# The Best of Both (Virtual) Worlds: Using Ethnography and Computational Tools to Study Online Behavior

NICOLAS DUCHENEAUT
NICHOLAS YEE
VICTORIA BELLOTTI
*Palo Alto Research Center*

*In recent years, many ethnographers have conducted participant observation studies in virtual worlds, whether in games like World of Warcraft or user-generated environments like Second Life. However, the acceptance of digital fieldwork as a legitimate form of ethnography does not make it strictly identical to its physical counterpart. In particular, the logistics of virtual ethnography offer both opportunities and pitfalls that practitioners must address. The virtual nature of the space also compounds traditional issues such as generalizability and coverage. In this paper, we will highlight several interesting opportunities and challenges in conducting ethnography in virtual worlds. Moreover, we will then argue that the common problems shared by quantitative and qualitative social scientists in virtual world research serve to bridge the methodological divide, such that virtual ethnography could be greatly enhanced with the use of computational tools usually more associated with quantitative research methods. We will rely on examples from our own research projects to illustrate the value of such a combination and describe how it could be concretely implemented in a "digital ethnography toolkit."*

## THE MERITS AND PITFALLS OF VIRTUAL ETHNOGRAPHY

Through a combination of widely accessible broadband Internet access, fast personal computing technology (in particular, ever faster graphical processing units or GPUs) and the development of increasingly complex software, millions of users worldwide are now participating in persistent, avatar-mediated online environments known as 3D virtual worlds. While many of these online spaces descend from games (e.g. Everquest, World of Warcraft), others offer more open-ended environments (e.g. Second Life). The potential social, cultural, and economic impacts of these spaces have attracted enormous attention in the past decade, but evaluating these impacts remains a core challenge.

Because these environments are persistent and open-ended, they are the site for the accumulation of their users' effort into currencies and items, which quite famously have led to real exchange economies (Castronova, 2003) as well as social relationships (Williams et al., 2006) and new competencies (Malaby, 2006). Moreover, users are beginning to form groups within and beyond these spaces (such as guilds in World of Warcraft), and these new social institutions are developing new kinds of governance and becoming an increasingly important part of their members' social lives (Ducheneaut et al., 2007). As such, the broad scope of human social experience captured by these online spaces makes them possibly the first examples of full-fledged electronic societies, reproducing social and cultural issues not far removed from those of the physical world (Castronova, 2005). Such rapidly changing circumstances call for exploratory, qualitative research, that dimension of social science inquiry best suited for examining social change on the ground. In other words virtual worlds

*EPIC 2010 Proceedings, pp.136-148, ISBN 0-9826767-3-8.* 

need ethnographers, much like the earlier unexplored societies that led to the founding of the methodology (Malinowski, 1922).

Ethnographic field research involves the study of groups and people as they go about their everyday lives. The term "participant observation" is often used to characterize this approach, since researchers seek to immerse themselves in others' worlds in order to grasp what they experience as meaningful and important. Ethnography therefore entails "some amount of genuinely social interaction in the field with the subjects of the study, some direct observation of relevant events […] and open-endedness in the direction the study takes" (McCall & Simmons, 1969). However, virtual worlds pose a methodological challenge to the ethnographer. Indeed most of this approach is based on the ethnographer "being there" in the field to observe – but this "there" is nebulous at best in the case of online spaces (Rutter & Smith, 2002). This issue of presence (or lack thereof) was the subject of intense debate in the late 1990s as ethnographers began to consider moving into digital spaces. While this debate was raging, the use of "virtual ethnography" remained controversial, and it took several years before it came to be accepted as a legitimate practice. A good summary of the debate can be found in Lyman & Wakeford (1999).

Today a majority of researchers have come to believe in the virtues of "virtual ethnography" (Hine, 2000), that is, an adaptation of traditional ethnography to the study of cyberspace. As Mason (1999) put it:

> *"A virtual ethnography is one that fully immerses the ethnographer into the consensual reality experienced by groups of people who use computer-mediated communication as their primary, and often only, means of communication. As such, the online or virtual personas of the participants are the main focus of the ethnographer. Generally, researchers have wanted to focus on the person at the keyboard; a virtual ethnography reverses this and works instead with the persona that has been projected into cyberspace by the typist."*

A virtual ethnography is then, simply, an ethnography that treats cyberspace as the ethnographic reality. However, the newly established legitimacy of virtual ethnography does not imply a straightforward transfer of ethnographic practices to the digital realm. Indeed digital environments, with their inherent affordances and constraints, reconfigure demands on qualitative researchers in significant ways. Some of the core logistical challenges include:

- *Coverage*: Virtual worlds abolish many of the constraints imposed on traveling in the physical world since avatars can often "teleport" (or at least move faster than they would if physical constraints were enforced) from one location to the next. They also allow people separated by large physical distances to share the same virtual space. The combination of these two factors makes observing social activities in a group like a World of Warcraft guild quite difficult: at any given time the group's members might be scattered across the virtual world and moving abruptly across large distances, and a large fraction of the group may also be absent (for instance, European players might be asleep while US players are active). Tracking all the members of a given group therefore becomes close to impossible, since they are not bound to

a common physical gathering point at specific times, as they would be in more traditional ethnographic settings (e.g. a tribe, an organization).
On top of this, social exchanges can take place across a variety of channels, some public, some private: members can "talk" to each other using public chat, semi-private guild chat, or private messages, among others. Directly observing the totality of a group's activity becomes therefore quite difficult: researchers using traditional participant observation techniques would be able to witness only a small fraction of a virtual world group's activities, constrained to public interactions taking place while the observer is present in a circumscribed virtual location.

- *Generalizability*: due to their broad appeal and accessibility, virtual worlds are now used by people with widely diverging socio-demographic backgrounds and aspirations (see for instance Williams et al., 2006). As such it becomes difficult, not to say impossible, to conduct a traditional ethnographic study of an entire virtual society: after a time-consuming participant observation in one guild, for instance (which would already suffer from the coverage issues mentioned above), researchers could at best claim to understand only the experience of a fairly circumscribed type of group (say, guilds composed of young adult males focused on large-scale warfare) while many other social units can exist (for instance, a guild composed of family members mostly interested in low-key world exploration conducive to chatting and sociability).
- *Data collection and processing*: accumulating data in virtual worlds can be deceivingly easy. Quite often textual interactions can be saved without limits to a researcher's hard disk, and capturing the entirety of a researcher's visual experience is feasible thanks to common video capture software. While this can sometimes prove useful (for instance, video logs are necessary for fine-grained conversation analysis – see Moore et al., 2007), the resulting data is more often a deluge of digital content that is hard to process using traditional fieldnote coding techniques (Emerson et al., 1995). Conversely, some new forms of data that ethnographers may not be familiar with are available: in game-like worlds there is information about combat performance ("kills", damage done, etc.), for instance (Ducheneaut et al., 2006). More generally a lot of a virtual world user's actions are automatically tracked (e.g. "emotes" such as the number of hugs given, total distance travelled, etc.) but retrieving them might require familiarity with specific software tools.

In spite of the difficulty of processing virtual world data, mentioned above, the abundance of this data has proven a boon to more quantitatively inclined researchers. Indeed, the relative ease of automated data collection and the sheer scale of the resulting data sets have enabled research projects that would have been otherwise infeasible. For instance, it has become possible to reconstruct the social interactions between hundreds of thousands of players from massively multiplayer online games (Ducheneaut et al., 2007). The results of these quantitative analyses, however, could easily be misinterpreted if they were not grounded in qualitative interpretations of the experienced reality of virtual world participants, which is precisely what ethnography is very good at providing. What is needed therefore is a way to combine both approaches and leverage the strengths of quantitative research (in particular, its broad array of tools and techniques for collecting and summarizing large-scale data sets) with the fine-grained, culturally sensitive perspective offered by ethnography.

This move, to find ways to incorporate technology into the best practices of ethnography, is consistent with one of the core insights of Science and Technology Studies – the mutual influence of tools and people in the course of technological practice. As Andrew Pickering (1995) and others (Latour & Woolgar, 1986) have suggested, the human relationship to technology is one of mutual influence yielding possibilities for innovation only through their contingent combination. Some of this paper's authors have recently explored the possibility of adopting a "technographic" approach by building information visualization tools aimed specifically at assisting ethnographic fieldwork online (Ducheneaut, 2005). The online communities that were the focus of this work (email-supported open source software development), however, were quite simple compared to full-fledged virtual worlds. In this paper, we therefore draw on our practical experience, combining quantitative tools and techniques with ethnography to argue that the study of virtual worlds would greatly benefit from a similar approach.

Rather than a more abstract discussion of the possible intellectual merits of this methodology, which we have already offered in previous work (see again Ducheneaut, 2006 and Ducheneaut, 2003 for an even deeper treatment of the subject), we focus instead on concrete examples from our ongoing study of social dynamics in World of Warcraft, a large game-based virtual world with more than 11 million paying subscribers. By presenting our solutions to the three major logistical issues outlined above, we hope to start a dialog with other practitioners about how it might be possible to experience the best of two worlds (quantitative and qualitative research) that are usually kept apart but have now been given an unprecedented opportunity to join forces in the study of full-fledged virtual societies. In particular, we show how such a hybrid, qualitative-quantitative approach can be put to task in the form of a "digital ethnography toolkit," that is, analytical software tools designed specifically to support digital fieldwork. We use the example of one such tool we developed, the Social Dashboard, to illustrate how this approach can provide concrete solutions to important business problems such as customer retention and targeted marketing.

## FIRST OBSTACLE: GENERALIZABILITY

We define generalizability as the degree to which a given set of findings reflect characteristics of the population sampled as a whole. Generalizability has always been a challenge in applying ethnographic methods due to the high cost and localized scope of participant observations. It is often difficult to assess or provide a sense of how generalizable ethnographic findings are outside of the observed sample (LeCompte & Goetz, 1982). But virtual worlds provide the opportunity to assess the generalizability of ethnographic findings because many social metrics are tracked automatically by these environments. For example, after conducting an ethnography of both a 10-person guild and a 25-person guild on one server in the online game World of Warcraft, it is possible to reference the companion site Armory (http://wowarmory.com) to calculate the average guild size on that server, as well as the average guild sizes across servers. Knowing how typical or atypical the sizes of the observed guilds are allows ethnographers to better estimate how their findings may or may not generalize to other guilds.

The problem here is that this data is most often not available in a synthetic format highlighting the most relevant variables for a given ethnographic project. In the context of our study of the social

dynamics of groups in virtual worlds we developed a suite of tools, the "Social Dashboard" (Ducheneaut et al., 2007), to summarize these high-level social metrics in order to guide ethnographers toward the best and most generalizable participant-observation sites. We used this tool in a recent study of guilds in World of Warcraft (Williams et al., 2006), where we first mapped out the range of guild sizes in the game and selected representative guilds that fell along the spectrum. We also selected particular individuals in those guilds to cover a range of social network metric variances: people who were well-connected, people who recently joined, etc. This information was then conveyed to our fieldworkers who conducted in-depth interviews with those individuals. Thanks to this approach, a small team of qualitative researchers (we deployed only two ethnographers in the game world) was able to collect highly generalizable data since we were confident they had interacted with the full-range of group types present in this virtual world.

## SECOND OBSTACLE: COVERAGE

We define coverage as the ability to be in many places at the same time, as well as the ability to quickly spot and observe low occurrence events wherever and whenever they occur. Virtual worlds also provide unique advantages in studying these low-occurrence or highly unpredictable events, such as an organization splintering or loss of a central member in a large group. This is because it is possible to instrument virtual worlds to monitor active groups and alert researchers to these events in real-time. In other words, computational tools can be used to track and flag the "key events" (Fetterman, 1998) that are essential in ethnographic practice. This way, ethnographers can track several sites in parallel, greatly reducing the time between critical incidents and therefore generating more useful observations in less time.

The same is also true for being able to monitor interesting events at the individual level that are impossible to schedule ahead of time. For example, after enlisting a cohort of individual players into a study, it is possible to monitor their characters in real-time for specific triggers of interest, such as: when they leave or join a guild, when they enter a large 25-person raid for the first time, or when they become a central member of a guild. Thus, instead of following individual players around and waiting for these events of interest to occur, a researcher could use a real-time monitoring tool and be alerted when their triggers of interest occur and they could, quite literally, beam themselves right next to their participants as the scenario unfolds or initiate an interview.

The "Social Dashboard" described above was also developed to address this need. Researchers first define the "key event" of interest – in our case, the death of a guild due to progressive loss of its entire membership. Based on the data accumulated to this point, the software then attempts to compute the best possible model of which available variables contribute to the event. In our case for instance, the size of a guild, its fragmentation (the number of cohesive subgroups), and its density (how interconnected the members are) all seemed to play a role in the guild's eventual survival. The software then highlights these key variables (Fig. 1) and lets researchers find groups that are nearing the key event of interest (here, a guild's disappearance – see Fig. 2). This way, researchers can quickly spot groups of interest across the entire virtual world and log into the world *before* the event occurs, giving them the opportunity to observe it and interact with the users to understand how the situation came about (something that the numbers alone cannot reveal).

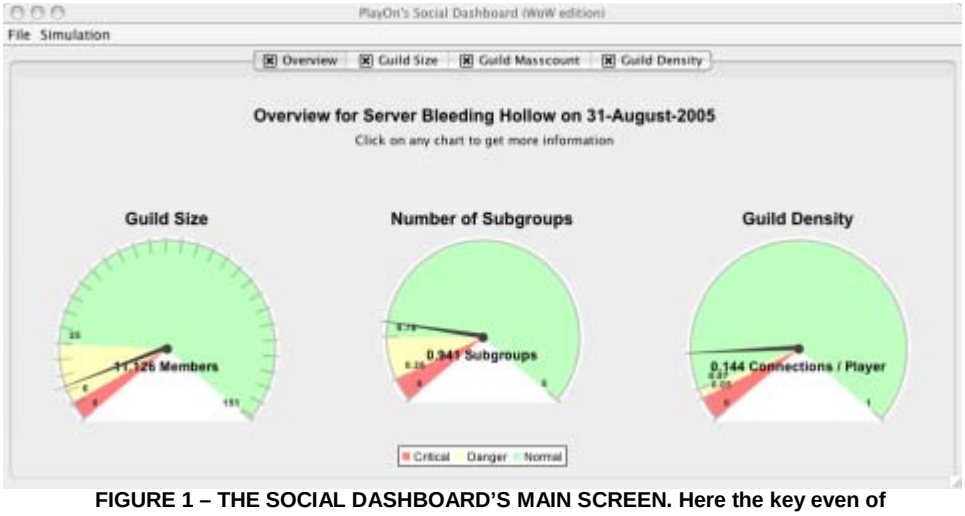


**FIGURE 1 – THE SOCIAL DASHBOARD'S MAIN SCREEN. Here the key even of interest is guild death, and the software is tracking three high level variables that appear to contribute to the event.**

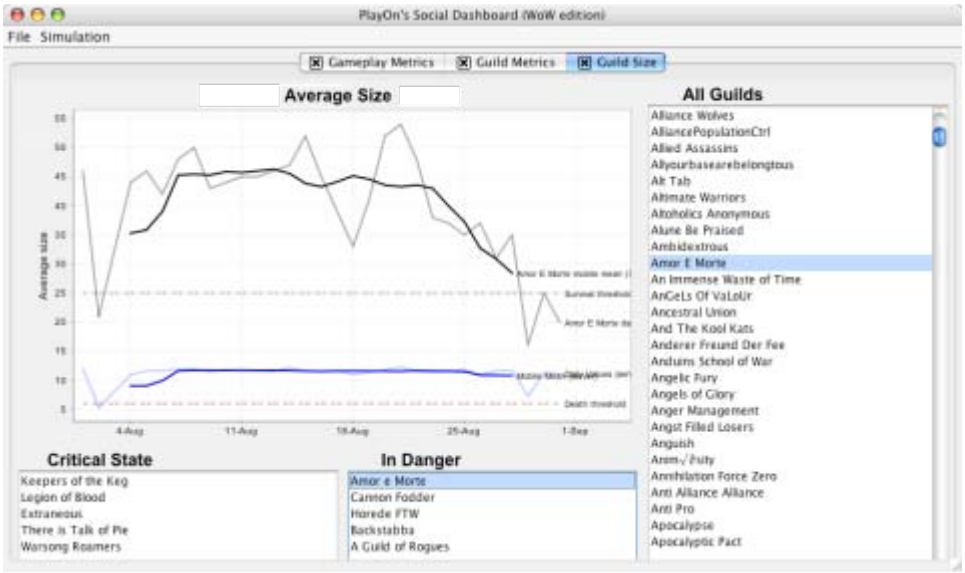


**FIGURE 2 – THE TRACKING SCREEN ALLOWS RESEARCHERS TO SPOT A KEY EVENT BEFORE IT OCCURS. Here the guild "Amor E Morte" is collapsing (its size has gone below the model's survival threshold) but has not totally disbanded yet. Researchers can immediately log into the virtual world to interact with the guild members and understand the root causes of the group's problems.**

## THIRD OBSTACLE: THE DATA DELUGE

We have illustrated above how the development of custom software, the "Social Dashboard", helped us leverage the strengths of quantitative research (quickly processing large data sets, visually representing norms and trends, modeling events) to guide the work of ethnographers such that their analyses do not suffer from the traditional problems of coverage and generalizability. This software, however, was tailored to the analysis of group dynamics since it was the main research issue we were investigating. As such, it only processed a small fraction of the deluge of data potentially available to guide ethnographers. Still, we believe some key components of our approach could be re-used in a much wider range of research projects, giving rise to the possibility of a "virtual ethnography toolkit" to support online fieldwork.

The core of our approach relies on having as many virtual "boots on the ground" to compensate for the inability of an ethnographer to be in several places at once. In virtual worlds, this means creating "robots", that is, avatars controlled by a machine for the sole purpose of recording activities in the world. Each robot then needs to 1) cycle through virtual observation sites of interest such that data is collected frequently enough to spot key events and 2) record data in a format suitable for visualization and the guidance of ethnographers.

While each game/virtual world is based on different avatar systems, the high-level logic of creating as many robots as needed and cycling them through observation sites is independent of the environment being studied. As such, the first component of a virtual ethnography toolkit should be a kind of "control tower" supervising the robots' activities. We have implemented (and currently use for an on-going project) such a piece of software in the Java programming language. Our software can launch an arbitrarily large collection of virtual machines (that is, independent copies of Windows and whatever virtual world is being studied) and orchestrate their activities to optimize data collection (for instance, given 10 World of Warcraft accounts, the software would spawn 10 virtual machines and assign 28 game servers to observe to each of them, thereby covering the entire US gaming population since Blizzard Entertainment currently runs 280 servers there). The software is designed such that each virtual machine could contain any other virtual world (say, Second Life or Everquest, for instance): the only additional software engineering required would be the in-world scripts tasked with recording the data.

Moreover, the "control tower" was also designed to extract and process the data collected in-world whenever a robot is cycling out of a given server. Provided the in-world scripts mentioned above conform to our software's application programming interface (API – in other words, a set of conventions about how the data should be formatted), data can be processed whatever the source might be. The combination of environment-independent control logic and common data formats can greatly speed up deployment across a wide variety of current and future virtual worlds.

As we illustrated earlier with the Social Dashboard, a virtual ethnography toolkit should eventually allow ethnographers to visualize high-level events in order to guide their activities. Visualizing the raw data, however, is often pointless: our WoW robots, for instance, simply record who is online at a given

time. This data becomes interesting only when pre-processed with the following heuristic: if two players from the same guild are online at the same time and in the same game zone, then they are playing together. This translates into a social network graph connecting players who usually play together, which in turn can be processed to reveal interesting properties of the group (e.g. how centralized it is, or how fragmented, etc.). We use this detailed example to illustrate the importance of a second component of the ethnography toolkit, preceding visualization: an intermediate layer mapping raw data to higher-level concepts of interest. This could be implemented using simple user interface techniques, such that no further software engineering is required. For instance, users could see a list of raw variables and be allowed to "pile" them into higher-level aggregates using simple drag-and-drop. Or users could create simple flow diagrams connecting relevant pieces of information. Whatever the approach, the goal is to translate the raw data into meaningful concepts without having to learn complex programming and/or data processing languages. We note that, in our current prototype, this mapping is hardcoded: the development of compelling user interfaces for the middle layer of the ethnography toolkit presents an interesting opportunity for further development.

Finally, the data should be visualized to reveal trends of interest. Here we simply rely on the many open-source Java libraries dedicated to data visualization. For instance, Mondrian (http://rosuda.org/mondrian/) offers compelling visualizations for general-purpose statistics (barcharts, scatter plots, etc.). JUNG (http://jung.sourceforge.net/) and Prefuse (http://prefuse.org/) can be used to visualize social networks. The third component of the ethnography toolkit is therefore a kind of glue allowing ethnographers to select a set of high-level variables and the type of graphs they would like to use. As we did in the Social Dashboard, it is also possible to automate the process further: for instance, automatically analyze the data with machine learning techniques (WEKA from http://www.cs.waikato.ac.nz/ml/weka/ is an open-source, Java-based solution we have used) and isolate trends of interest, which can then be visualized using a graph of the proper type as determined by the ethnography toolkit (e.g. use line graphs to map trends, pie charts for distributions, etc.).

## USING THE DIGITAL ETHNOGRAPHY TOOLKIT TO SOLVE BUSINESS PROBLEMS

A general-purpose digital ethnography toolkit and the variety of environment-specific tools it enables, such as the Social Dashboard, has benefits extending beyond research projects. Indeed, businesses dealing with online communities of any type (from guilds in World of Warcraft to customers exchanging reviews on Amazon.com) have long recognized the need to understand these customers better in order to respond to their demands more efficiently and in a timely manner. More often than not however, this translates into manpower-intensive customer relationship management, whereby many employees (often hundreds, if not thousands) are tasked with "keeping a finger on the pulse" of the community by reading their comments, interacting with them live (in a 3D virtual environment or other electronic settings, like a chat room), or inviting them to participate in activities like focus groups. Note how similar the problems faced by these community managers are to those of digital ethnographers: they are both trying to cover a lot of virtual ground (an entire digital community) to spot events representative of broader trends that they can analyze more finely (in one case, to confirm/inform a research question; in the other, to adjust marketing activities and fine-tune a

business' offerings). In fact, it is reasonable to argue that for Internet businesses, customer relationship management *is* a form of digital ethnography – the only way to understand a community is to be a part of it and observe ever-shifting moods and trends. As such, we believe that the approach we outlined in this paper could be used to solve concrete business problems – we illustrate how with two recent examples from our own work: customer retention and targeted marketing.

The users of virtual worlds (either game-based or more generic) often state that "it is the people that are addictive, not the [environment]" (Lazzaro, 2004). Accordingly, the designers of online social spaces have recognized the importance of such a "people factor" and create environments encouraging social interactions as much as possible. For instance, quests in World of Warcraft's many dungeons (instances) are purposefully designed to be too difficult to complete alone: players need to form groups of five to tackle them and progress, thereby creating opportunities for longer-lasting social bonds (Ducheneaut et al., 2006b). But if creating such bonds increases the environment's "stickiness" and promotes increased usage, it is logical that, conversely, the disappearance of the same bonds makes the environment less compelling and might lead to users eventually leaving it altogether. Therefore, monitoring sociability and interactions in groups can be a powerful way to predict whether or not a given customer will remain engaged for the foreseeable future or defect to another online space. Engaging with customers about to defect can help identify the root causes of their dissatisfaction and which might be remedied with incentives for them to stay (which is all the more important if the customer plays a central social role in their group and their defection would cascade into several more). Note how both automated approaches (monitoring interactions across thousands of groups) and qualitative ones (interacting with key customers to adjust marketing actions) are needed: tools like the Social Dashboard were designed with exactly such a hybrid approach in mind and could, therefore, greatly facilitate customer relationship management in large online spaces where social activities are key to a business's survival.

Efficient marketing also often depends on properly categorizing customers into "segments" with different needs and aspirations. Focus groups have historically been the tool of choice for such segmentation, with customers from different socio-demographic backgrounds being invited to discuss issues important to a business and how it might go about offering possible solutions. We believe that, in online spaces, the use of tools like the Social Dashboard can make such segmenting more straightforward and reliable. In a way, the approach is the reverse of the one described above. Instead of starting with high-level data analysis and monitoring, followed by ethnographic participant observation to uncover the root causes of an issue, businesses interested in refining their marketing segments might start here with digital ethnographic fieldwork to identify individuals in their community with interesting profiles (these vary greatly depending on the business and must be locally defined – to use a concrete example, interesting profiles in World of Warcraft could be players who seem to be very vocal in public chat and might influence opinions, assuming they get positive reactions from the community). The activity traces of these individuals can then be used in the Social Dashboard to see if they are representative of a larger group. This way, it is very easy to quickly see if interesting individuals are representative of a market segment that can be approached with a custom message, or if they are simply idiosyncratic and can be safely ignored. Such an approach therefore enables targeted marketing that is both responsive to newly emerging segments identified during fieldwork and more fine-grained than had been previously possible.

## ETHICAL CONCERNS

Before concluding this paper, we believe it is important to spend some time discussing the ethical implications of the approaches we discussed. While the affordances of virtual worlds provide unique opportunities for conducting research, they also create unique problems with regard to university IRBs as it is often unclear how specific rules and definitions rooted in the physical world translate to virtual settings. The scale and scope of data collected with an automated toolkit such as the one described above also introduces ethical concerns that "regular" ethnographies may not usually have to face. Here, we briefly note several problems to foreground these potential IRB concerns for other researchers.

Minors are a separate class of human subjects in the eyes of the IRB and require parental permission in addition to assent from the minor, in lieu of the standard informed consent. The problem is that it is difficult, if not impossible, to validate parental status or age in cyberspace (i.e., the minor could just lie and say they are 18 or above to begin with). On the other hand, there is a reasonable argument to be made about waiving certain standard requirements given the minimal risk nature of most social science studies and the impracticality of conducting such research otherwise. This, however, depends on the mindset of the local IRB members. We also note that online games and virtual worlds have historically been the province of mostly young users, which explains the IRBs' particular scrutiny of how underage participants are handled by researchers. However, it is important to note that the demographics of virtual worlds users has broadened immensely in the past decade (Williams et al., 2008) and now includes a large majority of adults (including female users, which historically were extremely under-represented). The natural evolution of the population in virtual worlds might therefore mitigate the issue of age and eventually reduce (but certainly not eliminate) the need for additional monitoring by IRBs.

Another provocative set of issues arises related to whether virtual worlds are public spaces. Under IRB rules, a study is exempt from review if it collects passive, observational data in a public space without personal identifiers. The difference between the physical and virtual world is that the latter comes perfectly instrumented (or that it can easily be instrumented, as we proposed earlier). It is possible to collect behavioral data from every avatar in a virtual world continuously, in a way that would be impossible to do in the physical world. However, it is also possible to use techniques (such as one-way hashes) to anonymize character names as the data is being recorded. While IRBs would have no privacy issues about recording people at crosswalks using hidden cameras, it is an entirely different issue when the hidden cameras are used to record people in their own private homes, even if the footage were later anonymized. The core question is therefore what reasonable expectations of privacy users of virtual worlds have, and whether a multi-user avatar space like World of Warcraft is more like a crosswalk or a living room.

This problem is compounded by the relatively recent practice of game developers to make public what would otherwise be private data. A good example would be Blizzard's Armory, mentioned earlier. This raises the possibility of, for example, an employer making judgments of an employee based on the latter's frequent WoW gaming as documented on the Armory. This would be like Facebook status updates that are automatically generated for you and which you cannot edit or hide. But, given that this

information is now in the public domain, can it still be considered private in any sense? And as such, should specific safeguards be enforced by IRBs when researchers make use of this data to guide their observations?

The affordances of virtual worlds are strange in that the lack of validated identity makes certain IRB rules difficult to abide by, but at the same time, the public and panoptic nature of digital systems seems to have created loopholes in our traditional notions of privacy grounded in physical reality. Much of the recent research in virtual worlds has shown that people are heavily invested in their online identity and many social norms carry into these virtual environments. The key question is this: in what ways is analyzing hidden camera data from virtual worlds different from analyzing hidden camera data from people's living rooms? If ethnographic research online is to benefit from the guidance provided by large-scale, automated data collection tools, then ethnographers will have to grapple with this question first-hand instead of obtaining consent from a limited set of well-known participants as is usually the case in the physical world.

## CONCLUSION

We have argued in this paper that the emergence and ever-increasing popularity of large, multi-user virtual worlds presents an opportunity for ethnographers and more quantitatively-inclined researchers to collaborate in ways that were not previously possible. Indeed, the affordances of virtual worlds provide the perfect bridge for quantitative and qualitative researchers. The ability to track and monitor data at both individual and server levels provides a way for mixed-methods team members to complement each other's research. The quantitative team can help identify representative individuals or groups for the qualitative team and the qualitative team can help provide ethnographic explanations for high-level differences identified by the quantitative team.

Moreover, such collaboration does not have to be ad hoc and constantly re-negotiated: our own experience studying virtual worlds shows that the high-level practices necessary for mixed-methods teams to collaborate can be encoded into a "digital ethnography toolkit" that concretely enforces important data collection and analysis techniques, such as: easily deploying an arbitrarily large number of "boots on the ground" (that is, digital robots collecting data directly in the virtual world) and synchronizing their activities; mapping from low-level observations to higher-level, meaningful concepts – guided by the intuition of the ethnographers and the requirements of a given research project; and finally being able to easily observe, at-a-glance, patterns in these concepts and their evolution over time using compelling visualizations. While we have made progress in this direction and implemented some of these steps in our own software, many opportunities remain for further research and development in this area of computer-supported ethnography. We believe, however, that such an approach already has potential for solving concrete business problems: we illustrated the possible benefits in two areas, namely, customer retention and targeted marketing.

While opportunities abound, we also took pain to emphasize that the newfound ease of data collection and analysis offered by an ethnography toolkit is not without ethical risks that practitioners of more "classic" ethnography may not have encountered before. Ethnographers will need to carefully

consider the impact of panoptic data collection on their participants before unleashing the full potential of the new tools at their disposal.

## REFERENCES CITED

Castronova, E.
2003 On virtual economies. *Game Studies*, 3(2).

Castronova, E.
2005 *Synthetic Worlds*. Chicago: Chicago University Press.

Ducheneaut, N.
2003 *The reproduction of Open Source software communities*. Unpublished Ph.D. thesis, University of California, Berkeley, CA.

Ducheneaut, N.
2005 Socialization in an Open Source software community: a socio-technical analysis. *Computer Supported Cooperative Work*, 14(4).

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J.
2006 Building a MMO with mass appeal: a look at gameplay in World of Warcraft. *Games and Culture*, 1(4), 1-38.

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J.
2006b "Alone Together?" Exploring the social dynamics of massively multiplayer online games. *Proceedings of CHI 2006* (pp. 407-416). New York: ACM

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J.
2007 The life and death of online gaming communities: a look at guilds in World of Warcraft. In *Proceedings of CHI 2007* (pp. 839-848). New York: ACM.

Emerson, R. M., Fretz, R. I., & Shaw, L. L.
1995 *Writing ethnographic fieldnotes*. Chicago, IL: The University of Chicago Press.

Fetterman, D.
1998 *Ethnography Step by Step*. Thousand Oaks: Sage.

Hine, C.
2000 *Virtual ethnography*. Sage Publications.

Latour, B., & Woolgar, S.
1986 *Laboratory Life: The Social Construction of Scientific Facts*. Princeton: Princeton University Press.

Lazzaro, N.
2004 *Why we play games: four keys to emotion without story*. Xeodesign.

LeCompte, M. D., & Goetz, J. P.
1982 Problems of reliability and validity in ethnographic research. *Review of educational research*, 52(1), 31-60.

Lyman, P., & Wakeford, N., editors.
1999 *Analyzing virtual societies: New directions in methodology*. Thousand Oaks: Sage.

Malaby, T.
2006 Parlaying value: Capital in and beyond virtual worlds. *Games and Culture*, 1(2), 141-162.

Malinowski, B.
1922 *Argonauts of the Western Pacific*: Waveland Press.

Mason, B.
1999 Issues in virtual ethnography. In K. Buckner (Ed.), *Proceedings of Esprit i3 workshop on ethnographic studies* (pp. 61-69). Edinburgh: Queen Margaret College.

McCall, G., & Simmons, J.
1969 *Issues in Participant Observation*. New York: Addison Wesley.

Moore, R. J., Ducheneaut, N., & Nickell, E.
2007 Doing virtually nothing: awareness and accountability in massively multiplayer online worlds. *Computer Supported Cooperative Work*, 16, 265-305.

Pickering, A.
1995 *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.

Rutter, J., & Smith, G.
2002 Ethnographic presence in nebulous settings: a case study. Paper presented at the ESRC virtual methods seminar series, research relationships and online relationships, CRICT, Brunel University, 19 April 2002.

Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N., & Nickell, E.
2006 From treehouse to barracks: The social life of guilds in World of Warcraft. *Games and Culture*, 1(4), 338-361.

Williams, D., Yee, N., & Caplan, S.
2008 Who plays, how much, and why? A behavioral player census of virtual worlds. *Journal of Computer-Mediated Communication*, 13(4), 993-1018.