the research depends on multiple factors, among them: how good the researchers are, how good the recruited participants are, how good the line of inquiry is, and how broad a net is cast. Fortunately, ethnographers practicing in industry do not rely solely on participant observation. We would argue that participant observation is the method with the most explanatory power of all of the methods we use, which is why it is so critical. Without it, and the deep insights it reveals about people's beliefs, views, and behaviors, other data in our tool box would be of limited value.

Let's consider survey research for a moment, one of the tools in an ethnographer's research kit. It is very scalable, and from a corporate perspective gives the best bang for the buck; it is relatively fast to collect the data, and per-participant costs are lower than for participant observation. One can ask as many or as few questions as one wants to, of as many or as few people are required for statistical purposes, and in theory, statistically significant findings imply generalizability and have predictive value. Large sample sizes, unfortunately, give business decision-makers a false sense of security in the validity of the data. Because, as with any research, the data is only as good as the researcher, how good the questions are, and how good the survey respondents are (which is a story for another day). The biggest problem with survey research is that taken on its own—without participant observation--it gives little to no insight into people's beliefs, views, and behaviors, because the responses are completely devoid of context and it is impossible to know anything about the thought processes that went into giving a particular response.

With all of this said, different business and technology related goals, require different research methods to achieve them, methods at different scales. For example, one cannot design a new product based on a large-scale market research segmentation. Products need to be designed with individual whole human beings in mind, with a detailed understanding of individual daily lives, motivations, situations, goals, beliefs, biases, etc., which is one area in which participant observation excels. The converse of this is that one cannot understand market sizing, price sensitivity, or build predictive models solely by doing participant observation. That said, participant observation and other qualitative methods used in ethnography are invaluable for properly framing large-scale quantitative studies.

Scale from the Business Perspective

From a business perspective scaling almost always refers to growing the market. The question is always, "How do we get more customers to purchase our product"? At Intel, this is ultimately about selling more chips, whether direct to consumers who build their own PCs, or to "Original Equipment Makers" (OEMs) who make and sell laptops and desktops with Intel chips inside. Over the past decade or so, compute options have diversified with smartphones and tablets now capable of nearly everything one once needed a PC to accomplish. In addition, overall penetration of PCs in the overall market has increased, while the general compute needs of most users is met or exceeded by the devices they already own or have access to. Thus while the extent to which the PC is dead or dying, as some predict, is very much up for debate, the overall market for PCs is no longer seeing the kinds of growth it once did, and consumers now have a wider array of options.

High compute needs of scientific and enterprise data analytics aside, video game play remains one of the most visible spaces where users and game developers continue to push the boundaries of what current devices can do. In other words, it is an important growth segment in a market that is otherwise relatively lackluster and under threat. For that reason, Intel is especially interested in protecting and cultivating this market of enthusiast users who spend more on their systems and renew or upgrade their equipment more often than other consumers. In that context, it became important for the Intel gaming team to better understand the *scale* of the problem of toxicity in gaming as an impact on the market. They wanted to know the extent to which harassment and abuse lead to otherwise enthusiast gamers choosing other kinds of hobbies versus switching to, say, other kinds of games, while

still enjoying video games overall and continuing to invest in it as a hobby. That question was one we decided to tackle through a quantitative study of current and former gamers that is was informed by our ethnographic work, but not part of the discussion here.

In addition to enthusiast gamers, Intel has known for some time that nearly everyone who owns a PC of any kind plays games on them. One of the questions that Intel's gaming team was interested in, then, was how we might encourage these "casual" or "mainstream" gamers to become enthusiast gamers, and what kinds of barriers are there that we could do something about. Thus in early research we actually focused on these kinds of mainstream, non-enthusiast gamers, exploring how gaming fit into their everyday lives, the choices they were making in terms of whether and when to play, what to play, and what device or platform to play on. In addition, we asked a number of questions about their affective relationship to gaming, and their identification with gaming, as well as their experiences with, and concerns about toxicity and harassment. What we found in that work was that mainstream gamers, if they played with others, tended to limit their play to family and pre-existing friendships such that toxicity, which tends to occur between players who don't know each other from in-person contexts, was of less concern. However, we also found that mainstream gamers actively shied away from identifying as "gamers" in part because they didn't want to be seen as "one of those people." Thus while they reported fewer bad encounters and worried less about harassment, the overall negative association of gaming with toxicity and other kinds of bad behavior did in fact represent a barrier to gaming.

Getting back to "scale," from the business perspective, then, scale is best captured in terms of the number of purchasers, the frequency of purchase, and the relative value, or profit, in that purchase. From the business point of view, more is nearly always better: more people (size) buying more often (frequency) and buying more expensive models (in Intel parlance "upsell").

SCALE IN ANTI-TOXICITY CHAT ALGORITHM

Background: Toxicity in Gaming

Social anxiety around the content of video games cropped up as early as the 1970s, when the game Death Race was pulled off the shelves due to public outcry over its depiction of running over pedestrians for points, but it wasn't until 2014 or so that interactions between gamers came under significant discussion. Far from the first example of harassment of women in video games, Gamergate was nonetheless amongst the first to garner widespread attention. A disgruntled ex-boyfriend posted a host of accusations against an independent game developer accusing her of a range of things, including trading sex for positive reviews of her video game. The man's post sparked a coordinated campaign of harassment and intimidation that went viral, spreading from the woman herself to her known associates, and to other women writing and posting about video games. The harassment of these women included rape threats, death threats, coordinated email campaigns against websites, revealing photos emailed to employers and relatives, and public posting of home addresses and personal information in a tactic now commonly known as "doxing."

While Gamergate played out primarily across non-gaming social media platforms such as Twitter, YouTube, reddit, and 4chan, the notoriety of the campaign brought increased attention to the culture of sexism and racism in the video game industry, where women and

people of color are radically under-represented, systemically excluded, and routinely harassed. In addition, it brought heightened awareness to player interactions, where likewise women, people of color, LGBTQ+, and other minority communities are routinely and often quite viciously attacked (see here, here, and here, for just a few examples).

A Scalable Technological Intervention

Against this backdrop, in early 2018, Intel drove a series of forecasting efforts focused on video game play. In that work, and in a range of subsequent studies including our own, toxicity between players emerged as a top concern for game players and was identified as a potential damper on market growth (limiting the scale of the market). The 2018 findings effectively shifted perceptions within the company from seeing player harassment in games as a moral and ethical problem facing individual players and game companies to perceiving player to player harassment as a business issue with potential implications for profit and sales, and a technical question regarding how and what we might "solve" using technological means.

While cognizant that the challenge is complicated, where harassment is frequently multi-modal and multi-platform, anti-toxicity work at Intel approaches the problem with a framework that is explicitly scaled, framed in terms of "crawl-walk-run." In this context, "crawl-walk-run" is an elastic concept applied both to the overall challenge of addressing social issues like this one by means of technical solutions, and more narrowly in terms of how a technical solution works by setting reachable goals ("minimum viable product") with the intention to add, grow, and improve over time.

From an overall domain perspective, team leadership focused in on voice chat as a key area for technical intervention. One of the (many) challenges in addressing toxicity is in identifying when and where it is happening. Today, most games rely on a combination of algorithms that screen text exchanges for transgressive vocabulary determined by individual game or organizational standards, and on users reporting of other user transgressions. The sheer volume of interactions, as well as the way that these interactions are fragmented across gaming and social media platforms make identifying incidents challenging. The Intel team began to focus on voice chat because, while text chatting is ubiquitous, there are existing tools for screening, and text itself leaves written records that game companies and moderators can review. By way of contrast, voice chat is transitory, the compute costs and privacy challenges of recording all exchanges prohibitive, and there are few to no existing tools. Here, the aspect of scale (or crawl in relation to walk and run) is about the focus on a relatively narrow context: screening voice chat interactions for "toxicity."

Within the voice chat anti-toxicity project, there are also internal notions of scale that are both additive and progressive. Initial efforts toward a "minimum viable project" focus primarily on the textual meanings of expressions and words identified by the algorithm in English (currently capturing UK and US based expressions). Future plans include areas for "scaling" such as the addition of community specific expressions, adding new game communities, adding new languages, and drawing on more vectors for analysis for better identification of the emotional tenor of an exchange – aspects such as tone, velocity, or decibel level have been mentioned.

Yet a third aspect of scale applies to the ways the team envisions the algorithm being used and by whom. Identifying offensive speech can be used by gamers to screen or silence

others, to document or report others for offensive behavior, and by gaming and game-related platforms to "triage" interactions for human moderation review, or to take automated action for specific kinds of transgression. The various ways the underlying capability can be applied, whether through unique applications or bundled into a single application with multiple features and stakeholders, represent an aspect of scale linked to, but not quite the same as, notions of scale embedded in the business perspective.

LOCAL SENSIBILITIES V. SCALABLE TECHNOLOGY

This paper draws on a series of 4 in person and remote ethnographic studies conducted over the course of a year from March 2019 to March 2020. In addition to these studies, one of us conducted a series of 8 informal interviews with employees in the US who self-identified as enthusiast gamers, and fielded a quantitative survey with digital game players, non-players, and former players in the US to better understand overall patterns in leisure time practices and decisions to play or not play digital games, including the role of harassment or toxicity in games.

The first of the ethnographic studies in March of 2019 focused on "mainstream" and "casual" players in the Greater Los Angeles, California area. In that study we were particularly focused on better understanding the choices these players made across gaming devices, and the factors that made them more, and less likely to play overall. For the purposes of that study we defined "mainstream and casual" as players who spent less than 5 hours per week playing digital games of any kind on any device.

The next two studies, conducted in China and India looked at more enthusiast gamers, people who played at least 5 hours per week on average, and whose primary platform for gaming was either a PC or a console such as Microsoft Xbox or Sony Playstation, or both. In that study we were similarly interested in deeper understandings of motivations to play or not play, and in platform and device choices throughout the week.

The final study in this series was initially planned for Katowice, Poland, to coincide with a major e-sports event, and focused on competitive players and fans attending that event. In wake of the emergence of Covid-19 as a global health concern (not yet considered a pandemic at the end of February 2019), we pivoted to a mix of in person and online virtual meetings with participants in Portland Oregon, and several we had already recruited and who are based in the US, Europe, and Iceland. By the time we planned and executed this study, focus for the Intel gaming team had shifted toward Intel's "traditional" market of enthusiast and competitive gamers and this study was specifically designed to dive deeper into the specificity of competitive e-sports gamers.

Across these four studies, we spoke to a total of 49 video game players from diverse backgrounds, abilities, and levels of engagement with video game play in 5 countries. A few broad patterns emerged. Gamers of all kinds and abilities perceived toxicity and harassment as a critical problem in a guided question, part of the 7-day diary. In daily un-guided tracking of challenges and frustrations throughout the week, competitive gamers were far more likely to mention unpleasant encounters with other players. Mainstream and casual gamers who were less likely to play with people they did not know well in person did not cite toxicity or harassment in any of the sessions tracked in the study and were much less likely to mention personal experiences of that kind in the ethnographic and conversational interview portions of the studies.

In our interviews where toxicity emerged as a primary concern for participants, we were struck by all of the different ways gamers talked about and experienced toxicity, and the fact that what was considered toxic in one place was not considered toxic in another place. Furthermore, we found that even within a geographic location, that the very essence of toxic behavior could differ based on regionalism, language, or ethnic identity. Below, we tell three stories about toxicity in three countries—the US, China, and India. Each story illustrates a significant local sensibility about toxicity that we didn't see in the other countries. With the exposition of these stories, we challenge the reader, and the EPIC audience to think about the following questions:

1. To what extent can ethnography's focus on the relationship between the local and the global facilitate generalizable recommendations for developing technical solutions?
2. Is it possible to create technologies that are sensitive to local distinctions, and yet scale globally?

As a reminder, one of the primary directives from the company, was to assist in the development of the anti-toxicity chat algorithm we described before, a scalable technology that was being designed for scalable business purposes. We will discuss some of the thought processes and recommendations that have emerged based on our research.

US STORY

Princess is an experienced competitive player in her late 20s who began playing video games with her father and at a young age. She has participated in competitive e-sports for several years as both a competitor and as coach of a local youth team. While the rampant sexism she encounters in gaming has made her more determined to play and to win, she describes it as exhausting and frustrating. Over time, she has developed tactics for avoiding toxic interactions directed at her. She uses a screen name that is non-gender specific. She generally avoids voice chat altogether, even in games where it is useful to the game, telling others that her microphone is broken (it's not). When she does use voice chat, she told us that when other players hear her, they frequently ask if she is "a girl or a squeaker" where squeaker means a pre-pubescent boy. Often, she said, she answers "squeaker" as she gets less hassled than if they think of her as a woman.

The way that Princess avoids voice chat was echoed by other women we spoke to in the United States, and men and women alike said that they prefer to use third party voice chat services such as TeamSpeak and Discord (today, primarily Discord) where they have more control over who is on the chat in lieu of in-game voice features. At the same time, using such platforms does not necessarily prevent players from experiencing toxicity, and in some games, in-game voice chat is an important element in the gameplay itself, forcing players to choose between exposure and success in the game.

In general, in the weighing of success vs the risk of exposure to offensive behavior, winning is frequently seen as more important, with toxicity as the price to pay. As Courtney, another female player in her 20s put it,

> it's one thing if you're talking offensively but you're still performing well; unfortunately, because of how toxic the overall community of League of Legends is, most people are like, eh, they said really racist stuff all game. But they contributed.

On the other hand, for Courtney, and others we spoke to, toxicity was experienced not just through language in speech or text, but through actions. She gave us the example of a team member who effectively sabotaged the match by actively helping the opposing team, or by failing to help the team. This kind of behavior, as she said "hits a lot harder" because losing a match has consequences not just for that day, but can also impact ones ranking and resources in the game overall.

Figure 1: In the US, sexism and racism are at the center of US players' experiences of toxic behavior in gaming. These experiences are tempered by perceptions of gaming overall, where some poor behavior is to be expected and tolerated.

While Courtney told us this story in the context of harassment she had experienced, for some users, match sabotage was a tactic used to enact revenge on harassers. Paula, a participant in a user feedback pilot focused on anti-toxicity told us that was her favorite response to encountering both racist and sexist harassment in gaming. She couldn't "just quit" she said, because quitting before the end of the match triggered consequences from the game, which could include reduced standing in match making and temporary bans, but that she would often just "stand back and let them die."

Courtney, it should be noted, is White. Gamers of color like Paula and also including several Intel colleagues who are African American and gamers themselves told me that they, like Princess, tended to stay off of voice chat, and to choose carefully who they played with. As with the women of color that Kishonna Gray studies, some of the them self-segregate (Gray 2012). For these women and African American players, qualities of voice were clear "tells" where speaking online over voice chat revealed them as female or non-white or both. Thus participation in voice chat was particularly fraught. Yet for those players who "passed" as Princess did, in saying she was a "squeaker," or Sam, a competitive e-sports player who is Asian, yet whose voice does not necessarily reveal his ethnic background, the default assumption of straight white male-ness, and the rampant use of ethnic and sexist language

could become deeply uncomfortable and cumulative over time, leaving players deeply ambivalent about those experiences. Sam, for example both downplayed the issue, saying "it's just part of the game" and that "people get really intense" but later said that people's anti-Black and anti-Asian comments would stay with him and really impacted his self-esteem and sense of self.

From these stories, it becomes clear that sexism and racism are at the center of US players' experiences of toxic behavior in gaming, but that at the same time, these experiences are tempered by perceptions of gaming overall, where some poor behavior is to be expected, and through tactics of avoidance (not using voice chat, not playing with strangers), revenge (letting them die), or, in the worst scenarios, quitting the game and finding a new game to play.

INDIA STORY

Toxicity in gaming in India plays out quite differently than it does in the US or in China. One major difference in India relates to the gaming infrastructure, which is the most diverse of the three countries. While many of the communications platforms used in India are the same as for the US, players game on servers that are located in different regions of their own country, where there are significant ethnic and linguistic differences, and they also play on servers in neighboring countries like Pakistan and Singapore, countries with whom India has complex historical relations. This diversity in infrastructure gives rise to opportunities for toxicity that are not as prevalent in the US and China. The biases that were articulated related to cultural sensibilities of hierarchy in the Indian social system, and regional identities and linguistic markers. India also had the most hostile environment for female players.

One of our participants was a young woman in her 20s who got into gaming after meeting her boyfriend, who organizes local gaming meetups and competitions. For Divya, while she has had a number of bad experiences while gaming, those experiences are of a continuum with the overall sense of powerlessness and frustration linked to her position as a young woman more broadly, both online and off.   In one story she told us, a player she didn't know cursed at her while playing PUB-G. Later, he joined a community linked to a different game and posted negative comments about her and her game play. He then left the community but seemed to return and then leave again, perhaps, Divya thought, to check on the effect of his postings. When she looked the other player up, she discovered he was a high ranking player in the game. She philosophized that perhaps, then, it wasn't so bad: a high ranking player had noticed her, and taken the time to seek out her community to comment on her play, and then returned later to see if she had replied. She brushes it off: "it comes to mind sometimes, but it doesn't affect me that much." In another story she told us, another player talked a lot about how bad she was. In that instance as well, she said, it was better to stay quiet because she didn't know that game as well as the other player, and also because she doesn't like to say much when she is angry. She told a similar story about her college life where, she said, she doesn't care much for her faculty, but when someone says something mean to her, she tries to distance herself. "I would like to say some things to them, but I can't because they are faculty." Instead she writes her anger down – about the faculty, about things that happen in the gaming community, and about how she can't go out at night for fear of her safety. For Divya, all of these frustrations are of a piece, shaped by her experience of being powerless in relation to other people or the overall context.

Figure 2. Competitive gaming tournament in India. As elsewhere in the world, competitive gaming is male dominated.

For Sushanth, an aspiring e-sports player in his 20s, male, the tensions are national and regional but not universal. As he said,

> *that actually depends on your luck. Like if you're really unlucky, you might encounter them every game you play... It's just random, there can be one game where one Singaporean hates Indians, whereas in the next game, they can be like, one more Singaporean who doesn't care at all and just wants to like play the game.*

In addition, these international encounters can have a positive resonance. For example when he told us about playing with Pakistani players. "I mean, you usually see Pakistan or India like big enemies, right, but when you actually play a game, those games actually turn out to be really good players who respect others and stuff."

For Sushanth, then, playing on regional servers with players from nearby countries is a matter of luck. In some cases he encounters negative perceptions of Indian players with repercussions from verbal abuse to throwing a match, whereas in others, international servers humanize players from "enemy" nations, people who just want to play the game and respect others.

As in the US, skill in the game makes a difference. Sushanth's positive perception of some Pakistani players is shaped by his experience of them as "really good."

CHINA STORY

The infrastructure story in China is colored by the fact that many game players make a concerted effort to get to servers outside of the country, accomplished by using VPN. Some players prefer playing on foreign servers. One young man, Ren, has met many foreigners this

way, and even went so far as to fly to Germany to meet one of his online friends. The theme of traveling to other places in games was prevalent among other players, as well, but for most it was virtual.

Another key difference in China, and one that might present challenges in the design of a scalable chat-based algorithm for identifying toxic behavior is the fact that they use Chinese proprietary applications, like YY, QQ, and WeChat for gaming communications instead of Discord and Twitch apps, which are more widely used in the rest of the world. They also play a number of games that are modified from the original games that are available elsewhere. For example, they play Game for Peace, a version of PUBG that was developed specifically to appease Chinese censors, and is played on Android mobile devices.

In China "mean people" or "aggressive people" were cited as a serious problem, but people were reticent to talk in detail about their personal experiences with negative in-game interactions. Because participants were somewhat evasive, we feel that we don't have an adequate understanding beyond knowing that online toxicity *is* a problem. We know that cursing and "rage" behavior are prevalent, and we also know that gender-based interactions can be strained.

Most men talk about toxic behavior in a couple of ways—the use of foul language, which did not seem to bother them much, and bad actors, people who did not pull their weight, or who disrupted play in some other way. One young streamer, Jun, told a story about being blocked when the system detected that he was inadvertently teamed up with a known hacker.

> *The system detected that I was playing with a hacker. I did not know that I was playing with a hacker. I never cheated nor played with hackers. My friends are all working on that day and I was randomly matched to a group. Eventually, the webcast was reported and my account got blocked.*

He viewed the hacker as a bad actor, but wasn't bothered by other behaviors he encountered. Mostly, toxic behavior seemed like a non-issue for the Chinese men in our study, even though players reported it as a nuisance in their diaries.

Many of the women gamers that we spoke with initially told us that they did not experience online discrimination or sexual harassment, but some of the stories we heard suggested otherwise. One young woman, Juniper said that because men viewed women as weak players, males always sought to help and to protect females. This was especially true when she was younger and just started out. Later, when she became a strong player, this gave her extra advantages in the competition. She and others reported that sometimes men even pretended to be women in games just to get help from other male players. In describing an early experience in the gaming world, she told us how her brother served as her protector and guide.

Figure 3. Interviewing Juniper with her boyfriend in the background. How she talked about herself and her gaming abilities changed with him in the room.

> *People[males] there [in the game] all regarded me as a little girl, so they took care of me. No matter if it was a certain task or something else, they would do favors for me. I felt happy.*

Juniper went on to become a successful and relatively high-ranking player in male dominated games, like World of Warcraft and League of Legends. While we were interviewing her, her boyfriend with whom she lived, came into the room. He had overheard her talking to us about her prowess and ranking. He scoffed and said, "You are not that good." He went on to tell us that men are innately better at playing computer games; he claimed they are genetically wired with faster reflexes. Juniper immediately acquiesced and was quick to say that she was not as good as he was. This in-person dynamic played out repeatedly in other stories we heard from women; they consistently said they were not very good.

Overall, most Chinese women players agreed that men were quite polite and solicitous of them in games. Most said that the condescension did not bother them; they accepted that this is the way of the world. Related to this is the role that game playing has in match making. Both men and women seek opportunities to meet and interact with potential mates in the games they play, and several told stories of how they met their partners through playing games. Sometimes these relationships went wrong, and other times unwanted advances by male players were problematic, but nobody wanted to share the details of these more contentious online experiences.

Most women players employ strategies for avoiding conflict of any sort with players of any gender. Rose, a young woman who enjoys a range of games from League of Legends to adventure games, said aggressive behavior by both women and men is common, usually in the form of cursing. She simply chooses to ignore aggressive behavior because she just wants to have fun.

Girls are also quite aggressive now. No matter in life or in game, women also become very aggressive....We couldn't let those bad people affect us. I wouldn't even think about having quarrel with them. I met people having quarrel in the game. They bought the in-game trumpet to curse on all channels. I don't even know what the meaning is behind that. They're so angry and still spend that much money. 100 [yuen] for 10 trumpets, just to say 10 words. I don't know how much they would spend just to have a quarrel with someone.

Mei, a young mother who enjoys playing games, but doesn't want to engage with hostile online game environments, limits her play to "people she knows," some from real life and others she has come to trust through playing on teams with them.

Because in online games like this, too many strangers are too complicated, that is to say. Sometimes they play games to look for having sex.

When pressed, Mei admitted to having had some negative experiences when she was younger, but she did not want to go into details. Mei tries, for the most part, to only play games with friends and friends of friends to avoid toxic online situations in games.

For example, my friend plays a game and invites me to join a chat group. My friend has some friends in real life and some net friends in the chat group. When we become familiar, we can find that some people are in the same city, so we can make an offline meeting appointment. For example, we can have coffee in a bar or play games while gathering.

It became clear that the general sentiment in China is that women should not be interested in playing competitive online games. One of the phrases that we became familiar with while doing our research in China was, "I am not a normal girl." Almost every woman we spoke to said something like this about herself. The implication was that "normal girls" should be interested in fashion, shopping, and K-pop, not in playing video games. Furthermore, most of the women we spoke with, even if they were quite good players in actuality, downplayed their abilities, especially if men were present during our interviews.

COMMON GROUND

In thinking across these three cultures, we were able to find some common ground when it comes to toxicity, most of which have implications when it comes to thinking about scalable technical solutions.

Sexism is an issue in all three countries. That said, how it is expressed and how it is experienced are different from place to place. For example, in the US and India sexism is frequently expressed in verbal insults which could possibly be detected by our chat algorithm, but in China it is expressed in less easily identifiable actions— offers of help and in condescension. The generalized problem at sexism at a macro scale is difficult to address through a single technical solution because of the nuances that are revealed at the micro scale.

In all three countries, the relationship between skill differences and toxicity is present. Potentially toxic interactions arise when there is a skill mismatch among team members or competitors. If someone on your team is not skilled enough, or is refusing to fulfill their role in the game, it can destroy one's ranking. One's ranking in a game is not trivial, often the

product of significant investments of time and sometimes money as well. The desire to win at any cost gives rise to heightened emotions that lead to toxic behavior, often in the form of verbal abuse in voice chat, potentially identifiable with a speech recognition algorithm. However, there are caveats.

For example, there are exceptions to the skill rule that are also generalizable. Toxic behavior is more likely to emerge and be experienced as toxic with encounters between people of different skill levels who have relatively loose social ties. Thus, if one tends to play with people who are familiar, the emergence of highly toxic moments is less likely to occur, and also the toxic moments that do emerge are less likely to be perceived as toxic. It is the difference between your best friend calling you a [imagine an insult] when you make a winning move against her, and having perfect stranger calling you the same thing. Having a stranger hurl an insult at you is usually perceived as a worse transgression than having someone known to you do it. This is getting into some shaky areas for our algorithm. Can it learn who is known to you?

A correlate to the relationship of disparity in skill levels giving rise to toxic behaviors is that toxic behaviors are tolerated if the perpetrator's skill or relative social power in the team is high; people will tolerate bad behavior of the person who is doing the most to ensure a win. Again, we are unsure what this might mean for how a speech recognition algorithm might judge the toxicity of an incidence. Should it ignore potentially toxic speech acts of the highest ranked player?

Across all three geographies where we did research we observed that nearly all toxic interactions were described as a combination of speech, text, and other actions that crossed over multiple social platforms outside of the game itself. While an in-game or in-voice-chat algorithm might catch certain types of toxic behavior, this general finding reveals that the problem will not be entirely solved by such a technical solution.

Finally, universally, bad actors in games—spoilers and hackers—represent a particular type of toxic gaming behavior, another behavior for which there are no perfect technical solutions, and one that was beyond the purview of our work. Some games already have systems for automatically ferreting out such bad actors, but as we saw in the case of Ren in China, they can be quite heavy-handed, penalizing players who accidentally drafted a hacker into their team.

THE PITFALLS OF OVER/UNDER GENERALIZING

Anthropology, as a discipline, has wrestled with the issue of generalization vs specificity since its beginnings as evidenced from theoretical and ethical positions that have swung widely from sweeping cross-cultural generalizations evident in early notions of evolution of cultures (Frazer 1890; Morgan 1877) and in direct comparisons between cultures (Mead 1928; Benedict 1934). Franz Boas, the "father" of American Anthropology favored the specific, although he argued for a historical comparative method (1940). Cultural relativism and the idea of ethnography as "thick description" have largely prevailed in the second half of the 20th century. However, in recent years, some anthropologists have argued that the lack of comparison has weakened the discipline (Borofsky 2019; Nader 2015).

Within the EPIC community, we have seen arguments arise regarding our ability to make meaningful cross-cultural comparisons, and about under or over generalizing. One thing is clear—it is paramount for businesses and institutions operating in a global market

place to design products and services that address both the general and the specific whether we are creating software and hardware solutions for automated driving (Rothmüller et al. 2018; Vinkuyzen and Cefkin 2016), trying to design technologies for domestic spaces (Pulman-Jones 2005; McClard and Dugan 2017), or trying to help our companies understand their customers (Anderson et al. 2017).

Our ethnographic research with gamers shows clear generalizable patterns across the three countries where we did research in the impact of abusive and harassing behaviors, including pervasive domains (e.g., gender, race, ethnicity), common behaviors (e.g., textual and verbal abuse, disruptive or hostile play tactics), and patterns of escalation where the most harmful behaviors transcend particular gaming episodes and platforms. The most troublesome behaviors are those that escalate to events on other social media platforms, expanding from individual hostilities to organized harassment, and from digital exchanges to actions in or affecting the physical safety, including doxing (publicizing personally identifiable information such as names and addresses) and swatting (sending emergency and other services to a physical address) in order to intimidate and harass.

By identifying these broad patterns, ethnography is instrumental in teasing out directions and trajectories for the movement from crawl to walk to run in the technical solution space. In other words, these patterns suggest not only relevant categories to identify, but suggest next steps for expanding or connecting to other kinds of solutions that, for example, trace behaviors across verbal exchanges and in-game actions, or from one game platform to another and to social media: solutions not available today, but important in defining how any given solution might fit into a larger picture.

At the same time, ethnography reveals levels of specificity that can feel overwhelming to technical and business teams, and clearly show the need for extensive localization that includes game specific and region specific language practices such but also goes beyond language to encompass region specific social tensions and common expressions and symbols related to those divisions. In addition, our interviews revealed distinctions across game genres and gameplay structures that impact the ways that harassment unfolds. For example, competitive team and match based games where players are often paired with strangers, stakes and emotions run high, and tempers flare. As related above, negative experiences tend to be most acute where social ties are weak. Role playing games, by way of contrast, tend to involve more sustained engagement over time, where players create and join guilds, playing together for weeks, months, even years. In those games, toxicity and harassment stories focus less on sudden and isolated outbursts, but can involve longer term targeting and collective actions to isolate, bully, and intimidate victims.

Finally, the relationship between local and global and the frequent mixing of known and unknown participants, strong and weak social ties raise important issues of power, ethics, and responsibility. Specifically, who decides what is, and is not "toxic?" Today, this is left to individual gaming platforms to decide, and frequently up to players themselves to report. However, when considering the idea of technology that might screen interactions for offensive behavior, many players worry about the distinction between "trash-talk" which can be fun between friends, and harassment or trolling which is distinctly not fun for the victim, though perpetrators may perceive and represent it as "play." At the same time, while participants themselves may be comfortable with certain kinds of biased (sexist, nationalist, racist) expression amongst themselves, does that mean it should be allowable? If sexism

expressed in games takes place in contexts where men and women alike share a perception of male and female roles and abilities as unequal – is it toxic?

OUTCOMES AND CONCLUSIONS

Sometimes the specificity of ethnographic research is at odds with telling a simple generalizable story, and our insistence on telling complex ethnographic stories may make it seem immediately more difficult to act from a developer and business perspective. However, the complexity of ethnographic research findings is what ultimately gives direction for evolution of a product by emphasizing the local experiences and social and cultural differences that exist in disparate markets. In other words, ethnographic research has the power help to create a roadmap for scaling a solution. So while an MVP may barely scratch the surface in the "Crawl" stage, later implementation will have the guidance embodied in the general and specific findings gleaned from research.

How has the complexity gleaned from our ethnographic research on gaming informed scaling of the anti-toxicity voice chat algorithm? We showed the development team that localization, not just language translation, matter. We showed them that regional tensions, social hierarchies, vocabulary, and gaming community cultures matter too. For the "minimum viable product" (MVP) the team is working on an algorithm that accurately interprets a range of vocal intonations and accents in American English, and they now have roadmap that includes local sensitivity.

To wrap up, while, participant-observation is the cornerstone of ethnographic research, it is not the only hammer in our toolkit, and using multiple methods across a research domain enables researchers to make impactful contributions that reveal and bridge between micro and macro perspectives. Participant-observation helps drive the development of scalable technical solutions by revealing underlying patterns and common ground, by teasing out the specificity of local interpretations and enactments. The ethnographic research we discussed here has formed a critical part of developing an effective strategic roadmap for meeting users where they live and play.

Anne Page McClard holds a doctorate in cultural anthropology, and has worked in the technology industry for more than 30 years. Anne uses ethnographic research to influence and drive product design and strategy, in both consumer and B2B markets. Throughout her career, she has sustained an interest in gender issues in academia and technical industries.

Jamie Sherman is a cultural anthropologist (Ph.D. Princeton. 2011) and research scientist at Intel Corporation. Her research background is in techniques and technologies of self-transformation, performance, and dynamics of race, gender, and play. Since joining Intel in 2012, her work has focused on emergent technological practices from quantified self, to live motion capture projection. Her current research develops usages and drives strategies for video game play, media creation, and online toxicity.

REFERENCES CITED

Adinolf, S. and S. Turkay
2018   "Toxic Behaviors in Esports Games: Player Perceptions and Coping Strategies." In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY '18 Extended Abstracts)*, pp. 365–372. DOI:https://doi.org/10.1145/3270316.3271545

Anderson et al.
2017   "Creating a Creators' Market: How Ethnography Gave Intel a New Perspective On Digital Content Creators," *Ethnographic Praxis in Industry Conference Proceedings,* pp. 425-443 .

Benedict, R.
1934   *Patterns of culture*. Boston: Houghton Mifflin Company.

Boas, Franz.
1940.   The Limitations of the Comparative Method of Anthropology. In Race Language, and Culture. Pp. 270-280. New York: Macmillan.

Boellstorff, T.
2008   *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*. Princeton: Princeton University Press.

Borofsky, R.
2019   Where Have the Comparisons Gone? *Society for Cultural Anthropology*.

Dewey, C.
2014   "The only guide to Gamergate you will ever need to read" *Washington Post*. https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read/

Dyer-Witherford, N. and G. de Peuter
2009   *Games of Empire: Global Capitalism and Video Games*. Minneapolis: University of Minnesota Press

Frazer, J. G.
1890   *The Golden bough, a study in comparative religion*. London: Macmillan.

Fredman, L.A.
2018   *Not Just a Game: Sexual Toxicity in Online Gaming Hurts Women*. Unpublished Dissertation, University of Texas at Austin. http://dx.doi.org/10.26153/tsw/2112

Gray, K. L.
2012   "Intersecting Oppressions and Online Communities," in *Information, Communication & Society*, 15:3, 411-428, DOI: 10.1080/1369118X.2011.642401

Gray, K. L., B. Buyukozturk, and Z. G. Hill
2017   "Blurring the boundaries: Using Gamergate to examine 'real' and symbolic violence against women in contemporary gaming culture" in *Sociology Compass*. 11:12458. https://doi.org/10.1111/soc4.12458

McClard, A. and T. Dugan
2017   "It's Not Childs' Play: Changing Corporate Narratives Through Ethnography," *Ethnographic Praxis in Industry Conference Proceedings.*

Mead, M.
1928   *Coming of age in Samoa.* New York: William Morrow & Company

Nader, L.
2015   *What the Rest Think of the West : Since 600 AD.* Oakland: University of California Press.

Nardi, B. A.
2010   *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft.* Ann Arbor: University of Michigan Press.

Pulman-Jones, S.
2005   "Using Photographic Data to Build a Large-scale Global Comparative Visual Ethnography of Domestic Spaces: Can a Limited Data Set Capture the Complexities of 'Sociality'?" EPIC *Ethnographic Praxis in Industry Conference Proceeding*, pp. 128-139.

Rothmüller, M. et al.
2018   "Designing for Interactions with Automated Vehicles: Ethnography at the Boundary of Quantitative-Data-Driven Disciplines" *Ethnographic Praxis in Industry Conference Proceedings*, pp. 482-517.

Vinkuyzen, E. and M. Cefkin
2016   "Developing Socially Acceptable Autonomous Vehicles," *Ethnographic Praxis in Industry Conference Proceedings*, p. 522–534.