New Media & Society

The virtual census: representations of gender, race and age in video games

Dmitri Williams, Nicole Martins, Mia Consalvo and James D. Ivory

The online version of this article can be found at:

Published by:
$SAGE

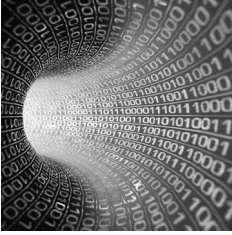**Additional services and information for *New Media & Society* can be found at:**

Email Alerts: http://nms.sagepub.com/cgi/alerts

Subscriptions: http://nms.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.co.uk/journalsPermissions.nav

Citations http://nms.sagepub.com/cgi/content/refs/11/5/815

ARTICLE

The virtual census: representations of gender, race and age in video games

DMITRI WILLIAMS
University of Southern California, USA

NICOLE MARTINS
Indiana University, USA

MIA CONSALVO
Ohio University, USA

JAMES D. IVORY
Virginia Polytechnic Institute and State University, USA

Abstract

A large-scale content analysis of characters in video games was employed to answer questions about their representations of gender, race and age in comparison to the US population. The sample included 150 games from a year across nine platforms, with the results weighted according to game sales. This innovation enabled the results to be analyzed in proportion to the games that were actually played by the public, and thus allowed the first statements able to be generalized about the content of popular video games. The results show a systematic over-representation of males, white and adults and a systematic under–representation of females, Hispanics, Native Americans, children and the elderly. Overall, the results are similar to those found in television research. The implications for identity, cognitive models, cultivation and game research are discussed.

815

Video games have become a widely popular and highly profitable medium, with more than 40 percent of Americans now playing them regularly (Slagle, 2006). A majority of adults age 18 and older (53%) play video games and approximately one in five adults (21%) play every day or almost every day (Lenhart et al., 2008). In fact, video games surpass television in terms of time spent among some populations (Sherry et al., 2006). It follows that if games are a significant portion of the media diet, they need to be understood as important systems of symbols which might have a broad social impact. In the same vein that television has been thought to create cultivation effects (Gerbner et al., 1994) and to have an impact on the cognitive modeling of social identity formation (Mastro et al., 2007), games also may be influencing players' impressions of social groups, including their own (Comstock and Cobbey, 1979). Content analyses of mainstream media have demonstrated where portrayals of gender, race and age have diverged from actual group proportions in the US population (Harwood and Anderson, 2002). However, despite the popularity of video games, there is a gap in our understanding of such portrayals across the wide range of game titles and as understood from the consumer's viewpoint. Past research has focused on convenience samples of game titles, but never in proportion to what is actually played. Sampling and weighting games according to popularity will allow a connection between research and actual social practice.

The existing content analytic work done on video games has focused on two topics of special interest to communication researchers: violence (Dietz, 1998; Heintz-Knowles et al., 2001; Schierbeck and Carstens, 2000; Shibuya et al., 2004; Smith et al., 2003; Thompson and Haninger, 2001; Thompson et al., 2006), and gender and sexuality (Braun and Giroux, 1989; Dietz, 1998; Downs and Smith, 2005; Heintz–Knowles et al., 2001; Janz and Martis, 2007). Although these studies are important steps in examining videogame content, there is still much left to discover, including a more basic study of representation. In the work presented here, this study seeks to obtain a baseline measure of race, gender and age distribution across the current universe of videogame characters. Because media character demographics and portrayals of social groups may influence players' likelihood of attending to and learning from game characters (Bandura, 1994), as well as players' perceptions of social reality (Gerbner et al., 1994; Shrum, 1999), establishing sound baseline measures of videogame character demographics is a necessary

step in applying theories of influence, identity construction and perceived social reality.

PRIOR CONTENT ANALYSES

Prior studies of game content have been concerned with the sensitive topics of violence and sexuality, while typically not focusing on the generalizability of the findings to the universe of actual content. Although violence is outside the scope of this article, there are strengths and weaknesses in this body of literature that inform the present study. Most notably, these concern the gender and race breakdowns of prior samples, the inclusion of the Entertainment Software Rating Board (ESRB) ratings system as an independent variable and the sampling frame used to collect the games.

For example, Dietz (1998) examined violence and gender stereotyping in her content analysis of 33 popular Nintendo and Sega Genesis videogames. She found that the most common portrayal for female characters was the complete absence of females. In fact, there were no female characters in more than 40 percent of the games that she sampled. Heintz-Knowles et al. (2001) also examined violence and gender stereotyping in their study of 70 video games. The results revealed that of the 874 characters coded, 73 percent were male and 12 percent were female. When females did appear, they were likely to be seen in secondary roles. Dill and colleagues (2005) coded the role of each character encountered in their content analysis of 20 top-selling PC video games of 1999. The results revealed that across all 20 games, 70 percent of the primary characters were male and 10 percent were female. For secondary characters, 55 percent were male and 31 percent were female. Dill and colleagues (2005) also found that more than two-thirds of the main characters were white (68%), followed by Latino (15%) and black (8%).

Game ratings have received attention in the literature and should be examined as a possible source of content variation. Created in 1994, the ESRB rates videogames with age-based symbols and content descriptors. Games rated 'E' (for everyone) have been deemed suitable for players aged six years and older. In contrast, games rated 'AO' are suitable for adults only. Studies of the ESRB ratings system suggest that games with different rating levels do have different kinds of content (Thompson et al., 2001), yet not all content analyses of games take ratings into account.

Sampling has proven to be a barrier in prior content analyses of video games, with many studies examining limited subsets of games or games released for systems that never became popular. For example, the comprehensive 'Children Now' study (Heintz-Knowles et al., 2001) featured games played on the Sega Dreamcast, a system that flared briefly and then sputtered in the late 1990s. To make things more complex, most studies rely on one or two systems, but in the current environment this severely limits

a sample. In the present very competitive games market there are no fewer than nine major and viable systems representing PCs, consoles and hand–held options for players. Any comprehensive content analysis would need to include each of these.

A key obstacle in prior work has been linking the small samples used to the games that are actually consumed by the wider public. In the absence of a Nielsen-like system, researchers have been unable to connect content data with the actual practices of players. The Heintz-Knowles et al. (2001) study examined the 10 top-selling video games for each of the six video game consoles available at the time (Dreamcast, Game Boy Advance, Game Boy Color, Nintendo 64, Playstation, Playstation 2 and PC), resulting in a total sample size of 60 games. Other pioneering studies have limited their samples to games from a handful of platforms. Dill et al. (2005) analyzed 20 top-selling games, but they were for PCs only. Downs and Smith (2005) had one of the largest samples when they analyzed 60 top-selling games for the Microsoft Xbox ($N = 20$), Nintendo Game Cube ($N = 20$) and Sony Playstation 2 ($N = 20$), yet none of these studies considered any of the games to be more influential than any other. However, when the most popular game (*Madden '06*) sells over 6 million copies and the least popular (game #150 in the present study's sampling frame, *BeyBlade*) 15,000 copies, it is safe to assume that one game will be played significantly more than another. Thus, if the goal is to measure what the public is actually consuming, content from the two should not be given equal weight in the analysis.

A final limitation concerns the absence of handheld game systems from content analytic samples. With the exception of the 'Children Now' study, no content analysis to date has examined character portrayals on handheld games. This is an important omission to note, because these systems are marketed to young children and adolescents. The present study sought to address these previous limitations by employing a large sample, including games from every major platform, utilizing ratings schemes and sampling with a scheme that includes the actual popularity of games in the resulting content analysis.

BACKGROUND

Why game representations matter

There are several reasons why the presence, absence or type of portrayal of social groups matter in a diverse society, ranging from social justice and power imbalance to models of effects and stereotype formation. Harwood and Anderson (2002) have suggested that representation on television is at heart a proxy for other social forces – that is, groups who appear more often in the media are more 'vital' and enjoy more status and power in daily life. Their use

of ethnolinguistic vitality theory argues that the media work as a mirror for existing social forces as much as a causal agent of them. Therefore, measuring the imbalances that exist on the screen can tell us what imbalances exist in social identity formation, social power and policy formation in daily life.

Moving past the media as a mirror for social power relations, several theories offer models and explanations for the reason why the consumers of media may be affected by them. Cultivation theory posits that the world of media exerts a broad, 'gravitational' pull on the viewer, systematically shaping their worldview to match that of the symbolic one on TV (Gerbner et al., 1994). This work has remained highly contested and controversial (Hirsch, 1981; Potter, 1994). Moreover, an experiment of cultivation in a video game (Williams, 2006b) has shown that the mechanism was precise and targeted rather than broad and spreading, supporting Shrum's (2002) cognitive processing version of the theory. In other words, it was a specific set of symbols that yielded cultivation effects rather than a broader set of values or cultures.

The theoretical mechanism in Shrum's approach suggests that the presence (or absence) of a set of images in media causes a set of impressions in viewers (or players) through well-studied cognitive mechanisms. Price and Tewksbury (1997) reviewed this literature on cognitive associations, priming and framing and generated a parsimonious model for the impact of media imagery. Viewing (or in this case, playing) media creates objects in what Price and Tewksbury term the 'knowledge store', which they describe as 'a network of constructs, including information about social objects and their attributes' (1997: 186). The frequency with which social objects will be recalled and used depends in large part on chronic accessibility. At the simplest level, constructs are accessible when they are reinforced repeatedly and recently. Thus, imagery that is viewed or played repeatedly is more accessible when a person is attempting to recall information about that class of social objects. This is consistent with Shrum's (2002) approach to cultivation, i.e. that a set of ideas about the real world are in large part based on the accessibility of constructs, which in turn are influenced by how often those constructs are viewed in media. In other words, social objects, like types of people, can be viewed or played in media and this action makes them more likely to be recalled later if they were more prevalent.

Theoretically, a media environment in which a particular type of person is highly represented will result in a viewer or player who is more likely to recall that type of person rather than a different type of person. The outcomes of such a system are very similar to the outcomes suggested by traditional cultivation, even while the causal mechanisms differ. Recently, work by Mastro and colleagues (2007) has made this connection with the mental models approach for the cultivation of Latinos on television. This work

reveals that a medium's general depiction of a group does have an impact on its users' perceptions of that group, albeit moderated by their real–world experiences. If such a consistent pattern of representation on television can have effects as Mastro et al. (2007) show, a consistent pattern in other media may do as well. This is especially relevant as games begin to displace prior media as the dominant symbol sets for many Americans. For gaming, groups repeatedly seen or seen in particular roles, will begin to be more accessible to the viewer or player. In keeping with prior video game content analyses as well as the Harwood and Anderson television work, the key group variables here are gender, race and age.

This is also relevant to the populations themselves, as representation can have identity and self-esteem effects on individuals from those groups (Comstock and Cobbey, 1979; McDermott and Greenberg, 1984). Tajfel's social identity theory (1978) suggests that groups look for representations of themselves and then compare those representations with those of other groups. The presence of the group – including within games (Royse et al., 2007) – serves as a marker for members to know that they carry weight in society. Conversely, the absence of portrayals should lead to a feeling of relative unimportance and powerlessness (Mastro and Behm–Morawitz, 2005). These effects may be more or less likely if those populations play games at higher or lower rates. Thus, population figures can be used as an expected value baseline for comparison with the actual numbers of characters. In addition, real–world demographic player data can suggest which groups might be accessing games at higher rates than others.

Gender is the first example: the US population is 50.9 percent female and 49.1 percent male (Smith and Spraggins, 2001). In contrast, 60 percent of gamers are male and 40 percent are female across all age brackets (Entertainment Software Association, 2009).

For race, the US population is increasingly multicultural, but still dominated by one major racial group. In the 2000 census, the ethnic breakdown was 75.1 percent white, 12.5 percent Hispanic/Latino, 12.3 percent African American, 4 percent Asian/Pacific Islander, 0.9 percent Native American and 2.4 percent biracial (Grieco and Cassidy, 2001).[1] African American and Hispanic youths play games at the highest rates (Rideout et al., 2005).

For age, children (under 13) are 21.41 percent of the population, teens (13–19) are 7.18 percent, adults (20 to 64) are 58.97 percent and the elderly (65+) are 12.43 percent. Players skew slightly younger than the general population, but not as much as they did 10 years ago (Williams, 2006a). According to industry statistics, the average game player is currently 35 years old (Entertainment Software Association, 2009).

RESEARCH QUESTIONS

The appearance of groups in the media environment can be contrasted with the appearance of that group in daily life. Exposure to different groups varies greatly according to socioeconomic status and geography, but simple census–level data provide a national comparative baseline. If a group appears more in a game than in real life, it is over-represented, which could make such groups more accessible to players recalling group-based schemas. Additionally, Price and Tewksbury's (1997) model implies a converse possibility: social objects which are not chronically accessible will be recalled less often. Demographic groups of people who are not represented are slowly rendered invisible by virtue of their relative inaccessibility in the knowledge store. For example, mainstream television shows rarely portray actors or characters from the Hmong ethnic group of South–East Asia. If asked to list ethnic groups, theoretically the people who viewed mainstream television would be less likely to name the Hmong because the social object 'Hmong' would be relatively inaccessible compared to the easily accessible objects 'white', 'black', etc. (that is, except for those viewers whose personal experiences, knowledge or location made the 'Hmong' social object accessible; Mastro et al., 2007). Hmong would be unlikely to be the object of recall for both positive associations (e.g. groups that are seen as trustworthy, hardworking or make good leaders) and negative associations (e.g. groups that are seen as untrustworthy, lazy or make bad leaders). In this sense, groups who appear relatively less often in media may be rendered invisible to the viewer. Therefore, with the data on users and the general population in place as a comparative baseline, we can formulate the following research question:

> RQ1: How frequently are different gender, race and age groups represented in games?

Just as in other media, all characters are not equally important to the story or action. Some appear briefly, while others remain on the screen for the entire game session. To simplify for the very broad sample used here, the most basic level of role is used: those who are played by the player and those that are part of the game world. Played characters, called 'primary' characters in the analysis, drive the action in games. Non–player characters are the 'secondary' characters in the analysis. They may drive some of the action, for example when they talk to or interact with the player, and some may feature strong artificial intelligence, but ultimately their role is secondary, subservient or simply as targets. For the mental models or accessibility and recall approaches, the groups which typically appear in secondary roles may be associated with having secondary roles in daily life. Within the game session, there is evidence to suggest that players regard secondary characters very differently than primary ones. Magnetic Resonance Imaging (MRI)–based research shows

that players react far differently to other characters when they think they are controlled by the computer rather than another real person (Reeves, 2007). Therefore, it is important to establish who appears in primary and secondary roles. Once these baseline measures are established, future research can use them to test their impact rigorously in causal models. Therefore:

> RQ2: Is there a difference between groups' appearance in primary and secondary roles?

As noted in the introduction, there has been a practical gap in research between the games which have been studied and the games actually played by the public. This is an issue of external validity; if research has occurred on media that are not used, it is of less practical import. Investigating the possible difference between all games and the games most played allows for two kinds of analysis. First, it creates a test of prior content analytic work, testing to see if those projects have been overestimating or underestimating outcomes in proportion to what games are actually played. Second, it creates a test with implications for player research. If there is a systematic difference between content in the games that are most played and the games that are made, it offers insight into player tastes and preferences. In past work, developers have been imputed as the agents driving content, but examining popularity places some of the causal agency in the hands of the player. If players systematically choose games with particular forms of social identity, developers are economically incentivized to create those games, creating a cycle of social identity creation and perpetuation. Therefore:

> RQ3: Is there a difference in character representation between the typical game made and the most popular games?

As noted earlier, ratings systems have been used in previous content analyses. The present study seeks to replicate and extend those findings with a larger sample which can be generalized:

> RQ4: Is there a difference in characters' social group representation between games with different ESRB ratings?

METHOD
A large–scale content analysis of video game characters was undertaken to address the research questions.

Sample
Sales data were obtained from the research firm The NPD Group for a calendar year from March 2005 to February 2006 for the nine major game systems sold in the USA during that timespan: Xbox 360, Xbox,

PlayStation 2 (PS2), PlayStation (PS), Nintendo Gamecube, PlayStation Portable (PSP), Nintendo Gameboy Advance (GBA), Nintendo Dual Screen (DS) and PC. Of these, the PSP, GBA and DS are portable systems, PCs are standard computer systems and the rest are TV-based console systems. The sampling frame included the top 150 games across all platforms, with a minimum of 15 titles per system. On multiple systems 17 games were available, leaving 133 games to be tested. In these duplicate system cases, the system with the most advanced graphical processor was used.[2] These 133 titles constituted a highly representative frame for the universe of games as made by developers by accounting for more than 95 percent of all game sales within the sampling period.

In order to determine the relative use and popularity of the games, sales figures were used as weights during the analysis. Each game was weighted by the number of copies that it sold, meaning that a game selling 4 million copies figured twice as heavily in the computations as a game selling 2 million copies. In this way, the sample could be weighted to represent not only the games made by developers, but also the games that were actually purchased and played. Giving each title equal weight would inaccurately represent the time spent on content across the country, and invalidate the theoretical approaches outlined earlier that depend on the volume of impressions as consumed, not as produced.

Each game was played by an expert game player – who was not one of the coders – for 30 minutes on the default difficulty setting, typically 'low' or 'easy'. These 30-minute segments were recorded digitally and stored on a high-end desktop computer for coding. The 30-minute segments were the largest unit of analysis and were used for the research questions that called on game-level data. Measures of representation as portrayed on games for a given system required a game-level number for computation. Similarly, portrayals by ratings code required a rating-level value of different titles. In these cases, the weighted data were used to create collapsed mean values for games rather than including all characters as equal units of analysis. Such an approach would have skewed the results, given that some games feature more characters than others.

Data analysis

The bulk of the analysis concerned characters within games. Every character in the sample was recorded as an individual unit and coded, for a total of 8572 characters. Each character was coded for their status as either a primary (player-controlled) or secondary (computer-controlled) figure prior to coding. Not all games featured visible primary characters. In the event that a game provided the option to choose from a list of primary characters, such characters were selected randomly so that characters from both genders, all

ethnicities and all age groups had an equal chance of being selected. Then, all characters were coded on the variables of gender, race and age. Because prior work has shown that viewers heavily discount social objects that they cannot confirm as human (Hoffner and Cantor, 1991; Reeves and Nass, 1996), in addition the characters were coded as either human or one of several non–human or quasi–human options. Only the human characters were retained for analysis ($N = 4966$).

To gain comparative baseline data from the actual US population, US census data were collected from the government census website (www. census.gov) and from special reports derived from the 2000 census (Grieco and Cassidy, 2001; Hetzel and Smith, 2001; Smith and Spraggins, 2001).

Two research assistants coded all of the 30-minute recordings in the sample. Prior to coding, all of the assistants participated in 10 hours of training sessions. Coding rates were assessed at regular intervals throughout the study. Cohen's kappas (Cohen, 1960) were .93 for gender, .92 for age, .89 for race and .97 for human/non–human.

RESULTS

The first question addressed what groups appear in games and which are over-represented and under-represented in relation to the actual population. These were addressed with respect to gender, race and age. Figure 1 shows the weighted data for gender representation across the universe of games. Figure 1 also addresses RQ2, which was concerned with the role (primary or secondary) played by the character.

As the figure shows, male characters are vastly more likely to appear than female character in general. The overall difference of 85.23/14.77 percent is also a large contrast with the 50.9/49.1 percent distribution in the actual population. This difference is heightened among the primary 'doer' characters, where males are even more likely to appear. As a general rule then, males appear more frequently in games than females, and even more so as drivers of the action. When females do appear, they are more likely to be in secondary roles than primary ones.
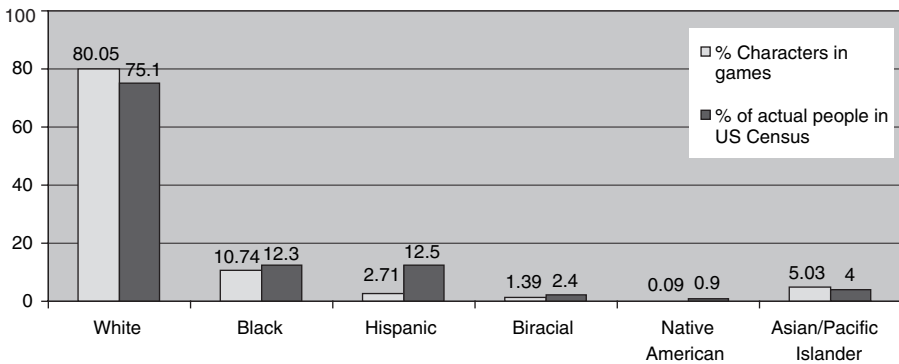
Figure 2 shows the weighted data for race representation across the universe of games, with comparison bars from the US population.

Contrasts between the proportion of characters appearing in games and those appearing in the general population can be observed for every race. Whites and Asians are over-represented and all other groups are under-represented. In proportional figures relative to their actual population, whites are 6.59 percent and Asians are 25.75 percent over-represented. All others are under–represented: blacks by 12.68 percent, Hispanics by 78.32 percent, biracials by 42.08 percent and Native Americans by 90 percent. When primary roles are considered, all groups appear less often except for whites,
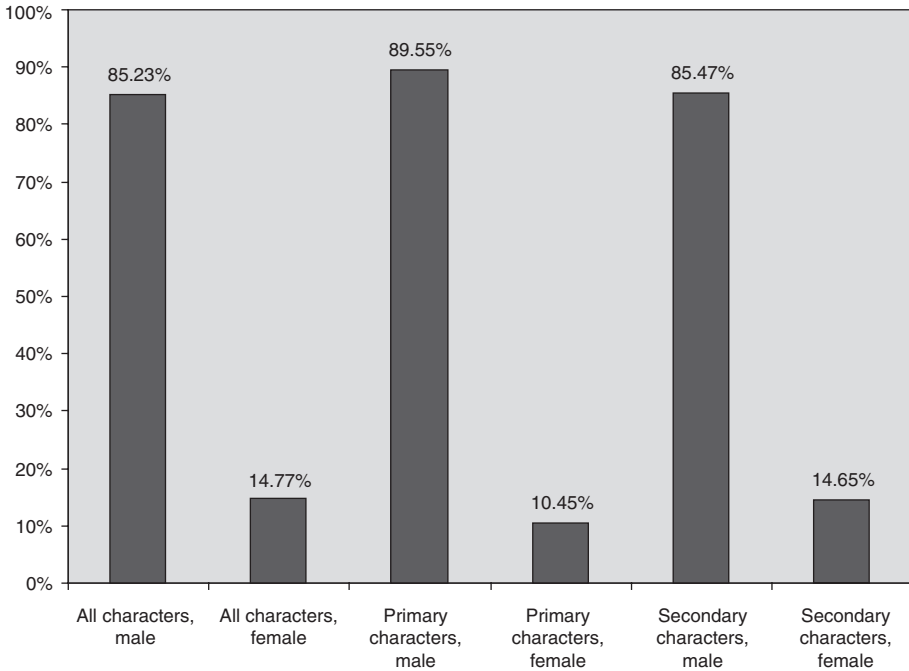
t-tests for male vs. female: for all characters, *t* = 224.31, *df* = 4343, *p*<.001; for primary characters, *t* = 22.05, *df* = 96, *p*<.001; for secondary characters, *t* = 175.331, *df* = 4231, *p*<.001

• Figure 1   Gender breakdowns by role

• Figure 2   Over-representation and under-representation by race

who appear more often than overall. White characters account for 84.95 percent of all primary characters, black 9.67 percent, biracial 3.69 percent and Asian 1.69 percent. Hispanics and Native Americans did not appear as a primary character in any game, they existed solely as secondary characters.

Figure 3 shows the weighted data for age representation across the universe of games, with comparison bars from the US population.

825
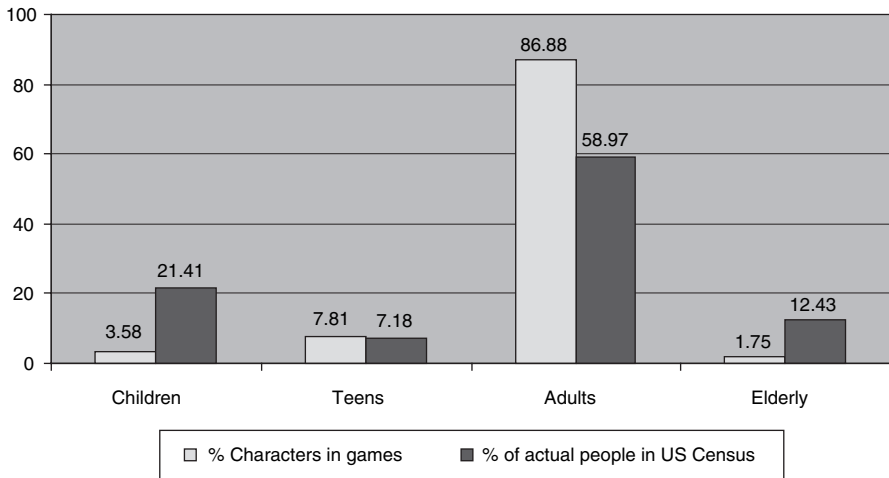
• Figure 3   Over-representation and under-representation by age

Contrasts between the proportion of characters appearing in games and those appearing in the general population can be observed by age bracket. Again, some groups differ systematically. Almost all of the differences can be explained by the over–representation of adult characters, who appear at a rate in games 47.33 percent higher than their prevalence in the actual population. Teens appear at a rate very similar to the population, but children and the elderly appear at substantially lower rates.

When primary and secondary roles are considered, these patterns change. Among primary characters, children rise to 8.9 percent, teens rise to 13.3 percent, adults drop to 76.5 percent and the elderly drop to 1.34 percent. These numbers remain substantially different from the actual population, but with the exception of the elderly, all ages move towards population values among primary characters. It is the great bulk of secondary characters that makes the universe of games appear less age representative, but regardless of the role, adults appear the most often and at the expense of children and the elderly.

The next research question asked whether representations varied by ESRB rating. There were no games rated AO, leaving the ESRB categories E = Everyone, E10 (everyone 10 or older), T = Teen and M = Mature (18+) (see Table 1).

The gender differences between ratings were statistically significant due to the large number of characters, but were substantively negligible. In contrast, there was a substantive race difference for the E-rated games, which featured far fewer white characters and included far more black and biracial characters. This was due almost entirely to the presence of very popular sports titles,

• Table 1   Representations by ESRB rating

GENDER AS A PERCENTAGE OF ALL CHARACTERS	EVERYONE	E10	TEEN	MATURE
Percentage of all characters, male	86.09	79.10	85.03	86.55
Percentage of all characters, female	13.91	20.90	14.97	13.45
Race as a percentage of all characters:				
White	59.53	75.07	79.60	72.39
Black	32.64	18.78	9.09	11.68
Hispanic	1.63	1.61	1.70	7.59
Biracial	5.30	0.53	3.14	0.15
Native American	0.00	0.00	0.11	0.71
Asian/Pacific Islander	0.89	4.00	6.36	7.47
Age as a percentage of all characters:				
Child	5.14	16.61	2.14	0.41
Teen	10.76	0.00	3.66	1.62
Adult	81.74	83.06	93.27	96.15
Elderly	2.37	0.33	0.93	1.82

Total game $N = 68$ for Everyone, 8 for E10, 38 for Teen and 19 for Mature. 95% confidence interval tests performed at the character level show that none of the values in the table overlap with another column (with the exception of the 0 values for Native Americans and the E and E10 values for Hispanics).

which contained a much higher proportion of black and biracial characters – all actual athletes drawn directly from real-world sporting leagues. Age differences showed the E and E10 rated titles having more child characters than the older titles, with a decrease in the ranks of adult characters.

RQ4 involved the possible difference in content between the games that were made and those that were most popular. Using the unweighted data, an examination of game representations by gender shows that the split is 81.24/18.76 percent male/female. This is in contrast to the weighted data, which showed an 85.23/14.77 split.[3] In other words, gamemakers created games that heavily featured male characters, but the games that were actually purchased were even more heavily male. This finding varied by ratings, with smaller gender imbalances between games made and games bought occurring as the ratings increased (E-10 rated games were omitted because there were too few in the sample year for substantive analysis): E-rated games featured 79.03 percent male characters as made, but 86.19 percent as purchased (7.16 percent difference); T-rated games featured 81.69 percent male characters as made, but 84.71 percent as purchased (3.02 percent difference); M-rated games featured 88.07 percent male characters as made, but 86.99 percent as purchased (-1.08 percent difference). In other words, games for young or general audiences had more gender equity, but the games actually purchased by players (and assumedly by parents) largely negated it.

DISCUSSION

The goal of the study was to determine whether there were systematic patterns of representation for videogame characters for gender, race and age across all games and within ratings categories. The large sample – nearly the universe of titles for a year – and the weighting approach allow for the first statements which can be generalized about the characters contained in popular games. The results show that males, whites and adults are over-represented in comparison to the actual US population. These over-representations come at the expense of women, some minority groups – chiefly Latinos and Native Americans – and children and the elderly. The most popular games are less representative than the typical game produced by developers, indicating that players also play a role in the cycle of creation and consumption. In addition, the mere difference in numbers between the weighted and unweighted data suggests a likelihood that prior content analyses have been systematically off if the goal was to analyze the games that the public actually plays. These several outcomes have implications for theories of game effects and identity and for comparative analyses of media.

The differences between the game world, the player base and the US population have implications for both self-identities and considerations of other groups. The Latino case provides one strong example from the data. Although there are no race data for older game players, we do know that Latino children play more video games than white children, so it is conceivable that they play more as teenagers and adults. Nevertheless, Latinos are unlikely to see representations of their ethnic group among game characters and never as primary characters. According to social identity theory (Tajfel, 1978), this lack of appearance is a direct signal to Latinos that they are relatively unimportant and powerless compared to more heavily present groups. In addition, perceptions about Latinos may change for members of other groups. The data here show that – unlike television, which shows modest gains for Latinos (Mastro and Behm-Morawitz, 2005) – members of these two ethnic groups continue to be most present in secondary roles. The same issues are present for blacks, Native Americans, females and children, all of whom are under-represented compared to the population and among game players. The data on elderly players are not reliable, but given that the average age of game players continues to rise steadily, it is a safe assumption that more older players are picking up the control pad over time. When they do, they will find a game universe which does not feature them. For children, the stakes may be slightly higher than for television. Many have suggested that games function as crucial gatekeepers to interest in technology, which translates into education and careers in mathematics and science-related fields (Lin and Lepper, 1987; Williams, 2006a). If Latinos or any other groups become disenchanted with games due to poor representation,

subsequently they may have less interest in technology and its opportunities for class advancement. Ironically, they would be less likely to become gamemakers themselves, helping to perpetuate the cycle.

For those outside the groups, representations have been shown consistently to lead to effects (Dixon and Linz, 2000; Pan and Kosicki, 1996), although the valence of the effects depends on the context on the portrayals (Bodenhausen et al., 1995) and the viewers' other real-world exposure to that group in relation to the amount of media consumed (Mastro et al., 2007). In the case of the game data, the under-represented groups were both scarce and, when present, featured in secondary roles. The Price and Tewksbury (1997) 'knowledge store' and Shrum's (2002) cultivation recall model both suggest that this makes those groups seem less visible, while social identity theory additionally suggests that they will be seen as less important. Of course, a content analysis cannot show a link between images and resulting opinion or behavioral change. However, it is a necessary condition which has now been demonstrated and can be tested for causality.

Comparing games with prior media and tracking the causes

The findings have a striking similarity to those typically found in content analyses of television, suggesting that whatever causal forces are at work stretch across media and are not unique to games. With regard to race, the findings here nearly mirror Harwood and Anderson's television data, which found 82.9 percent white, 2.6 percent Latino, 11.4 percent black and 2.6 percent Asian characters. It follows that Latinos and Native Americans can be said to be systematically under-represented across media. This is a particularly surprising finding for Latinos, who are a growing minority ethnic group off-screen. According to the Census Bureau (Grieco and Cassidy, 2001), Latinos are projected to double in population over the next 20 years. Any theoretical effects arising from their invisibility in media could be expected to worsen unless the basic causes of under-representation change. The age outcome disparity is nearly identical to what Harwood and Anderson found for television, i.e. that adults are over-represented at the expense of children and the elderly. The only exception to the TV rule is with gender, where the disparity is much larger for games than for television.

There are several possible explanations as to why game characters are disproportionate to the general population. First, the most obvious is that game titles are driven by consumer demand, an explanation that finds support in the previous analysis; the most popular games were the ones with larger imbalances. Williams (2006a) has suggested that games and gender work as a cycle: games feature more males and so attract more young males to play. Those males grow up and are more likely to become gamemakers than women, perpetuating the role of males in game creation, etc. What is new in

these data is the recognition that the patterns could be far more imbalanced if they were driven entirely by sales. After all, if the most popular games contain more males than the average game does, there is an economic incentive for developers to make games with 85 percent males, not 81 percent; if developers were aware of these data and were interested in matching their clientele with representations accurately, they would add slightly more males to game titles in general – that is, if their goal was only to satisfy the current audience rather than to expand it.

Second, ethnolinguistic vitality theory offers a further explanation for the disparity: that games are merely a mirror for underlying systematic social inequalities. However, this line of thinking ignores the process by which media are created by individuals and groups with their own identities and influences. Third, creators of media simply make media that reflect their own identity. Because the game industry studies diversity within its ranks, this ecological hypothesis can receive some basic investigation. With the exception of African Americans, the representation in games bears a strong resemblance to the game developer workforce itself. Game developers are 88.5 percent male and 11.5 percent female (Gourdin, 2005). They are 31 years old on average, with nearly all between the ages of 20 and 40. Ethnically, they are 83.3 percent white, 7.5 percent Asian, 2.5 percent Hispanic and 2.0 percent black. This would explain the low numbers of Hispanic game characters, who have few internal advocates, but not the black characters. Blacks are far more represented in games than their numbers among developer ranks would suggest. However, this may be simply an artifact of the prevalence of black characters in sport-based games, a pattern which is reflected in the larger culture of professional sports. Indeed, outside of sports games, the representation of African Americans drops precipitously, with many of the remaining featured as gangsters and street people in *Grand Theft Auto* and *50 Cent Bulletproof*. Not coincidentally, the popularity of sports games is also a primary driver of the maleness of characters, since the sports titles are all game versions of real-world men's sports leagues (*Madden Football*, *NCAA Football*, *NBA Live*, *MVP Baseball*, *WWE Wrestling*), with no games derived from women's sports leagues. The developer demographic explanation is made stronger by considering that the number of female characters (15%) comes much closer to the number of female gamemakers (11.5%) than the number of female players (38%). If the process were entirely player-driven, there would be far more female characters, especially among the primary, playable characters (where the proportion is nearly identical to the developers' ranks). Research on gender among game developers would shed light on this process.

Still, the most likely cause for the representation patterns in this study is a combination of developer demographics and perceived ideas about game

players among marketers. The stereotype of game players as only young, white males who want to be powerful white adults may be driving the content-creation process, even as the player base becomes older and more diverse. From a business and marketing viewpoint, game developers would be missing substantial opportunities for making games for different audiences. Women, at 38 percent of game players but only 15 percent of characters, are the most underserved. Latinos, who play more per day than whites and form 12.5 percent of the population, are only 2 percent of characters. These two groups would be profitable existing groups to serve better. This rests on the assumption that those underserved groups do in fact want to play characters like themselves. It is possible that those relatively marginalized populations would prefer to play the more empowered groups: women preferring to play men, blacks preferring to play whites, etc. However, research on female gamers generally has found the opposite: that female players in educational (DeJean et al., 1999), online (Taylor, 2006) and more general game contexts seek out feminine, sexy and strong characters to play (Royse et al., 2007).

CONCLUSION

As with any project, there are limitations in the current work. Most notably, the representations here do not give context beyond primary and secondary roles, merely appearance. Groups systematically shown in negative contexts in media can become stereotyped by other groups (Dixon and Linz, 2000). Future research could explore this in more depth, with the appearance of blacks as either athletes or gangsters as an obvious starting point. The measure of exposure (units sold) is also a proxy rather than a perfect time log-based one. However, there is no such system available, making this the best available option. Lastly, while this study examined the largest sample of top-selling games to date, we have only one massively multiplayer online role-playing game (MMORPG) within the year sampling frame: *World of Warcraft*. MMORPGs are becoming increasingly popular and regularly give players multiple options for choosing gender, race, age and other increasingly fine-grained identity markers. This highlights the fact that if the level of player choice continues to grow, researchers should pay special attention to the players themselves as important sources of character variation. Nevertheless, the current study demonstrates that the world of game characters is highly unrepresentative of the actual population and even of game players. For developers, this is a missed opportunity. For players, it is a potential source of identity-based problems.

Notes

1   Note that the Hispanic/Latino classification system used by the census allows for overlapping entries among Latinos who self-classify as Latino in one scale and another race elsewhere, making the total not a perfect 100 percent.
2   An exception was made for Nintendo DS games, which were substantively different due to the unique dual screen. The DS games were all measured, whether duplicated or not. The master game list is available from the first author.
3   When measured at the character level, the 99 percent confidence intervals for the weighted and unweighted values do not overlap for any of the measures in this paragraph.

References

Bandura, A. (1994) 'The Social Cognitive Theory of Mass Communication', in J. Bryant and D. Zillmann (eds) *Media Effects: Advances in Theory and Research*, pp. 61–90. Hillsdale, NJ: Erlbaum.

Bodenhausen, G., N. Schwarz, H. Bless and M. Wänke (1995) 'Effects of Atypical Exemplars on Racial Beliefs: Enlightened Racism or Generalized Appraisals?', *Journal of Experimental Social Psychology* 31(1): 48–63.

Braun, C.M. and J. Giroux (1989) 'Arcade Videogames: Proxemic, Cognitive and Content Analyses', *Journal of Leisure Research* 21(2): 92–105.

Cohen, J. (1960) 'A Coefficient of Agreement for Nominal Scales', *Educational and Psychological Measurement* 20(1): 37–46.

Comstock, G. and R. Cobbey (1979) 'Television and the Children of Ethnic Minorities', *Journal of Communication* 29(1): 104–15.

DeJean, J., R. Upitis, C. Koch and J. Young (1999) 'The Story of Phoenix Quest: How Girls Respond to a Prototype Language and Mathematics Computer Game', *Gender and Education* 11(2): 207–23.

Dietz, T.L. (1998) 'An Examination of Violence and Gender Role Portrayals in Videogames: Implications for Gender Socialization and Aggressive Behavior', *Sex Roles* 38(5–6): 425–42.

Dill, K., D. Gentile, W. Richter and J. Dill (2005) 'Violence, Sex, Race and Age in Popular Videogames', in E. Cole and J. Daniel (eds) *Featuring Females: Feminist Analysis of the Media*, pp. 115–30. Washington, DC: American Psychological Association.

Dixon, T. and D. Linz (2000) 'Overrepresentation and Underrepresentation of African Americans and Latinos as Lawbreakers on Television News', *Journal of Communication* 50(2): 131–54.

Downs, E.P. and S. Smith (2005) 'Keeping Abreast of Hypersexuality: A Videogame Character Content Analysis', paper presented at the International Communication Association, New York, 26–30 May.

Entertainment Software Association (2009) 'Video Games and Families', URL (consulted 5 April 2009): http://www.theesa.com/gamesindailylife/families.asp

Gerbner, G., L. Gross, M. Morgan and N. Signorielli (1994) 'Growing Up with Television: The Cultivation Perspective', in J. Bryant and D. Zillman (eds) *Media Effects: Advances in Theory and Research*, pp. 17–41. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gourdin, A. (2005) *Game Developer Demographics: An Exploration of Workforce Diversity*. Mt Royal, NJ: International Game Developers Association.

Grieco, E. and R. Cassidy (2001) *Overview of Race and Hispanic Origin: Census 2000 Brief*. Washington, DC: US Department of Commerce.

Harwood, J. and K. Anderson (2002) 'The Presence and Portrayal of Social Groups on Prime-time Television', *Communication Reports* 15(2): 81–97.

Heintz-Knowles, K., J. Henderson, C. Glaubke, P. Miller, M.A. Parker and E. Espejo (2001) *Fair Play? Violence, Gender and Race in Videogames*. Oakland, CA: Children Now.

Hetzel, L. and A. Smith (2001) *The 65 Years and Over Population: 2000: Census 2000 Brief*. Washington, DC: US Department of Commerce.

Hirsch, P. (1981) 'Distinguishing Good Speculation from Bad Theory: Rejoinder to Gerbner et al.', *Communication Research* 8(1): 73–95.

Hoffner, C. and J. Cantor (1991) 'Perceiving and Responding to Mass Media Characters', in J. Bryant and D. Zillman (eds) *Responding to the Screen: Reception and Reaction Processes*, pp. 63–101. Hillsdale, NJ: Lawrence Erlbaum Associates.

Janz, J. and R. Martis (2007) 'The Lara Phenomenon: Powerful Female Characters in Videogames', *Sex Roles* 56(3–4): 141–8.

Lenhart, A., S. Jones and A.R. Macgill (2008) 'Video Games: Adults Are Players Too, *Pew Internet & American Life Project*, URL (consulted 5 April 2009): http://pewresearch.org/pubs/1048/video-games-adults-are-players-too

Lin, S. and M.R. Lepper (1987) 'Correlates of Children's Usage of Videogames and Computers', *Journal of Applied Social Psychology* 17(1): 72–93.

McDermott, S. and B. Greenberg (1984) 'Parents, Peers and Television as Determinants of Black Children's Esteem', in R. Bostrom (ed.) *Communication Yearbook Vol. 8*, pp. 164–77. Beverly Hills, CA: Sage.

Mastro, D. and E. Behm-Morawitz (2005) 'Latino Representation on Primetime Television', *Journalism and Mass Communication Quarterly* 82(1): 110–30.

Mastro, D., E. Behm-Morawitz and M. Ortiz (2007) 'The Cultivation of Social Perceptions of Latinos: A Mental Models Approach', *Media Psychology* 9(2): 347–65.

Pan, Z. and G. Kosicki (1996) 'Assessing News Media Influences on the Formation of Whites' Racial Policy Preferences', *Communication Research* 23(2): 147–78.

Potter, W.J. (1994) 'Cultivation Theory and Research: A Methodological Critique', *Journalism Monographs* 147: 1–17.

Price, V. and D. Tewksbury (1997) 'News Values and Public Opinion: A Theoretical Account of Media Priming and Framing', in G. Barnett and F. Boster (eds) *Progress in the Communication Sciences Vol. 13*, pp. 173–212. New York: Ablex.

Reeves, B. (2007) 'Doing Research with Avatars: Current Trends, Predictions and Warnings', paper presented at the International Communication Association, San Francisco, CA, 24–28 May.

Reeves, B. and C. Nass (1996) *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. New York: Cambridge University Press/CSLI.

Rideout, V., D. Roberts and U. Foehr (2005) *Generation M: Media in the Lives of 8–18 Year-olds*. Washington, DC: Kaiser Family Foundation.

Royse, P., J. Lee, B. Undrahbuyan, M. Hopson and M. Consalvo (2007) 'Women and Games: Technologies of the Gendered Self', *New Media & Society* 9(4): 555–76.

Schierbeck, L. and B. Carstens (2000) 'Violent Elements in Computer Games: An Analysis of Games Published in Denmark', in C. Vonfeiletzen and U. Carlsson (eds) *Children in the Media Landscape: Games, Pornography, Perceptions*, pp. 127–31. Nordicom: UNESCO International Clearinghouse on Children and Violence on the Screen.

Sherry, J.L., Lucas, K., B.S. Greenberg and K. Lachlan (2006) 'Videogame Uses and Gratifications as Predictors of Use and Game Preference', in P. Vorderer and J. Bryant (eds) *Playing Computer Games: Motives, Responses and Consequences*, pp. 213–24. Mahwah, NJ: Erlbaum.

Shibuya, A., A. Sakamoto, R. Shiratori, K. Arai and F. Kato (2004) 'The Quantity and Context of Videogame Violence in Japan: Toward Creating an Ethical Standard', in R. Shiratori (ed.) *Gaming, Simulations and Society*, pp. 111–20. Tokyo: Springer-Verlag.

Shrum, L.J. (1999) 'Television and Persuasion: Effects of the Programs between the Ads', *Psychology and Marketing* 16(2): 1119–40.

Shrum, L.J. (2002) 'Media Consumption and Perceptions of Social Reality: Effects and Underlying Processes', in J. Bryant and D. Zillmann (eds) *Media Effects: Advances in Theory and Research*, pp. 69–95. Mahwah, NJ: Lawrence Erlbaum Associates.

Slagle, M. (2006) 'Poll: 4 in 10 Americans Play Videogames', *Washington Post*, 8 May, URL (consulted 20 June 2007): http://www.washingtonpost.com/wp-dyn/content/article/2006/05/07/ar2006050700172.html

Smith, D. and R. Spraggins (2001) *Gender: 2000: Census 2000 Brief*. Washington, DC: US Department of Commerce.

Smith, S., K. Lachlan and R. Tamborini (2003) 'Popular Videogames: Quantifying the Presentation of Violence and its Context', *Journal of Broadcasting and Electronic Media* 47(1): 58–76.

Tajfel, H. (1978) *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*. London: Academic Press.

Taylor, T. L. (2006) *Play between Worlds: Exploring Online Game Culture*. Cambridge, MA: MIT Press.

Thompson, K. and K. Haninger (2001) 'Violence in E-rated Videogames', *Journal of the American Medical Association* 286(5): 591–8.

Thompson, K., K. Haninger and F. Yokota (2001) 'Violence in Children's Films and Videogames: Reply', *Journal of the American Medical Association* 286(16): 1971–2.

Thompson, K., K. Tepichin and K. Haninger (2006) 'Content and Ratings of Mature-rated Videogames', *Archives of Pediatrics and Adolescent Medicine* 160(4): 402–10.

Williams, D. (2006a) 'A (Brief) Social History of Gaming', in P. Vorderer and J. Bryant (eds) *Videogames: Motivations and Consequences of Use*. Mahwah, NJ: Erlbaum.

Williams, D. (2006b) 'Virtual Cultivation: Online Worlds, Offline Perceptions', *Journal of Communication* 56(1): 69–87.

DMITRI WILLIAMS is an assistant professor at the Annenberg School for Communication, University of Southern California. His research interests include community and networked games.
Address: Annenberg School for Communication, University of Southern California, 734 W. Adams Boulevard, Los Angeles, CA 90089, USA. [email: dcwillia@usc.edu]

NICOLE MARTINS is an assistant professor in the Department of Telecommunications, Indiana University. Her research interests include the social and psychological effects of the mass media on youth.

MIA CONSALVO is an associate professor in the School of Media Arts and Studies, Ohio University. Her research interests include the influence of Japan on the digital games industry, women and games, and fashion in virtual worlds.

JAMES D. IVORY is an assistant professor in the Department of Communication, Virginia Polytechnic Institute and State University. His research interests include the content and physiological, psychological and social effects of new communication technologies, particularly videogames and virtual environments.