Something In The Way We Move: Quantifying patterns of exploration in virtual spaces

John Murray
SRI International
Menlo Park, CA
jxm@sri.com

Edmond Chow
SRI International
San Diego, CA
ed.chow@sri.com

Christopher Connolly
SRI International
Menlo Park, CA
connolly@ai.sri.com

ABSTRACT
In this paper, we examine how behavioral characteristics related to spatial mobility and movement in a multi-player game may differ between male and female players. The study formed part of a large international research project called VERUS. Data was gathered from 380 players in a specially-designed multi-level fantasy game. A machine learning approach was used to identify association rules to differentiate the respective gamer behaviors. Using appropriate threshold levels, the rules achieved a precision of 99% for males and 89% for females. We believe that this quantitative technique offers an innovative way to characterize the gender differences associated with how people interpret and experience their virtual settings.

Categories and Subject Descriptors
H.1.2 [**Information Systems**]: User/Machine Systems – *Human factors, Human information processing.*

K.8.0 [**Personal Computing**]: General – *Games.*

General Terms
Measurement, Performance, Experimentation, Human Factors.

Keywords
Multiplayer gaming behavior, Machine learning, Gender studies.

1. INTRODUCTION
The *Virtual Environments/Real Users Study*, or VERUS, project was a recent multi-year research initiative that focused upon video game players, specifically exploring the relationships between their in-game activities in massively multiplayer online (MMO) games and their external real-world (RW) characteristics.[7,16] The underlying premise was that certain RW qualities of gamers, in particular their demographic features, are correlated with their virtual world behaviors. VERUS researchers examined particular in-game behaviors that could predict the RW characteristics of individuals or groups, including potential in-game attributes that would act as gender differentiators among players.

Gender differences in access to, uses of, and competence with game technologies and similar virtual worlds (VWs) has been the subject of research interest for a considerable time.[11] From a socio-cultural perspective, virtual environments can be seen as a context for players to "perform" their gendered identities.[3,12] Various prior studies of MMO player behavior noted that the selection of male or female avatar characteristics is often closely linked to a player's real-world gender.[10,19] Within the VERUS project, we sought to move beyond this generic pattern by investigating what additional in-game behavioral differences might be evinced through gameplay.

In particular, we conjectured that certain behavioral characteristics related to spatial mobility and movement in VWs might provide further insights, since studies of gender and mobility in the real world suggest that the "rightful" allocation of space to women is measurably less than men's. Historically and culturally, women have been consistently less able ("free") than men to move about in general, and to explore new territory in particular; this is true across ages, over time, and across racial and cultural contexts.[2] In addition, in terms of objectively measured spatial abilities, females consistently prove less adept at navigation than males.[1] Given that virtual-world game communication and interaction are proving to be hostile to female gamers, we reasoned that these consistently gendered spatial restrictions and mobility norms would be similarly evident, or even more so, in player/avatar movement in virtual space.[16]

Recent gender studies researchers have used a virtual water maze to examine differences in spatial navigation search strategies. [5,17] They found that female participants – even 'experts' – generally covered less space than males in trying to locate a hidden objective. This insight led us to design an expanded study of movement features with greater face validity in a gaming context. We used *Guardian Academy (GA),* a multi-level fantasy game that was specially designed for the VERUS project.

In the following sections, we discuss the data collection process, our data refinement and analytical techniques, the results and conclusions, and some notes on future research directions.

2. METHODS AND PROCEDURES

2.1 Data Collection Process

2.1.1 Real and Virtual World Settings
VERUS data was collected at multiple locations in several countries, including Canada, the United Kingdom and the United Arab Emirates. The in-world movement data was gathered in a specially-designed persistent multi-player virtual world called *Guardian Academy (GA)*, where players interacted via their avatars and comprehensive game telemetry data on all activity was logged. Logged activity included social exchanges and text chat interactions with non-player characters (NPCs) and among players, as well as combat, trading, and crafting, and other behaviors. *GA* was a multi-level sword-and-sorcery fantasy world implemented as a 2.5-dimension web-centered, Flash-based, game. See Figure 1.

Proceedings of the 10th International Conference on the Foundations of Digital Games (FDG 2015), June 22-25, 2015, Pacific Grove, CA, USA. ISBN 978-0-9913982-4-9.

Players initially selected and named a male or female avatar. NPCs provided quests that were completed to earn points. Players could form groups with each other to complete quests or to explore the world together. *GA* included an in-game exchange for trading health potions, weapons, and other items. It also featured several separate player-versus-player (PvP) arenas where private duels could be held.

Figure 1: Guardian Academy

2.1.2 Participants

Data was collected from a total of 380 participants, of whom 97 (25.5%) were female. The population comprised both adults and youth under 18 years of age. The study settings included research laboratories and informal settings such as video-gaming clubs and events. Approximately one-third of the participants were youth (aged 8-17) who were recruited through schools, and summer camps. After informed consent (and parental agreement where appropriate) was received, a web-browser-based demographic survey was administered to all participants, to obtain their RW characteristics as ground-truth.

Only our registered participants and research staff members had access to the *GA* world. Gameplay was observed for at least 20 minutes and up to 45 minutes in the primary experimental sessions. Since the game client was web-based and the world was persistent, participants had the opportunity to re-enter the game and continue playing elsewhere, after the primary sessions ended.

2.2 Derivation of VW Movement Features

2.2.1 Event Log Structure

The *Guardian Academy* event log files recorded the two-dimensional position versus time of avatars as they move in the virtual world. For each user, the log entries detail a location identifier along with the avatar's time-stamped *x*, *y* position expressed in a location-specific coordinate frame. Metrics that distinguish the movement of different players can be distilled from this position information, yielding an important class of motion-derived features that encapsulate behavior associated with the position and velocity of avatars.

2.2.2 Histograms

Once the location data are extracted, it is possible to compute spatial and velocity histograms. These histograms can be computed over all observations in the data set, or they can be segregated by other properties like VW region or avatar or user. These enable us to compute the ***Area Covered***, i.e. the extent or range of a player's movement, and the ***Distance Travelled***, which is the length of the path traced over time. We also compute the 2-dimensional *change* of user movement at each time step, and aggregate each user's relative displacements into a *step histogram*. Figure 2 depicts two step histograms; the upper histogram has high entropy, while the lower has low entropy. The *x* and *y* axes represent horizontal and vertical displacement. Each histogram is colored by the frequency of visits to the corresponding positions: black is 0, blue indicates a low number of visits, and red indicates a high number. We can also compute *velocity histograms,* which are constructed from a three-point approximation to the derivative of smoothed motion.

2.2.3 Entropies

While *Area Covered* and *Distance Travelled* provide useful movement information, they do not adequately capture the frequency of a player's visits to certain locations. We therefore characterize the frequency as well as the extent of travel using *Shannon entropy* [4]. This gives us a measure of the "variability" or "disorder" of the travel pattern, and accounts for the frequency of visits to specific positions, as well as the variety of positions visited. The Shannon entropy $H(X)$ of random variable X is

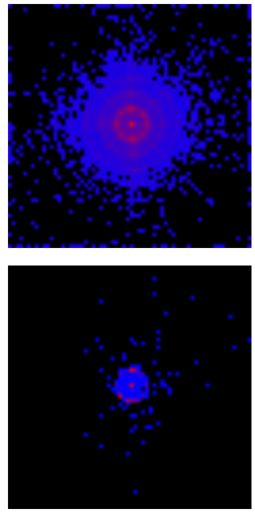

Figure 2: Step histograms for two players

$$H(X) = -\sum P(x_i) \log P(x_i) \quad (1)$$

where $P(x_i)$ is the normalized frequency for value x_i. In the present context, $P(x_i)$ is the fraction of the total histogram mass corresponding to bin i. The Shannon entropy can be seen as the minimum number of bits needed to encode the distribution of X. Figure 3 shows the qualitative effect of different entropies, where the concentrated green (lower entropy) histogram can be more easily encoded than the wider-spread red (higher entropy) one.

Entropy measures are used to assess exploratory behavior. We define the ***Propensity to Explore*** as the extent to which a player travels throughout the overall environment, and quantify it as the entropy of the position histogram (the *place entropy*). This represents not just the area travelled, but also frequency of visits to specific places within the game. We also define the ***Propensity to Wander*** as the variability in the player's direction and magnitude of movement from one time point to the next, and quantify this variability by the entropy of the step histogram (the *step entropy*).

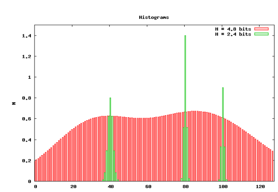

Figure 3: Effect of histo-gram shape on entropy

Formally, for a given player, we calculate the propensity to explore using Equation (1) for x_i ranging over the possible positions observed at each time step and $P(x_i)$ the observed frequency of position x_i. Similarly, we calculate the player's propensity to wander using Equation (1) with x_i ranging over the possible displacements observed from one time step to the next.

While these entropy quantities are difficult for a human observer to calculate numerically, one can argue that they capture concepts that can be perceived rather readily: place entropy captures the tendency to *fully* explore one's surroundings, while step entropy measures the tendency to wander in undirected, irregular paths. It is nonetheless possible to familiarize the observer with place entropy and step entropy by presenting examples of player movement and the associated entropy values. Given exemplars illustrating the range of movement, one can gain a sense of the distribution of propensity, or, at minimum, what behavior exceeding the 70th percentile or 90th percentile looks like.

2.3 Analysis Using Machine Learning

2.3.1 Association Rules Methodology

In our approach to analyzing the data, we sought to identify machine learning **association rules** of the form IF PREMISE, THEN CONCLUSION, where the premise is a **conjunction** of **assertions** about in-game features, and the conclusion is a statement about a real-world characteristic of the player. An example rule would be "IF *Avatar Sex* is male AND *Area Covered* exceeds *[defined-threshold]*, THEN *Real-World Gender* is male," which has two assertions and one conclusion.

Our association rules are produced using the ***WEKA*** workbench of machine learning tools [9], via the ***Tertius*** rules-discovery extension [8], using the following methodology: Input data consists of features derived from multiple data streams, adjoined with the binary target of interest (such as "Real-World Gender is male"). For real-world characteristics that are multivalent, i.e., possessing more than two class values, we convert these RW characteristics into the requisite number of binary targets, and seek rules to predict each class value of interest. We submit only records for which every feature and target value is non-missing. Tertius then returns a list of rules that predict the target. The list is reviewed by subject matter experts, who retain only rules meeting a target precision threshold and that, in their opinion, also pass the test of validity and defensibility.

An important methodological decision was spurred by a limitation of the Tertius module. Tertius treats all input variables as categorical, and moreover returns only assertions of the form 'Variable = value'. While this form of assertion is acceptable for features with few possible values, it yields rules that are far too specific (and with low recall) if continuous numeric variables are involved. Our solution is to replace each numeric variable with a small number of binary features of the form 'Variable > value', for various cutoff values chosen from the observed range of the variable. This coarsening of the original variable not only speeds computation but also prevents the rule-mining from overfitting.

For this study, we discretized the VW movement measures into deciles, such as 'Variable exceeds the 60th percentile'. By reporting the cutoff in decile form, we obtain an assertion that is scale- and distribution- independent: the assertion is directly portable to another virtual world where the units of measurement are different. The cutoffs are equally spaced along the distribution of the measure of interest, yielding a maximally discriminating list of assertions. Reporting cutoffs in decile form also enables a human to perceive what is being asserted: is the player at the extreme end of the measure distribution, or middling? To determine the truth of such an assertion for a particular game player, one can sort all players according to the measure value, and place the player of interest along that continuum. The decile specification puts the assertion in proper context, without unnecessarily narrowing the discrimination factor.

2.3.2 Ensuring Validity of Rules

Rule learning was performed on all observations; no data was set aside for validation. Our discretization of continuous variables into a limited number of binary features serves to maintain predictive power while reducing the risk of finding spurious rules. We further increase generalizability by limiting each rule to at most two assertions, and adopting several best practices.

For example, we select rules in conformance with both statistical experts and subject matter experts. Thus, the former verify that confounding factors do not lead to non-portable rules, while the latter assess how defensible an attribute-predicting rule is in the context of relevant social-science-based theories. In the current case, rules that successfully predict gender from VW player movement echo recent research that found gender differences in mobility norms.[5,17]

Similarly, we minimize the effect of prior game experience and skill. Some participants will possess knowledge of, and be quite skilled in, other VWs and MMOGs. The confounding effect of skills transferred from prior game play experience can be reduced, but not eliminated, by introducing participants to unfamiliar virtual worlds, so that all players start with the same amount of knowledge of these new worlds, and by observing gameplay up to the same time limit. Using *Guardian Academy* – a game new to all participants – helped to address this issue. However, we recognize that familiarity with movement-based activity in a game, as opposed to its ludic features, may favor players who have experience with MMOGs or questing and combat.

Another key strategy was to construct and examine features that are world-independent. By reporting features as percentages of time played, and/or expressing assertions in percentile terms, we ensure features do not depend on the duration of the observation period or on the mechanics or constraints of gameplay.

2.3.3 Analytical Performance Measures

We evaluate the performance of each derived rule according to familiar binary classification metrics. For the association rule "IF A, THEN B", we count the number of True Positives (instances for which both A and B are true), False Positives (A is true, B is false), False Negatives (A is false, B is true) and True Negatives (both A and B are false), and calculate Coverage = (TP + FP)/n, Precision = TP/(TP + FP), Recall = TP/(TP+FN), and F-measure = 2 × Precision × Recall / (Precision + Recall), with n the total number of instances studied.

For each real-world characteristic we augment the final list of association rules with the **union** rule that combines the premises of the component rules. The rules selected for combination are chosen to maximize the recall of the union rule, while meeting or exceeding a target threshold for precision. It is important to report the performance of this union rule, since presenting a list of rules is tantamount to proffering their union. In this regard, it is worth remarking that precision almost always goes down when rules are united, while coverage and recall can only improve.

3. RESULTS AND DISCUSSION

3.1 RW Gender Prediction Rules

Prior game studies projects have noted typical generic mappings between a player's real-world gender and their choice of the avatar's sex.[10,19] More recent research has explored how other in-game behaviors may be linked with RW gender.[13,18] Other VERUS work found that language usage and avatar naming can be effective predictors of RW gender, especially for males.[14,15]

In the present study, we reapply the simple avatar sex predictive rule to *Guardian Academy* data. The precision of this baseline rule – that RW gender matches the VW choice – is 93.5% (with recall 81.2%) for predicting RW male gender, and 60.7% (recall 84.5%) for predicting RW female gender. We improve upon this base by identifying additional in-game behavioral differences. For predicting male RW gender, augmenting the baseline rule with movement features yields rules with precision exceeding 98%, but at the cost of a significant drop in recall (and coverage). The implication is that, for the relatively small proportion of players that satisfy the assertion, the prediction of RW male gender is near perfect. For the prediction of female RW gender, the augmented rule raises the precision very close to 90%, with good recall.

3.2 Analytical Results

Machine learning association rules (two assertions and one conclusion) were used to characterize the RW-VW relationship, with the goal of achieving a precision threshold goal, while also maximizing the associated recall level. The *GA* player population is 380 total – 282 males and 97 females. (The gender of one player is unidentified.) For males, predictive precision is substantially improved with a 60th percentile lower cutoff for each of the three movement metrics. Table 1 displays the rules and their precision, recall, F-measure, and coverage evaluated on n = 380. For females, adopting a 40th percentile upper threshold for each movement metric yields correspondingly improved precision; see Table 2. In each case, the union of the rules for the individual metrics has improved recall over any individual rule.

Without a holdout set, there is a risk that these results are optimistic; but as discussed earlier, we believe our methodology implements sufficient safeguards to minimize overfitting.

Table 1: Association rules to predict RW gender Male

In the RW, Player is Male if......					
Rule	**In the VW.....**	**Pre**	**Rec**	**F**	**Cov**
Base	Avatar is male (AM)	93.5	81.2	88.9	64.5
1	AM AND AreaCovered > 114 (i.e. 60 %ile)	99.0	35.5	52.2	26.6
2	AM AND Prop.Explore > 4.46 (i.e. 60%ile)	99.0	35.5	52.2	26.6
3	AM AND Prop.Wander > 3.40 (i.e. 60%ile)	99.0	35.8	52.6	26.8
Union	RULE 1 OR 2 OR 3	98.4	44.7	61.5	33.7

Table 2: Association rules to predict RW gender Female

In the RW, Player is Female if......					
Rule	**In the VW.....**	**Pre**	**Rec**	**F**	**Cov**
Base	Avatar is female (AF)	60.7	84.5	70.7	35.5
1	AF AND Area Covered < 37 (i.e. 40%ile)	87.3	49.5	63.2	14.5
2	AF AND Prop. Explore < 3.27 (i.e. 40%ile)	87.0	48.5	62.3	14.2
3	AF AND Prop. Wander < 2.89 (i.e. 40%ile)	88.3	54.6	67.5	15.8
Union	RULE 1 OR 2 OR 3	88.9	57.7	70.0	16.6

3.3 Discussion

For the population and VW studied, these findings suggest that in-game movement patterns can be used analytically to supplement other behavioral indicators of gender – in this case, the sex chosen for the avatar. The 40th and 60th percentile cutoffs for the movement metrics provide a means to partition players into the predominantly RW male and predominantly RW female bins, with a group of unknowns between these two cutoff values.

The entropy quantities may be difficult to calculate numerically, but they embody behaviors that can be perceived rather readily: the tendency to fully explore one's surroundings, and the tendency to wander in undirected, irregular paths. We believe that providing human observers with exemplars from a continuum of behaviors can enable them to develop an ability to assess the level of a particular movement metric exhibited by a given player.

4. FUTURE DIRECTIONS

This work forms a small part of the overall VERUS project, which involved a much larger exploration of the mappings among numerous behaviors in a range of RW and VW settings. In this paper, we have discussed just one form of behavior (movement) where players used a single avatar in one multiplayer game. Much of the data was gathered in naturalistic settings rather than formal laboratories, and this basic analysis does not fully address potential confounding due to location, context and surroundings, player age group, time played, etc. Likewise, technical issues such as hardware performance, network latencies, browsers used, etc. remain unaddressed. All that being said, we hope that the concept of movement entropy contributes an innovative quantitative technique to the field, one which helps to characterize how different genders may interpret and experience their virtual settings. We look forward to future opportunities to apply similar analytical techniques to player movement in other virtual worlds.

We note that the affordances of different VWs have a significant influence on players' in-world activity, and it is therefore difficult to generalize from one world to others. We are currently working on techniques to normalize game space data, so that it becomes more feasible to make meaningful player movement comparisons across different VWs. However, on a broader level, we also we need to be conscious of the inherent methodological challenges that research in this overall field entails. In particular, we recognize the danger of conflating insights about the game with insights about the players, and we echo our VERUS colleagues' calls to establish firmer theoretical foundations for this work.[6]

A key topic of interest for future work concerns a single player's potential use of multiple avatars in a given virtual environment. *Guardian Academy* offered participants very few personalization choices, and for this study we restricted them to playing with just one avatar, either male or female. However, as noted above, VWs may be seen as venues where players have opportunities to "perform" their – potentially multiple – identities [12]. There is much scope for studying players' use of different avatars to introduce and explore diverse aspects of their RW identity in a virtual space. The extent to which their patterns of movement in-world may be part of that exploration remains an open question; perhaps the present work offers quantitative support for such research.

Another related research area concerns the movement patterns of two or more gamers playing together in the same virtual space. The VERUS project embarked upon an initial exploration of this topic, using a framework for coding in-game observed behaviors that can yield predictive characteristics for the real-world relationship of a given pair of players.[17] The coding framework includes constructs for proximity, collaboration, and behavioral synchronicity. Supplementing this observational approach with quantified characterizations of in-world movement may be a fruitful area of research, not just for examining pairs of gamers, but also for studying larger groups playing in unison as teams.

5. ACKNOWLEDGMENTS

We would like to acknowledge the extensive work of our VERUS colleagues at York and Simon Fraser Universities in Canada, and at Nottingham University in the United Kingdom. We also thank the anonymous reviewers for their helpful insights and suggestions for improvement. This research was undertaken with support from the U.S. Air Force Research Laboratory under contract FA8650-10-C-7009.

6. REFERENCES

[1] Allen, G. 2000. Men and women, maps and minds: Cognitive bases of sex-related differences in reading and interpreting maps. In S. Nualláin, *Spatial cognition: Foundations and applications: Advances in consciousness research.* John Benjamins Publishing, Amsterdam Netherlands.

[2] Ardener, S. 1981. *Women and Space: Ground Rules and Social Maps*. Croom Helm, London.

[3] Bryce, J. and Rutter, J. 2005. Gendered gaming in gendered space. In J. Raessens and J. Goldstein (Eds.) *Handbook of Computer Games Studies*. MIT Press, Cambridge, MA.

[4] Cover, T. and Thomas, J. 1991. *Elements of Information Theory*. New York: John Wiley and Sons.

[5] de Castell, S., et. al. 2010. The Eyes Have It: Measuring Spatial Orientation in Virtual Worlds to Explain Gender Differences in Real Ones. *Technical Report, Simon Fraser University Library,* Vancouver, BC.

[6] de Castell, S., Jenson, J., Taylor, N. and Thumlert, K. 2014. Re-thinking foundations: Theoretical and methodological challenges (and opportunities) in virtual worlds research, *Journal of Gaming & Virtual Worlds* 6:1.

[7] Dieterle E. and Murray J. 2010. Virtual Environment Real User Study: Design and Methodological Considerations and Implications. *Journal of Applied Learning Technology*, 1/(1)

[8] Flach, P. and Lachiche, N. 2001. Confirmation-Guided Discovery of First-Order Rules with Tertius, *Machine Learning*, Vol. 42, 1/2, pp. 61-95.

[9] Hall, M. et al. 2009. The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[10] Huh, S. and Williams, D. 2009. Dude looks like a lady: Gender swapping in an online game, in W. Bainbridge (ed.), *Online Worlds: Convergence of the Real and the Virtual,* New York: Springer.

[11] Jenson, J. and de Castell, S. 2010. Gender, simulation and gaming: Research review and redirections. *Simulation and Gaming,* 41, (1), pp. 51-71.

[12] Jenson, J., Taylor, N., de Castell, S., and Dilouya, B. 2015. Playing With Our Selves, *Feminist Media Studies,* DOI: 10.1080/14680777.2015.1006652

[13] Kennedy, T., et al. 2014. Predicting MMO Player Gender from In-Game Attributes Using Machine Learning Models. *Predicting Real World Behaviors from Virtual World Data.* Springer International Publishing.

[14] Lawson, A. and Taylor, N. 2012. The Names We Play: Exploring MMOG players' avatar naming conventions. Canadian Games Studies Association Symposium, Waterloo, Canada.

[15] Lawson, A. and Murray, J. 2014. Identifying User Demographic Traits through Virtual-World Language Use. In *Predicting Real World Behaviors from Virtual World Data,* Springer International Publishing, Heidelberg.

[16] Lerios, H. 2011. Gender Differences in Virtual Route Learning, PhD Dissertation, School of Interactive Arts and Technology, Simon Fraser University, Vancouver, Canada.

[17] Murray, J. et al. 2012. Reynard VERUS Research Project – Final Report, US Air Force Research Laboratory, Wright-Patterson AFB OH, RY-WP-TR-2012-0286.

[18] Symborski, C. et al. 2014. The Use of Social Science Methods to Predict Player Characteristics from Avatar Observations. In *Predicting Real World Behaviors from Virtual World Data*, Springer International Publishing, Heidelberg.

[19] Yee, N., Ducheneaut, N., Yao, M. and Nelson, L. 2011. Do men heal more when in drag? Conflicting identity cues between user and avatar, *CHI 2011 (Proceedings), Vancouver, Canada*. ACM New York.