

Springer Proceedings in Complexity

Muhammad Aurangzeb Ahmad  
Cuihua Shen  
Jaideep Srivastava  
Noshir Contractor *Editors*

---

# Predicting Real World Behaviors from Virtual World Data

 Springer

# **Springer Proceedings in Complexity**

For further volumes:

<http://www.springer.com/series/11637>

Muhammad Aurangzeb Ahmad • Cuihua Shen  
Jaideep Srivastava • Noshir Contractor  
Editors

# Predicting Real World Behaviors from Virtual World Data

 Springer

*Editors*

Muhammad Aurangzeb Ahmad  
Dept. Computer Science and Engineering  
University of Minnesota  
Minneapolis  
Minnesota  
USA

Jaideep Srivastava  
Department of Computer Science  
University of Minnesota  
Minneapolis  
Minnesota  
USA

Cuihua Shen  
Emerging Media & Communication  
Program School of Arts & Humanities  
University of Texas at Dallas  
Richardson  
Texas  
USA

Noshir Contractor  
Dept. of Communication Studies  
Northwestern University  
Evanston  
Illinois  
USA

ISSN 2213-8684

ISSN 2213-8692 (electronic)

ISBN 978-3-319-07141-1

ISBN 978-3-319-07142-8 (eBook)

DOI 10.1007/978-3-319-07142-8

Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014940069

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

*To my father, Mushtaq Ahmad Mirza, you  
will always be missed; to my mother, Khalida  
Parveen, may you always be with us*

–Muhammad Aurangzeb Ahmad

*To my daughter Amelia Lingyun*

–Cindy Shen

# Preface

In September 2013, I was honored to be the keynote speaker at a workshop at the ASE/IEEE International Conference on Social Computing. The workshop was titled “Predicting Real World Behaviors from Virtual World Data.” This book is an outcome of that workshop.

The realm of virtual worlds (VW) and massively multiplayer online games (MMOG) is a fascinating one for social scientists. This online activity has seen extensive growth in the last decade, both in the number of VWs and MMOGs on the market, and the number of players and participants worldwide. The worldwide popularity of this online medium has increased the opportunities for communicating and socializing between individuals and groups.

Researchers showed early enthusiasm for the study of VWs and MMOGs. Scientists postulated that much could be learned about offline behavior by studying behavior online, assuming that the behavior would typically “map” from one environment to the other. The online communities provided a rich natural laboratory where all actions could be captured via the digital footprints left by the participants.

In one example, researchers studied the spread of the “Corrupted Blood” virus in the popular MMOG *World of Warcraft*. The virus was designed by the game developers to spread from one player character to the other, and soon erupted into a full-scale epidemic, with thousands of characters infected in the game.<sup>1</sup> The researchers hypothesized that we could learn much about the spread of viruses in the real world by studying the patterns of infection in virtual environments.

Of course, this “mapping principle”<sup>2</sup> may not apply in all situations. The players in the Corrupted Blood incident may have been trying to deliberately infect one another, just for the fun of it. Then there are experts who postulate that we should not assume too much about the real identity of the individuals that we meet online. People may be exploring alternative personas and ways of behaving in this relatively anonymous setting.<sup>3</sup> The attractive, youthful female avatar in the VW *Second Life*

---

<sup>1</sup> Lofgren, E. T., & Fefferman, N. H. (2007). The untapped potential of virtual game worlds to shed light on real world epidemics. *The Lancet Infectious Diseases*, 7, 625–629.

<sup>2</sup> Williams, D. (2010). The mapping principle, and a research framework for virtual worlds. *Communication Theory*, 20, 451–470.

<sup>3</sup> Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet*. Simon and Schuster.

may be controlled by an older male player. After all, “On the Internet, nobody knows you’re a dog”.<sup>4</sup>

This view of online behavior suggests that players act out fantasies, deceptions, and exploration of alternate identities. However, maintaining the consistency of this façade requires a large mental effort on the part of the player. When caught up in the thrill of a dungeon raid in *World of Warcraft*, the heat of battle in *Eve Online*, or the pleasure of conversation and music in *Second Life*, it is difficult to maintain a false persona while also concentrating on the task (or play!) at hand. This is simply too much cognitive load for most of us.

As a program manager at IARPA, I had the opportunity to test this in a research program called Reynard. The premise of the program was that our real world (RW) characteristics strongly influence our behaviors, both offline and online, and so those RW characteristics and ways of behaving “bleed over” into VW and MMOG behaviors. Those online behaviors can be digitally captured, analyzed, and modeled to create quantifiable behavioral indicators of RW demographic and personal characteristics of the players. Over a 3-year period, 5 research teams studied over 15,000 players from 9 countries in 12 different MMOGs and VWs. The results of some of this research, along with other related work, are presented in this book.

The “mapping” of the VW and RW can be bi-directional. That is, RW characteristics and experiences can influence an individual’s behavior in VWs. Similarly, experiences in VWs can have an impact on behavior and attitudes in the RW. A sampling of the research literature shows that both premises can be supported.

As an example of the RW influencing the VW, studies of various types of social media have found that the offline personality characteristics of the user are portrayed in online behavior. Researchers found they could predict the personality characteristics (as measured by a Big 5 personality scale) of *World of Warcraft* (WOW) players by examining their game play behavior as recorded in the WOW Armory.<sup>5</sup> Behaviors such as questing, playing solo, doing dungeon runs, and completing achievements, were indicative of personality characteristics such as introversion, openness, or conscientiousness.

This premise does not apply just to games. In a study of Twitter, language features were found to be predictive of personality characteristics such as openness and agreeableness.<sup>6</sup> A study of the Chinese messaging service Renren found that type and frequency of blog postings could be predicted from Big 5 personality characteristics.<sup>7</sup>

The research described here has concentrated on predicting a general class of characteristics such as personality, gender, and age. At times, however, an individual may reveal much more about themselves that can lead to specific identification.

---

<sup>4</sup> Steiner, P. L. (1993). <http://www.plsteiner.com/> *The New Yorker*.

<sup>5</sup> Yee, N., Ducheneaut, N., Nelson, L., & Likarish, P. (2011). Inverted elves & conscientious gnomes: The expression of personality in World of Warcraft. *CHI-2011*, 753–762.

<sup>6</sup> Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011) Predicting personality from Twitter. *Proceedings of the 3rd IEEE International Conference on Social Computing*, Boston, Massachusetts, 149–156.

<sup>7</sup> Bai, S., Zhu, T., & Li Cheng, L. (2012). Big-Five personality prediction based on user behaviors at social network sites.

The choice of character name or screen name is one example. Name choices not only reveal gender and nationality,<sup>8</sup> but also can tie an individual person to multiple accounts. Players reuse the same name, or variations of the name, over and over in multiple games, such as “Gimli”, “G!ml!”, or “GimLi”. This is possibly because (1) the players wish for their friends to be able to find them online in multiple games and social media, and (2) most of the social media sites require unique “handles”. Once a user has hit upon a unique combination of letters, numbers, and symbols that no one else has used, they are likely to stick with it.

The influence of the players’ experiences in virtual environments can also bleed over into the real world. Researchers at the Stanford University Virtual Reality Lab have found, for example, something that they call the Proteus Effect. By representing the player as a more physically attractive avatar, they were able to increase subsequent positive dating behaviors.<sup>9</sup> Participants in another study watched their avatar exercising online. Those individuals were more likely to report exercising in the RW in the days after the study, as opposed to those whose avatar did not exercise.<sup>10</sup>

The potential to influence RW behavior through the use of VWs and games has not escaped the attention of the medical community. As another example, young cancer patients were randomly assigned to play either a commercial video game (the control condition), or a special videogame called *Remission*, which challenged the young players to destroy cancer cells with a variety of in-game weapons. During the months-long course of the study, researchers found that the participants who played *Remission* had a higher adherence to their treatment protocols than did the control group.<sup>11</sup>

The RW impact of VW activities may not always be beneficial, however. The potential harm from playing videogames has in fact received the majority of the attention from mainstream media. Multiple studies have shown some relationship between playing violent videogames and violent tendencies.<sup>12</sup> This begs the classic scientific question: is this a case of correlation or causation? In those cases where the study has been experimental, with players randomly assigned to play violent versus nonviolent videogames, the effect is still found. If we posit that virtual experiences can have positive effects on RW behaviors, then we must accept that the negative case may also be true.

---

<sup>8</sup> Mateos, P., Longley, P. A., & O’Sullivan, D. (2011). Ethnicity and population structure in personal naming networks. *PLoS ONE* 6(9): e22943.

<sup>9</sup> Yee, N., & Bailenson, J. N. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, 33, 271–290.

<sup>10</sup> Fox, J., & Bailenson, J. N. (2009). Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology*, 12, 1–25.

<sup>11</sup> Kato, P. M., Cole, S. W., Bradlyn, A. S., & Pollock, B. H. (2008). A video game improves behavioral outcomes in adolescents and young adults with cancer: A randomized trial. *Pediatrics*, 122, e305–e317.

<sup>12</sup> Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal and prosocial behavior: A meta-analysis. *Psychological Science*, 12, 353–359.



As a final example of the positive influence of experience in VW environments, consider the large amount of interest in the use of games as training tools. Multiple organizations are experimenting with the use of games for Science, Technology, Engineering, and Math (STEM) education. IARPA, in the Sirius program, is experimenting in this area as well, exploring the use of serious games for teaching individuals to recognize and mitigate cognitive biases in critical thinking.<sup>13</sup> Early research results show that the Sirius games have powerful, long-lasting impacts on judgment and decision-making choices.

As a final thought before you read this book, I will mention another concern that scientists must carefully consider: research ethics. The ability to access large data sets from a variety of social media brings with it a responsibility to the users of that social media. Privacy advocates warn of the possible abuse of large-scale data-mining, and the ability to identify supposedly anonymous individuals through their online activities. Research shows that this identification may indeed be possible.<sup>14</sup>

The question of whether the individuals who have signed up to play MMOGs, or participate in other sorts of social media, have consented to have their captured data used for research, is a vexing one.<sup>15</sup> Most of us bypass reading the Terms of Service, clicking the “I Agree,” without understanding that we may have given permission for our online activities to be recorded, sorted through, given to a research team, or sold to marketing companies. Researchers in the area of VWs and MMOGs must continue to hold themselves to the highest ethical standards, complying with applicable human subject research regulations. It is only through the exercise of great care in the protection of human subjects’ privacy that research in this area can continue.

Program Manager  
Intelligence Advanced Research  
Projects Activity (IARPA)  
Office of the Director of National Intelligence

Rita M. Bush, Ph.D.

---

<sup>13</sup> Insert URL for Sirius.

<sup>14</sup> Jones, R., Kumar, R., Pang, B., Tomkins, A. (2007). “I know what you did last summer”—Query logs and user privacy. *CIKM’07*, Lisboa, Portugal.

<sup>15</sup> Fairfield, J. A. T. (2012). Avatar experimentation: Human subjects research in virtual worlds, *U.C. Irvine Law Review*, 2, 695–772.

# Contents

<b>On the Problem of Predicting Real World Characteristics from Virtual Worlds</b> .....	1
Muhammad Aurangzeb Ahmad, Cuihua Shen, Jaideep Srivastava and Noshir Contractor	
<b>The Use of Social Science Methods to Predict Player Characteristics from Avatar Observations</b> .....	19
Carl Symborski, Gary M. Jackson, Meg Barton, Geoffrey Cranmer, Byron Raines and Mary Magee Quinn	
<b>Analyzing Effects of Public Communication onto Player Behavior in Massively Multiplayer Online Games</b> .....	39
Kiran Lakkaraju, Jeremy Bernstein and Jon Whetzel	
<b>Identifying User Demographic Traits Through Virtual-World Language Use</b> .....	57
Aaron Lawson and John Murray	
<b>Predicting MMO Player Gender from In-Game Attributes Using Machine Learning Models</b> .....	69
Tracy Kennedy, Rabindra (Robby) Ratan, Komal Kapoor, Nishith Pathak, Dmitri Williams and Jaideep Srivastava	
<b>Predicting Links in Human Contact Networks Using Online Social Proximity</b> .....	85
Annalisa Socievole, Floriano De Rango and Salvatore Marano	
<b>Identifying a Typology of Players Based on Longitudinal Game Data</b> . . . .	103
Iftexhar Ahmed, Amogh Mahapatra, Marshall Scott Poole, Jaideep Srivastava and Channing Brown	
<b>Index</b> .....	117

# Contributors

**Muhammad Aurangzeb Ahmad** University of Minnesota, Minneapolis, MN, USA

**Iftekhar Ahmed** University of North Texas, Denton, TX, USA

**Meg Barton** Leidos, Arlington, VA, USA

**Jeremy Bernstein** Sandia National Labs, Albuquerque, NM, USA

**Channing Brown** University of Illinois at Urbana Champaign, Urbana, Illinois

**Noshir Contractor** Northwestern University, Evanston, IL, USA

**Geoffrey Cranmer** Leidos, Arlington, VA, USA

**Komal Kapoor** Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA

**Tracy Kennedy** Department of Communication, Popular Culture and Film, Brock University, St. Catharines, Canada

**Kiran Lakkaraju** Sandia National Labs, Albuquerque, NM, USA

**Aaron Lawson** Speech Technology and Research (STAR) Lab, SRI International, Menlo Park, CA, USA

**John Murray** Computer Science Laboratory, SRI International, Menlo Park, CA, USA

**Amogh Mahapatra** University of Minnesota, Minneapolis, MN, USA

**Salvatore Marano** Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica (DIMES), Università della Calabria, Rende, Italia

**Nishith Pathak** Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA

**Marshall Scott Poole** University of Illinois at Urbana Champaign, Urbana, Illinois

**Mary Magee Quinn** Leidos, Arlington, VA, USA

**Rabindra (Robby) Ratan** Department of Telecommunication, Information Studies and Media, Michigan State University, East Lansing, MI, USA

**Byron Raines** Leidos, Arlington, VA, USA

**Floriano De Rango** Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica (DIMES), Università della Calabria, Rende, Italia

**Jaideep Srivastava** University of Minnesota, Minneapolis, MN, USA

**Cuihua Shen** University of Texas at Dallas, Richardson, TX, USA

**Annalisa Socievole** Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica (DIMES), Università della Calabria, Rende, Italia

**Jaideep Srivastava** Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA

University of Minnesota, Minneapolis, MN, USA

**Carl Symborski** Leidos, Arlington, VA, USA

**Jon Whetzel** Sandia National Labs, Albuquerque, NM, USA

**Dmitri Williams** Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, CA, USA

# On the Problem of Predicting Real World Characteristics from Virtual Worlds

Muhammad Aurangzeb Ahmad, Cuihua Shen, Jaideep Srivastava  
and Noshir Contractor

**Abstract** Availability of massive amounts of data about the social and behavioral characteristics of a large subset of the population opens up new possibilities that allow researchers to not only observe people’s behaviors in a natural, rather than artificial, environment but also conduct predictive modeling of those behaviors and characteristics. Thus an emerging area of study is the prediction of real world characteristics and behaviors of people in the offline or “real” world based on their behaviors in the online virtual worlds. We explore the challenges and opportunities in the emerging field of prediction of real world characteristics based on people’s virtual world characteristics, i.e., what are the major paradigms in this field, what are the limitations in current predictive models, limitations in terms of generalizability, etc. Lastly, we also address the future challenges and avenues of research in this area.

## 1 Introduction

When the number of factors coming into play in a phenomenological complex is too large scientific method in most cases fails.

—Albert Einstein in *Out of my later years* [12]

Although somewhat simple but it would not be a great exaggeration to state that the Sciences consist of two important parts—the descriptive and the predictive. Yet, for the most part, predicting human behaviors has been a more challenging task than describing them. The reasons behind this are twofold: first, until very recently, it has

---

M. A. Ahmad (✉) · J. Srivastava  
University of Minnesota, Minneapolis, MN, USA  
e-mail: mahmad@cs.umn.edu

C. Shen  
University of Texas at Dallas, Richardson, TX, USA  
e-mail: shencuihua@gmail.com

J. Srivastava  
e-mail: srivasta@cs.umn.edu

N. Contractor  
Northwestern University, Evanston, IL, USA  
e-mail: nosh@northwestern.edu