The mapping principle, and a research framework for virtual worlds

Dmitri Williams

Annenberg School for Communication

University of Southern California

Virtual worlds have emerged from the faddish corners of the video game industry to become economic powerhouses. At the same time, virtual worlds have become important sites of community (Ducheneaut, Moore, & Nickell, 2007; Steinkuehler & Williams, 2006). The overwhelming success of fantasy-based titles in both the West and the East now places virtual worlds on a par with a console games industry that rivals film and television for consumer dollars and attention. In the West, subscriptions to virtual worlds now number over 47 million (White, 2008), lead by the success of Activision/Blizzard's *World of Warcraft*, which claims over 10 million active accounts. In the East, data are less available, but the game *Maple Story* alone claims to have over 70 million accounts (Sheffield, 2008). These large-scale groups of players have attracted the attention of game developers and advertisers, and the envy of television executives who have seen a steady decline in viewership (Schechner, 2008). It is perhaps not surprising that researchers have also become interested in virtual worlds. Academic research has taken two distinct approaches to virtual worlds. The first is a simple attempt to understand the phenomenon: who is in these spaces and why (D. Williams, Yee, & Caplan, 2008; Yee, 2006)? What kinds of human interactions are taking place there (D. Williams et al., 2006)? Do these extend or conflict with our theories of computer mediated communication (Walther, 2006)? The second is a more radical form of scholarship. Recognizing that these are large-scale, complex, and highly social environments populated by thousands of people, some researchers have realized that they might present parallels to the offline, “real-life” (“RL” to many players) experience (Bainbridge, 2007). And if there is enough of a parallel between our online and offline worlds, tests of human behaviors in one might be able to tell us something about human behaviors in the other.

This approach assumes that these worlds are not only potential parallels, but that since they are far easier to change, tweak and adapt than a parallel offline space, they represent an entirely new venue for social science and humanistic inquiry. The advantages of the space over real-world labs are potentially immense. Testing market forces in the real world is an experiment that risks the livelihood of millions of real people and changes the balance of power between them. In an online space, that same test is relatively low-cost. This has been dubbed the “petri dish” approach to virtual worlds research in which the worlds are altered and studied from the outside (Castronova, 2005, 2006; D. Williams, 2006). Ideally, they will also be measured in comparison to an unaltered control condition world, making the research a true controlled experiment.

The first handful of experiments have been attempted primarily by economics and educational researchers. In some cases the researchers have had the ability to modify the world, but usually not. An astrophysicist has explored the possibility of using virtual worlds to plot out a parallel version of the universe (Evrard, 1999). In her work on science learning among adolescents in a virtual world, Kafai has studied interactions and learning, with the occasional ability to have the world altered to test a principle (Kafai, in press; Neulight, Kafai, Kao, Foley, & Galas, 2007). In one recent case, Castronova and colleagues implemented a treatment and control condition version of a single-player game to test economic behavior (Castronova et al., 2008). These initial outcomes beg an obvious question: Are these behaviors "real?" And by real, we mean, Are these behaviors similar to what we would observe had they occurred in the flesh and blood offline world? This approach asks when and how we might leverage virtual spaces to learn about human behaviors offline.

This is the *mapping principle*. Mapping is the extent to which human behaviors occur in virtual spaces in the same way they occur in real spaces. It is important to note that at this very early stage, mapping is not taken as a given. When starting from scratch, mapping is something that must be established, and most importantly, validated. It is a starting assumption that not all behaviors, not all virtual worlds, and not all contexts within them—and perhaps only a small few—will truly involve mapping. Learning which ones do and which ones do not will be the ultimate goal in establishing virtual worlds as valid spaces to test human behaviors.

A cautionary tale: Virtual plague

It is hard to overstate the importance of establishing validity. The case of the "Warcraft Plague" provides an important cautionary tale. A mistake by the game managers of *World of Warcraft* introduced the equivalent of a highly infectious disease into the game world. Players who had fought a particular monster were infected by it, and the virus was able to jump to any nearby player (Sydell, 2005; Ward, 2005). This virus was supposed to be time-limited and constrained to one particular area of the virtual world, but an oversight by the developer allowed players to teleport back to busy cities with the virus. Once there, they infected the other players, typically killing everyone. This incident garnered academic attention because it offered a possible parallel to real-world epidemiological patterns (Balicer, 2007; Lofgren & Fefferman, 2007), yet some of these researchers took the mapping principle on faith, assuming that the virtual behaviors had fidelity to a real-world context. Some focused instead on the mathematics of the potential tests. The excitement is understandable: If virtual world-based reactions and disease vectors could be used to test how viruses spread and how people react to them, it would be a substantially safer way to test disaster readiness procedures than using a real virus in real space. Would players avoid others or infect them? Would information flow freely or become confused? Would first responders react appropriately? Would people act how they say they would in hypothetical situations given to them on surveys?

The problem was, and remains, that no one knows whether these behaviors map or not. Beyond whether the players would react in one way or another, the more important question for anyone who hopes to use virtual worlds to study diseases and disasters should be “Are these behaviors realistic?” That is, will players in a virtual space dealing with a deadly virus react as people would offline? This would of course be a necessary condition for any “petri dish” form of testing that might take place.

There are, unfortunately, several reasons to be suspicious of this. First and foremost, the risks and rewards of virtual worlds do not always map to those offline. The most obvious example is pain and death. In *World of Warcraft*, players die, but there is no physical discomfort. There is emotional discomfort, which may be powerful in some circumstances, but it does not replicate actual nerve-induced pain. More importantly, though, there is no real death in *Warcraft*. When a player dies, they can be resurrected. The cost is primarily to the ego. In other words, the risks do not map. It should not be surprising then, that players reacting to the plague behaved in ways that on their face were clearly not mapping the offline world. Players were frequently seen dancing and laughing, and trying to hunt down and infect their friends. If the costs and emotional impacts had mapped to real life, the event might have offered enough similarity to real life to be of use to disaster managers and epidemiologists. As it was, the *Warcraft* case clearly did not provide the underlying risks and rewards to parallel the offline case of a disease outbreak. This is not to say that some virtual space might not achieve the sort of mapping required to truly simulate a disease outbreak. What it illustrates is the important need to understand mapping before moving forward with any test. To this end, there are two key necessary conditions to establish mapping. The first is validity, and the second is generalizability. Once those are explicated, it will be appropriate to present a research framework.

Virtual validity and generalizability

Validity is at the heart of using virtual worlds to study real behaviors because of the most basic definition of the term: Validity is the extent to which an instrument measures what it is intended to measure (Carmines & Zaller, 1979). In virtual world research, the instrument might be the entire world, or some particular portion of it. The component types of validity are threefold: face, concurrent and predictive validity (Stamm, 1989). Face validity is simply whether the measure appears, on its face, to measure the phenomenon in question. For virtual worlds, this test is dependent on the phenomenon. If the researcher is interested in studying the cultivation of perceptions about violence in media (Gerbner, Gross, Morgan, & Signorielli, 1994), the virtual world under examination should have some violence and variation within it. *Club Penguin* would lack face validity for such a test, while a violent fantasy MMO might have it (D. Williams, 2006). Face validity is the simplest, and most necessary condition, but it is not sufficient to establish validity.

Concurrent validity is the extent to which the measured phenomena correlate with other measures of the same phenomena (Stamm, 1989). The measuring instrument is checked against some present criterion (Wimmer & Dominick, 1997). For example, if a new test is designed to detect intelligence, giving it to a group of people who have been found by previous tests to be highly intelligent should yield higher scores than giving it to a group of people who had been found to have low intelligence. In the virtual case, this presents a new challenge. Since the virtual world is often self-contained, it can’t be measured with a parallel to the “real world” without leaving it. For example, an economist seeking to validate a measure of inflation (GDP) for a virtual world might be tempted to calculate that measure and then say that since it has worked offline, the measure has concurrent validity. But the essence of concurrent validity is a concurrent test in the same basic area. Therefore, it calls for another measure of a related activity *also within* the virtual world. In this case, a measure of virtual inflation should be able to predict when players will feel that prices are particularly high or low.

The final form of validity, predictive validity, is the most important of all for the mapping principle. Predictive validity tests whether a measure relates to other measures as expected. The virtual researcher has a double burden here. First, the measure must predict the appropriate outcome within the virtual world. Staying with our economist, it is not enough to detect inflation. The measure will only have predictive validity if it also correctly forecasts when spending patterns will change as a result of the inflation. The mapping principle introduces a second, additional test: For the virtual economist, detecting in-world inflation is only of use if that inflation behaves the same way we would expect of real-world inflation. For example, when prices rise, purchasing should drop, as has been observed many times in real space (Landefeld, Seskin, & Fraumeni, 2008; Smith, 1977). A measure of virtual inflation would only be valid if this parallel exists. The economic example is actually fairly easy in this regard since behaviors stemming from inflationary pressures have been well-studied, i.e the offline parallel already exists. In that sense, the test is similar to those in the physical sciences where a model or simulator must function according to known, invariant real-world properties. An example is a scale model test of an ocean liner that must hew to the known properties of fluid dynamics (Rott, 1990). If it does not, there can be no predictive validity for related tests. But what if the test is something that is not well established in the real world? What if, for example, the economist wants to make predictions about what will happen to prices if all of the banks suddenly start charging negative interest to their customers, i.e. borrowers suddenly made money? This scenario is feasible within a virtual space, but highly unlikely in real space. For tests such as these, this predictive validity step may not be possible. It would be roughly similar to an animal trial of a drug that would be difficult to test on humans for ethical or practical reasons.

This form of predictive validity with virtual tests quickly approaches the concept of external validity, i.e. how well the results can be generalized across populations, settings and time. Although external validity is typically used to describe the extent to which lab research is valid outside the lab, it is germane here. In the virtual case, the virtual world becomes the lab, and the test is only externally valid once the pattern detected in the virtual world is also borne out in the real one. And as many have noted, labs are often not like the real world (Cook & Campbell, 1979; Webb, Campbell, Schwartz, & Sechrest, 1966). If a laboratory result cannot be replicated outside of the laboratory, it is relatively useless; the drug that cures cancer only in a petri dish cannot improve anyone’s health. So too, the result found within a virtual world that does not exist outside of it is useless. Worse, like the virtual plague case illustrated, it may be dangerous if taken seriously.

The most obvious defense against this issue is to make the virtual world as similar to the real one as possible with regards to the phenomenon in question. If a test involves human responses to risk, danger, love, velvet, sugar or rewards, then the virtual version of these things must approximate the real one. Some of these may be more likely than others—it is more feasible to make a player feel danger in an online world than to taste sugar. The various potential confounds will need to be tested over time, but some are obvious even at this early stage. For mapping tests involving human interactions, the people involved need to have the same kinds of relationships, interdependencies, contexts, and social networks as their offline versions. For example, if a test of a virtual team is intended to map to a workplace team, then its members should have the same kinds of set up as the workplace. They should have the same kinds of prior engagement, the same forms of hierarchies, similar reward structures, etc. For tests involving difficult choices, the risks, opportunities and rewards need to be as close to real as possible as well (and notably, may face the same kind of IRB scrutiny that the offline test would). For example, if a risk-avoidance study wants to tests sexual promiscuity among virtual partners, then the risks for promiscuity need to be as strongly felt as they are in real life. Of course, no virtual system is going to give a real-life player an STD, but in the hands of a creative experimenter it might deliver annoyances, barriers and social shaming similar to the offline case, and empirically demonstrated to be experienced the same. The key to external validity would be whether the people involved perceived the risks and costs to be as powerful as those experienced offline. For questionable tests like this, a parallel predictive measure would be needed to serve as a virtual/real treatment check. Did a subject in the virtual test report the same levels of fear, danger and social opprobrium as a subject in a parallel real-world trial did?

It is worth noting that some variables may *not* need to be controlled. For example, tests of economic behavior might not require storefronts and cash registers. These are empirical issues that can be largely settled with tests of concurrent and predictive validity. If the tests had strong predictive power for offline economies (the mapping worked well), then it may not matter whether the storefront was a virtual medieval tavern or a virtual Walmart.

The mapping research framework

It is an assumption of this work that not all virtual situations will map to the real. Indeed, since many fail the simple face validity test, it may be that a minority of virtual contexts will map. The goal of the balance of this paper is to lay out a blueprint for discovering which work and which do not. Ideally, researchers in the future will have a rulebook of sorts to know what situations, contexts, levels of analysis and types of human interactions can successfully be tested within virtual spaces, and which cannot. Using the framework, patterns of findings may emerge which suggest the presence of consistent underlying variables that lead to mapping or discourage it. There is another key reason to use this framework; at present, research on virtual worlds is only in its infancy, and there are no central journals, few shared theories and few common practices. Researchers select worlds based on access, personal interest and convenience (the author is no exception) rather than for some systematic purpose. Conclusions are drawn which may or may not apply to other virtual spaces, let alone map to the real world. For the field as a whole to make progress, it is important that there be synthesis between the handful of past studies and future work. This synthesis must coordinate the questions, the levels of analysis and the different forms of virtual worlds. In an effort to organize the work within a larger research agenda, this paper presents a proposed systematic framework resting within the tradition of computer mediated communication (CMC) research.

CMC work has long sought to describe the human social experience when mediated through media technology. The questions have been those previously explored in communications research: what are the effects on community, self-perception, group effectiveness, interpersonal psychology, media effects, etc. (Lowery & DeFluer, 1995; McQuail, 1994). In each case, CMC researchers have asked the McLuhan-inspired (McLuhan, 1964) question of whether and how the new medium or new technology alters the fundamental patterns of human communication (Spears & Lea, 1992; Walther, 2006), group interactions, or individual behaviors. The work here is a direct extension of this tradition, substituting virtual worlds for now-older technologies such as email and phones. It begins with the mapping principle, but can be used as an organizational template for any theory-based program of research on virtual worlds. The purpose of this framework is to answer the deceptively simple CMC question: Do behaviors in virtual worlds map on to behaviors in the "real" world? If they do sometimes and not in others, what predicts the differences?

To answer these questions properly we have to spell out the several ways in which human interactions take place in virtual worlds, the several kinds of behaviors we are interested in, and the many contextual variables that make one virtual world different than the next. Spelling out these differences creates a framework which the community of scholars can begin (and in some cases has begun) to fill.

Table 1 provides the framework, breaking down the four major factors of group size, traditional controls and independent variables, contextual factors, and directionality. Any study of a virtual world should be able to identify where it sits in this table. This will allow meta-level understanding of the research by the larger community, and allow researchers to couch their claims appropriately.

Table 1

A research agenda for virtual mapping, and other tests

Group size	**Traditional controls and independent variables**	**Contextual and social architecture factors**	**Directionality**
Individual	Psychological profile	World size	Online to offline
Dyads	Motivations	Persistence	Offline to online
Small groups	Demographics	Competitive vs. Collaborative	Endogenous
Large groups	Communications medium	Role play	
Communities	Network-level variables	Sandbox vs. linear	
Societies		Representation	
		Interaction affordances	
		Costs of a behavior	

The first factor is straightforward. The unit of analysis of a study describes what level of the hierarchy the researcher is examining. Since virtual worlds are typically similar in their social structures to the real world, they also have collections of people ranging from individuals to very large groups. Group size is important to identify for the simple reason that some researchers tend to study large-scale community and society-level processes while others are focused on individuals or small groups. For example, several studies have focused on why individual players choose to play (Ducheneaut, Yee, Nickell, & Moore, 2006a; D. Williams et al., 2008; Yee, 2006), while others focus on the characteristics of small groups (Bailenson, Beall, Blascovich, Loomis, & Turk, 2005) and still others are concerned with large groups or whole societies (Castronova, 2001). One study sought to lay out the distribution of these demographic group patterns (Ducheneaut, Yee, Nickell, & Moore, 2006b). The point of the first factor is that these studies are not interchangeable. Conclusions drawn from the study of individuals cannot automatically be applied to collective groups because behaviors may change as the group size increases. Likewise, studies of large groups should not be used to make predictions about individuals because behaviors found in large groups may change as the group size decreases. Some results may well travel up and down the group size factor, but these must be tested and validated. Until then, findings should be plugged into the framework at the appropriate level of analysis, and any claims should be limited.

The second factor is considered essential within the CMC tradition. Any study of human behaviors and interactions must account for the background of those involved, if only because demographic categories may predict a large portion of the outcomes. This is to say that virtual world research should not be treated any less rigorously than other CMC work. Humans online are still humans, and the standard psychological issues of profile, personal background, and motivation found in other communication research will still apply. Early work on Internet use showed that personality (Kraut et al., 2002; Kraut et al., 1996) and self-disclosure (Bargh, McKenna, & Fitzsimmons, 2002) had large effects on relationships and community. Likewise, the traditional tests of communications medium must be confronted. Although it is not always obvious to outsiders (and sometimes contentious among players), virtual world users communicate with each other through a wide variety of modalities ranging from text to audio (VoIP) to video, and in “locations” ranging from in-world to message boards to face-to-face meetings. The moderating value of these different channels is not well understood. Moreover, the Web 2.0 world of players is comprised not of mutually exclusive channels and modalities, but of mixes and combinations. One initial test suggested that the use of voice and text, compared to text only, caused significant differences in how much virtual group members trusted each other (D. Williams, Caplan, & Xiong, 2007). Lastly, the new and burgeoning field of network science offers special appeal to virtual world research. Theorists have suggested that position within social networks has great predictive power for understanding motivations, group behaviors, information flows and many other outcomes (Monge & Contractor, 2003). Noting the position of the subjects within a network thus becomes another key variable to identify when comparing studies. Network measures are more attractive still when unobtrusive behavioral data become available. Networks can be more easily built from virtual world data than from real-world data as they are based on recordable actions such as chatting or teaming.

The third factor introduces the most complexity, and the biggest challenge for generalizability. To the new or inexperienced user, virtual worlds may seem to be all very similar: there is a space, some representation of the user, and some apparent goal. Yet a more careful examination reveals that even these apparent basics vary from one world to the next. Spaces can be text-based, two dimensional or three dimensional. Representations of the user might be as simple as a stick figure, a line of text, a hand holding an object in the immediate foreground, or a fully detailed avatar of a figure. And while some worlds have clear goals, e.g.

slaying a dragon in *World of Warcraft*, others have no explicit goals at all, e.g. the "open sandbox" style of virtual worlds as exemplified by *Second Life* or *There*.

Therefore a strong assumption of this framework is that we cannot automatically treat virtual worlds as equivalent to one another. The reasons for this lie in the concepts of code and social architecture. Lessig (1999) first articulated the concept "Code is law" as a way of drawing attention to the fact that behavior in virtual spaces is governed by software as much as by laws, markets or social norms. In real space we can take it for granted that we can walk, but not fly, or talk with people who are in hearing distance. These abilities and limitations are not safe to assume in virtual spaces where the affordances and limitations of human actions and interactions are whatever the code says they are. Indeed, they may all be flipped. Code may also control who can interact with whom, when and how. For example, some virtual worlds create subgroups where one set of people may communicate with only others in their subgroup. Some worlds have templates for social organization and hierarchies within groups. The sum total of these interaction affordances and limitations can be thought of as the "social architecture" of the space (Kim, 2000). Like regular architecture, it governs and impacts behaviors and interactions, but unlike a wall, bar or podium, it is less apparent to the users.

If research from a highly restrictive social world is used to predict behaviors within a highly open social world, the potential for mistakes increases. Even within single games there are distinct spaces and versions of the rules. Among the most common is the difference between "PvE," "PvP" and "RP" servers. These allow the ability to attack other players or not, or the expectation that players will maintain a fictitious persona. These differences could have a large moderating effect on processes such as aggression, interpersonal dynamics, group effectiveness and the motivations of the players who self-select into these spaces. Yet even when comparing two games with similar structures and limiting the analysis to PvP, care should be taken to identify the costs and risks. One game may offer a light penalty for being killed by another player, while another may have harsher consequences. For example, players of *Lineage II* who kill another player are flagged to the rest of the players, who are then given an incentive to hunt down the killer and possibly take a valuable item from them as a trophy. The defeated player may also lose an item and suffers a large time-based penalty (up to an hour) to regain their position. In contrast, players of *Warcraft* who kill another player gain points for doing so, and the defeated player typically suffers a one-minute penalty to regain their position. These architectural differences are likely to lead to different behaviors.

In some game worlds these rule sets are bright lines, while in other worlds they may be less clear or not zoned off from each other. Sandbox-style worlds—most popularly *Second Life*— represent a drastically reduced set of rules and versions from the MMO model. *Second Life* is made up of thousands of distinct subcultures, whose activities are sometimes governed by code, but more often governed by local norms (Robbins & Bell, 2008). Until we have verified that outcomes are the same or different between various kinds of worlds and their social architectures, care should be taken not to generalize too widely. Certainly there are many commonalities between the most popular spaces; fantasy-based MMOs make up over 85% of the virtual world market (White, 2008), and many feature similar game mechanics. However, a test of economic activity in these similar spaces might still find that they generate very different outcomes because their markets are organized differently. And because local cultures and code can have a large impact on study outcomes, participant observation work takes on a much larger significance than is the norm in communication research (D. Williams, 2005).

The last factor is directionality. Some virtual world research is concerned with the impacts that time online might have on our offline lives (D. Williams & Skoric, 2005), while other research is concerned with how our real lives impacts what goes on within virtual worlds (Yee, 2007). If the patterns of more general Internet research (Wellman & Haythornthwaite, 2002) are borne out again in virtual world work, future investigations will examine the more cyclical process of how these spaces are embedded within the context of everyday life. For researchers interested in causal issues, this will be particularly important to identify.

Case study for the framework

Again, the major issue at hand is whether behaviors in virtual worlds map on to those in the real world. An important and innovative series of studies by Bailenson, Yee and colleagues can be used as a test case for the framework. One subset of findings and papers has outlined instances where the space has impacted human behaviors; the “Proteus Effect” demonstrates that assuming a particular kind of avatar (e.g. tall) leads to different perceptions and behaviors during interactions (Yee & Bailenson, 2007). Tall avatars are treated with the same preference as tall people are offline—whether the unseen user was tall or short. In this case, the answer to the original question is that the virtual world did map onto the real one, at least for impressions of appearance. That same team also found that many interpersonal dynamics from the real world (e.g. social distance and eye contact) map nearly perfectly into virtual spaces, including offline gender differences (Bailenson et al., 2005; Yee, Bailenson, Urbanek, Chang, & Merget, 2007)—although it was not clear if the user’s gender matched the avatar’s. Tellingly, the study’s goals were a paraphrase of the mapping principle: “to explore whether social norms of gender, interpersonal distance (IPD), and eye gaze transfer into virtual environments” (Yee et al, 2007, p. 115). These several findings can now be categorized and put into the framework to enable to meta-level statements about mapping.

For the first factor (group size), the work was concerned with the interpersonal, small-level realm of activity, typically focusing on dyads. It is unclear whether the same findings would hold up as group sizes increase. Would individual “physical” appearance still lead to differences in liking, affect, etc. as group sizes increased to five, thirty or two thousand? Perhaps they would lessen, or widen. Bailenson and Yee’s work in this area is in many ways a product of the earlier work done on a virtual reality environments. Because virtual worlds have much in common with virtual reality (typically 3D avatars and human interactions), it is tempting to apply the body of results from VR to the mapping principle. And indeed, there is a tradition of immersive environment research within psychology that shows effects on small scales and in limited environments (Blascovich et al., 2002). In behaviors as varied as compassion (Gillath, McCall, Shaver, & Blascovich, 2008), leadership (Hoyt & Blascovich, 2003), ostracism (K. Williams, Cheung, & Choi, 2000) and racism (Dotsch & Wigboldus, 2008), this work has shown mapping-style results with dyads, where offline expectations were met in virtual spaces. However, these behaviors may not apply to larger groups or (as the third factor will demonstrate) to larger-scale and more open-ended social settings such as MMOs or large avatar-based sandbox worlds. Another key difference is that some of this research involves humans and computer-based agents rather than human-human interactions (Dotsch & Wigboldus, 2008; Gillath et al., 2008; Guadagno, Blascovich, Bailenson, & McCall, 2007). Although the authors sometimes suggest that the work has high ecological validity (Gillath et al., 2008), this should not be confounded with the mapping of human behaviors online and off. There is a key distinction to be made when the studies use computer agents rather than real people (confederates or otherwise) driving the avatars. The outcomes may be the same as reactions to human-controlled avatars, or not. Research has suggested that humans work hard to learn whether an avatar is human, and may react differently when it is not (Nass & Gong, 2000), with visceral negative reactions if they feel as if they are trying to be tricked (Mori, 1970; Serviss, 2005). Until this work is demonstrated to have offline predictive validity, these forms of study contribute to our understanding of presence and our interactions with virtual agents, but not to the online-offline mapping concept.

For the second factor (controls and traditional independent variables), the tests were not focused on the profiles of the users. This is perfectly reasonable as the intent of the experiments was to establish the presence of the phenomenon, not to explore the nuances right away. Only gender was examined, leaving aside possible motivational or personality-based differences such as extroversion or social control. The communications medium was also relatively poor, with most users relying on text for conversations. Perhaps an awareness or impression of real-life characteristics disclosed by the use of voice would change the outcomes of the research, e.g. the finding that female-female avatar dyads remained physically close might change if one of the females revealed their real-world maleness by speaking in a deep voice. It is also conceivable that the relative position of avatars within some social hierarchy or network might impact the amount of deference or familiarity felt, which might impact eye gaze or closeness.

For the third factor (social architecture), it is important to contextualize exactly where the studies took place, what the social architectures were, and how they might differ from other virtual worlds. The 2007 eye gaze study (Yee et al., 2007) examined naturally occurring (if such a thing can be said of virtual spaces) interactions within *Second Life*, while the 2005 eye gaze study used undergraduate volunteers in a more controlled lab setting (Bailenson et al., 2005). Since both found similar eye gaze results from different virtual settings, the power of the conclusions is stronger than if had been tested in just one space. However, as has been mentioned earlier, *Second Life* is a "sandbox" style virtual world, with great freedom of movement and choice. It, and the lab setting are substantially different from a game-based virtual world such as *Club Penguin* or *Warhammer Online*, which differ in several ways. Most obviously, those spaces are filled with more stock fantasy characters such as penguins, orcs and knights, whereas *Second Life* tends to the slightly more human (exceptions certainly occur). Would penguin avatars have the same eye gaze outcomes, or would controlling a non-human avatar change the mechanism driving the outcomes? Would being in a world with violence or no violence make a difference? Would the presence of game-based tasks change these interactions, i.e. will players on a mission to slay the evil dragon have the same kinds of interpersonal outcomes as users relaxing in a virtual nightclub?

For the final factor (directionality), this series of research experiments can be described as exploring whether the real impacts the virtual. That is, in asking what real-life interpersonal behaviors carry over into virtual worlds, the causal arrow points from the real to the virtual. These studies are not exploring whether the virtual impacts the real, or whether the two realms might end up reinforcing one another.

To sum up the findings from the meta-level perspective of the framework, the authors made a substantial contribution by establishing the presence of the phenomenon, but we cannot interpret their work as automatically applying to difference group sizes, for all kinds of people, in all virtual settings, with all communication modalities, and in all kinds of networks. Rather, their work should be taken as an important baseline, or starting point. As future researchers tackle these issues of virtual interpersonal behaviors, they would contribute most by examining other group sizes, exploring different kinds of users, looking in different kinds of spaces, and asking whether these same findings might occur in other causal directions. As each factor is tackled and the results added into the framework, our overall understanding of interpersonal avatar-based behavior will gain nuance and power.

Considerations for future work

Recent work has begun to tap into not only the small-group dimensions of virtual worlds, but to start exploring the community and society-level as well. The XEROX PARC team has begun analyzing entire servers of players, charting interactions and group dynamics with software bots and spiders (Ducheneaut et al., 2006a, 2006b). The *EverQuest II* studies use unobtrusive behavioral data to study the recorded behaviors of both individuals and aggregates within a fantasy setting (D. Williams et al., 2008). Such methods allow data retrieval and analysis on scales of immense size. Typically, large-scale analysis in social science relies on estimates, surveys and sampling, while smaller-scale work utilizes direct observations such as controlled experiments, participant observation and ethnography. Although both ends of the scale spectrum are useful, they also have well-known advantages and disadvantages. Small-scale work often risks sampling biases and generalizability issues, while large-scale work rarely incorporates detail and particular social contexts. In other words, one end of the spectrum typically suffers from internal validity challenges, while the other typically suffers from external ones. For both, human subjects are usually aware of the research being conducted, and so the natural attitudinal and behavioral responses collected from human subjects are often inseparable from those induced by Hawthorne Effects (Cook & Campbell, 1979) and social desirability.

The new methods being used by virtual world researchers can reduce or eliminate these problems. Data for an entire population, instead of a sample, have been collected and accessed. And because all data are collected, there is no need for estimation or sampling—given enough storage space, every single action, transaction, and interaction can be tracked, accumulated and analyzed. Virtual worlds are excellent sites for observation and data collection, where researchers become nearly omniscient about every aspect of player behaviors, at least within the virtual world. In addition, this recording happens without the players' knowledge, so demand conditions cannot be a factor (Webb et al., 1966). Practically speaking, this is an unobtrusive methodology that could not have existed before.

If there is an immediate drawback to using this approach for virtual world research, it is that there is rarely an opportunity to introduce control—unless the researcher becomes fortunate enough to work with a world operator to do so, or until such tools become more accessible to universities. Natural experiments may offer some potential, as when a commercial developer introduces a change into a virtual world that relates to a theoretical question. Tinkering and altering the very structure of a world has obvious appeal to experimentalists, but we may not always have a full appreciation of the moving parts involved. While virtual world researchers have become recently aware of the immense power of social architectures, code and rule sets to govern and influence human behaviors (Kim, 2000; Lessig, 1999), these are hardly new concepts to urban planners, architects, anthropologists and the like. Indeed, it is a hallmark (often contested) of those many fields that structure influences behaviors.

Social scientists and humanists of all stripes might find fertile ground to test, replicate and extend the many theories of human behavior that are too costly or unwieldy to perform in the real world. Perhaps urban planners could model traffic flows, community interaction and commerce patterns in a virtual setting rather than rolling out hotly contested subdivisions. Perhaps political scientists examining trust and social diversity could see the implications of changes in the law without having to persuade actual legislators—instead simply adjusting a few lines of code to change the rules and incentives of a community. The possibilities are both exciting and daunting, but there are two caveats. First, the mapping principle has to be tested for internal and external validity before any result can be said to truly “matter” beyond the confines of virtual spaces. A virtual result is just that, and until there is a mapped real-world parallel of conditions, incentives and assumptions, the result will remain relatively trivial. Second, the ethics of virtual research should be no different than real-world ones. Milgram’s shock experiments were essentially virtual in nature as well (Milgram, 1963), and while they provided deep insight into human nature, they did so at a high ethical cost. Humans in virtual spaces are still humans, and their psychology and life contexts are still important to protect.

There is one further issue virtual methodologists might encounter. If the tests are validated, undertaken with care and rigor, and then find results which challenge some existing theory, it is nearly a given that the method will be ignored or attacked. This is the norm in modern science (Kuhn, 1961). However, Kuhn notes that leaps in progress are always made when there is a gap between what is expected and what is found. Done conservatively, virtual methods have the potential to reveal many such gaps, and practitioners should be prepared for the reaction.

In conclusion, the field of virtual worlds research is poised to take off. To newcomers to the field or to laypersons, virtual worlds are likely a black box. *World of Warcraft* and *Habbo Hotel* are no different to those who do not understand the conventions and options found from world to world. Unlike television or books, which nearly everyone understands have different forms, genres and audiences, virtual worlds are not nearly as well understood. Moreover, the differences between worlds are made more complex by the presence of very different social architectures. Books, for example, are generally read by one person and not experienced in dozens of ways by dozens of different combinations of readers, each of whom impacts the way the story turns out. It is up to researchers to educate the public and ourselves about the key distinctions between spaces. Otherwise, it will be easy for journalists or novice researchers to find a result and to make large and irresponsible claims to a public which may not know better. This current work has suggested some of the many ways in which we might advance the field responsibly with scientific rigor, coordination and systematic inquiry.

References

Bailenson, J., Beall, A., Blascovich, J., Loomis, J., & Turk, M. (2005). Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Human Communication Research, 31*, 511-537.

Bainbridge, W. (2007). The scientific research potential of virtual worlds. *Science, 317*, 472-476.

Balicer, R. (2007). Modeling infectious diseases dissemination through online role-playing games: Virtual epidemilogy. *Epidemiology, 18*(2), 260-261.

Bargh, J. A., McKenna, K. Y. A., & Fitzsimmons, G. M. (2002). Can you see the real me? Activation and expression of the "true self" on the Internet. *Journal of Social Issues, 58*(1), 33-48.

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., & Bailenson, J. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*(2), 103-124.

Carmines, E., & Zaller, R. (1979). *Reliability and validity assessment*. Newbury Park, California Sage Publications.

Castronova, E. (2001). Virtual Worlds: A First-Hand Account of Market and Society on the Cyberian Frontier. *The Gruter Institute Working Papers on Law, Economics, and Evolutionary Biology, 2*(1), 1-68.

Castronova, E. (2005). *Synthetic worlds: The business and culture of online games*. Chicago: University of Chicago Press.

Castronova, E. (2006). On the research value of large games: Natural experiments in Norrath and Camelot. *Games and Culture, 1*, 163-186.

Castronova, E., Bell, M. W., Carlton, M., Cornell, R., Cummings, J. J., Emigh, W., et al. (2008). A test of the law of demand in a virtual world: Exploring the Petri dish approach to social science. CESifo Working Paper No. 2355. . from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1173642

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin Company.

Dotsch, R., & Wigboldus, D. (2008). Virtual prejudice. *Journal of Experimental Social Psychology, 44*(4), 1194-1198.

Ducheneaut, N., Moore, R., & Nickell, E. (2007). Virtual "third places": A case study of sociability in massively multiplayer games. *Computer Supported Cooperative Work, 16*, 129-166.

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. (2006a, April 22-28). *"Alone together?" Exploring the social dynamics of massively multiplayer online games.* Paper presented at the CHI 2006, Montreal, Quebec.

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. (2006b). Building a MMO with mass appeal: A look at gameplay in World of Warcraft. *Games and Culture, 1*(4), 281-317.

Evrard, A. (1999). Perspective: Real or virtual large-scale structure. *Proceedings of the National Academy of Science, 96*, 4228-4231.

Gerbner, G., Gross, L., Morgan, M., & Signorielli, N. (1994). Growing up with television: The cultivation perspective. In J. Bryant & D. Zillman (Eds.), *Media Effects: Advances in Theory and Research* (pp. 17-41). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Gillath, O., McCall, C., Shaver, P., & Blascovich, J. (2008). What can virtual reality teach us about prosocial tendencies in real and virtual environments? *Media Psychology, 11*(2), 259-282.

Guadagno, R., Blascovich, J., Bailenson, J., & McCall, C. (2007). Virtual humans and persuation: The effects of agency and behavioral realism. *Media Psychology, 10*, 1-22.

Hoyt, C., & Blascovich, J. (2003). Transformational and transactional leadership in virtual and physical environments. *Small Group Research, 34*(6), 678-715.

Kafai, Y. (in press). Understanding virtual epidemics: Children's folk conceptions of a computer virus. *Journal of Science Education and Technology*.

Kim, A. J. (2000). *Community building on the Web: Secret strategies for successful online communities*. Berkeley, California: Peachpit Press.

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgegson, V., & Crawford, A. (2002). Internet paradox revisited. *Journal of Social Issues, 58*(1), 49-74.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukhopadhyay, T., & Scherlis, W. (1996). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist, 53*, 1011-1031.

Kuhn, T. (1961). The Function of Measurement in Modern Physical Science. *Isis, 52*, 161-190.

Landefeld, J., Seskin, E., & Fraumeni, B. (2008). Taking the pulse of the economy: Measuring GDP. *Journal of Economic Perspectives, 22*(2), 193-216.

Lessig, L. (1999). *Code and other laws of cyberspace*. New York: Basic Books.

Lofgren, E., & Fefferman, N. (2007). The untapped potential of virtual game worlds to shed light on real world epidemics. *The Lancet, Infectious Diseases, 7*(9), 625-629.

Lowery, S., & DeFluer, M. (1995). *Milestones in mass communication research: Media effects*. White Plains, New York: Longman Publishers USA.

McLuhan, M. (1964). *Understanding media: The extensions of man*. Cambridge, Massachusetts: The MIT Press.

McQuail, D. (1994). *Mass communication theory*. London: Sage Publications Ltd.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal Social Psychology, 67*, 371-378.

Monge, P. S., & Contractor, N. S. (2003). *Theories of communication networks*. Oxford: Oxford University Press.

Mori, M. (1970). Bukimi no tani (The uncanny valley). *Energy, 7*(4), 33-35.

Nass, C., & Gong, L. (2000). Speech interfaces from an evolutionary perspective: Social psychological research and design implications. *Communications of the ACM, 43*(9), 36-43.

Neulight, N., Kafai, Y., Kao, L., Foley, B., & Galas, C. (2007). Children's participation in a virtual epidemic in the science classroom: Making connections to natural infectious diseases. *Journal of Science Education and Technology, 16*(1), 47-58.

Robbins, S., & Bell, M. (2008). *Second Life for dummies*. Hoboken, New Jersey: Wiley Publishing, Inc.

Rott, N. (1990). Note on the history of the Reynolds number. *Annual Review of Fluid Mechanics, 22*, 1-11.

Schechner, S. (2008, October 22). Network audience keeps eroding; Upswing in delayed viewing on DVRs isn't likely to offset prime-time declines. *Wall Street Journal*,

Serviss, B. (2005). Escaping the world: High and low resolution in gaming. *IEEE MultiMedia, 12*, 4-8.

Sheffield, B. (2008, April 11). Understanding free-to-play: Nexon's Min Kim speaks out. *Gamasutra,* 5.

Smith, A. (1977). *An inquiry into the nature and causes of the wealth of nations*. Chicago: University of Chicago Press.

Spears, R., & Lea, M. (1992). Social influence and the influence of the 'social' in computer-mediated communication. In M. Lea (Ed.), *Contexts of computer-mediated communication* (pp. 30-65). New York: Harvester Wheatsheaf.

Stamm, K. (1989). Measurement decisions. In G. S. III & B. Westley (Eds.), *Research methods in mass communication* (pp. 90-108). Englewood Cliffs, New Jersey: Prentice Hall.

Steinkuehler, C., & Williams, D. (2006). Where everybody knows your (screen) name: Online games as "third places". *Journal of Computer-Mediated Communication, 11*(4).

Sydell, L. (Writer) (2005). "Virtual" virus sheds light on real-world behavior [Radio Broadcast]: NPR.

Walther, J. (2006). Nonverbal dynamics in computer-mediated communication, or :(and the net :('s with you, :) and you :) alone. In V. Manusov & M. Patterson (Eds.), *The Sage handbook of nonverbal communication* (pp. 461-480). Thousand Oaks, CA: Sage.

Ward, M. (2005). Deadly plague hits Warcraft world. Retrieved September 26, 2005, from http://news.bbc.co.uk/2/hi/technology/4272418.stm

Webb, E., Campbell, D., Schwartz, R., & Sechrest, L. (1966). *Unobtrusive measures: Non-reactive research in the social sciences*. Chicago: Rand McNally and Company.

Wellman, B., & Haythornthwaite, C. (2002). *The Internet in everyday life*. Malden, Massachusetts: Blackwell Publishing.

White, P. (2008). *MMOGData: Charts*. Gloucester, United Kingdom.

Williams, D. (2005). Bridging the methodological divide in game research. *Simulation & Gaming, 36*(4), 447-463.

Williams, D. (2006). Virtual cultivation: Online worlds, offline perceptions. *Journal of Communication, 56*(1), 69-87.

Williams, D., Caplan, S., & Xiong, L. (2007). Can you hear me now? The social impact of voice on internet communities. *Human Communication Research, 33*(4), 427-449.

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

Williams, D., & Skoric, M. (2005). Internet fantasy violence: A test of aggression in an online game. *Communication Monographs, 72*(2), 217-233.

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

Williams, K., Cheung, C., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the Internet. *Journal of Personality and Social Psychology, 79*(5), 748-762.

Wimmer, R. D., & Dominick, J. R. (1997). *Mass Media Research: An Introduction* (5th ed.). Belmont, California: Wadsworth Publishing Company.

Yee, N. (2006). The demographics, motivations and derived experiences of users of massively-multiuser online graphical environments. *PRESENCE: Teleoperators and Virtual Environments, 15*, 309-329.

Yee, N. (2007). Motivations of play in online games. *CyberPsychology & Behavior, 9*, 772-775.

Yee, N., & Bailenson, J. (2007). The Proteus effect: Self transformations in virtual reality. *Human Communication Research, 33*, 271-290.

Yee, N., Bailenson, J., Urbanek, M., Chang, F., & Merget, D. (2007). The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior, 10*, 115-121.