Using Social Simulators to Assess Social Learning

Learning is fundamental to human growth and functioning. Through learning, we are able to develop a more comprehensive understanding of our environment and discover ways to interact with it that are more effective and efficient. Oftentimes, the word "learning" is associated with the acquisition of knowledge related to facts, concepts, or procedures (e.g., learning numbers, learning the scientific method, learning how to tie a shoe), but learning is equally important in shaping our understanding of people, as well. The social nature of being human provides countless opportunities for learning by interacting with or simply observing others engaging in activities across contexts. Combinations of personality, experience, and culture creates significant variability in social knowledge that in turn, influences individuals' beliefs, attitudes and future behavior. This type of learning, herein referred to as social learning, is what enables individuals to interact with one another successfully. It is through social learning that individuals learn to build relationships, communicate intent, understand others' perspectives, and predict others behavior. However, realizing these outcomes relies on people's ability to be flexible in response to changes in their social environment. This adaptability (which is often a function of the application of higher-order processes, such as metacognition) is crucial in nearly any interpersonal situation, but especially so in contexts that are complex, novel, or lack predictability, as individuals must be able to make sound judgments and decisions in the absence of familiar cues (Department of the Army, 2013).

As with non-social learning, the ability to learn how to interact effectively with others varies from individual to individual (Riggio, 1986), with some individuals acquiring social information more readily than others. Unlike non-social learning, however, social learning is not easily measured, as the knowledge required for effective performance changes in accordance with contextual factors (e.g., time, place) and interpersonal factors (e.g., goals, expectations, cultural background, personality). To date, most assessments of social learning have focused on small cross-sections of individuals' interpersonal skills, via either self-report (Sodowsky, Taffe, Gutkin & Wise, 1994) or behavioral observation (Ruben, 1976; Graf & Harland, 2005). While self-reported assessments can be helpful in providing holistic judgments of interpersonal effectiveness, they are insufficient at capturing variations in performance within and across interpersonal interactions. Likewise, they also preclude the observation of adaptive behavior that may occur in response to feedback or changes in one's social environment. Field-based observations, on the other hand, allow for more ecologically valid measurement opportunities, but are far from standardized and often costly to implement. One tool that may potentially address these methodological shortcomings is the social simulator. A social simulator is a computer-based environment in which a human learner interacts with one or more computer agents that model human behavior across one or more situations. Recently, social simulation technology has advanced through the creation of social simulation engines in the video game industry and game design research labs (Shapiro, Tanenbaum, McCoy, LeBron, Reynolds, Stern, ... and Coon (2015)). These technologies support a wide range of constraint from fully

autonomous behaviors by AI social agents (also called non-playing characters or MPCs) to author branching where designers define networks of social interaction (Khandaker-Kokoris, 2015). Based upon our preliminary work, we believe that people's patterns of interactions with game-based social simulators over extended periods of time or at multiple intervals can provide strong evidence of many aspects of social learning that have not been possible before. In the proposed research, we explore the viability of game-based social simulators as a method to observe and assess individual differences in social learning.

The use of social simulators in education and training is not a new concept. Many organizations, including the military, have used social simulators to train individuals on a variety of skills such as negotiation (Kim, Hill, Durlach, Lane, Forbell, Core, ... and Hart, 2009).), leadership (Riedl and Stern,2006),.), and cross-cultural competency (Shapiro, Tanenbaum, McCoy, LeBron, Reynolds, Stern, Mateas, Ferguson, Diller, Moffitt, Coon, and Roberts, 2015; Taylor and Sims, 2009). In most of these contexts, however, social simulators have been employed as learning interventions, with performance outcomes serving as the primary focus. For example, Lane, Hays, Core, and Auerbach (2013) found that interacting with the social simulator, BiLAT, fostered learning of intercultural communication skills, such as the use of appropriate greetings and win/win negotiation techniques, which influenced subsequent performance on a situational judgment test (SJT). While assessing performance outcomes is undoubtedly an important function of social simulators, the patterns of actions and choices players make as they interact with a social simulator provide a potential rich source of evidence of social learning as well (Shute, Ventura, Bauer, & Zapata-Rivera, 2009). To date, few empirical studies have systematically explored the value of this kind of evidence for identifying, describing, and predicting learning in simulated interpersonal contexts.

The proposed project will explore the viability and validity of evidence that can be gathered from participants' interactions with social simulators. Answering these fundamental questions is critical for their use in learning and assessment:

1. Are game-based social simulators a viable method to collect evidence of human performances that reflect patterns and strategies of social learning?
2. Can the evidence from performance in social simulations be aggregated into a valid and reliable assessment of social learning?

We will answer these questions in two studies using and expanding upon an existing game-based social simulation prototype of cross-cultural competence (3C). We will first describe aspects of the prototype and then discuss the research studies.

A Social Simulator Prototype for Assessing Social Learning

We have developed an initial prototype to evaluate aspects of social learning that focuses on how people learn to respond appropriately when interacting with others in an unfamiliar social environment, namely an artificial foreign culture. In the prototype, players must overcome

a series of game challenges while learning as much about the culture as they can through observation, action, and reflection. There are several important elements inherent to our design of the game-based environment that make it an ideal method to address fundamental questions about the nature of social learning. First, because the cultures in the simulation, though realistic, are artificial, they can be manipulated to achieve different learning and assessment goals. In other words, while the game-based cultures have no exact real-world counterpart, the core cultural values can be varied to resemble cultures that appear familiar and authentic to the player. The initial cultures were designed using dimensions such as individualism-collectivism, hierarchy/egalitarianism, etc. that are universal to nearly all real-world cultures (Hofstede, 2001; Nolan, LaTour, & Klafehn, 2014). Thus, the simulation's cultural environment includes elements or behavioral patterns that test-takers may find familiar but will not be able to attribute to any particular real-world cultural group or region. This enables us to study social learning and the effects of different cultural factors in a systematic fashion.

Second, the underlying social simulation engine employs a new authoring tool (currently under development by researchers at American University) to support the rapid creation and revision of social simulations. Together the engine and the author tool allow for the rapid creation of a robust range of interactions between human players and NPCs within the game. Specifically, gameplay interactions are defined by overarching sociocultural rules and social practices, as opposed to scripted situations and response options. Thus, gameplay evolves and adjusts in response to choices made in-situ by the player. This approach is ideal for a number of reasons. First, by authoring rules and social practices, the amount of situation- or NPC-specific authoring that is required is substantially reduced (Treanor, McCoy, and Sullivan, 2015). Second, this type of flexible and diverse gameplay more accurately reflects how interactions transpire in the real world, but in an environment that is controllable (through the standardization of rules and social practices) and measurable (through player decisions and NPC responses). Finally, the adaptive nature of gameplay affords opportunities to measure the adaptability of players' responses. As gameplay evolves, it is possible to observe whether players notice, evaluate, and incorporate feedback they receive from NPCs and the extent to which this processing of social information influences overall performance.

Research Studies

The two research studies we propose address the two fundamental questions about the nature of performance and learning in social simulation games and their relation to broader measures of cognitive ability and competency in social learning. The studies build on one another, addressing whether social learning occurs at all in human interactions with social simulators, what "in-game" metrics (a.k.a. process data) from the interactions are indicative of social learning, the internal stability and reliability of the in-game metrics, and to what degree they are associated with cognitive and non-cognitive skills. We are interested in how well specific in-game metrics are associated with outcome measures of learning and performance in the simulation, and for generalizability, the strength of association (e.g. correlation) of these metrics with external

measures like cognitive ability and non-cognitive social measures. Overall, we are interested in determining the validity of an aggregated set of game-related metrics as a measure of social learning.With that in mind, across the two studies we propose to identify and evaluate in-game metrics indicative of social learning that fall into two categories:

1) metrics of on-going performance that may be used to track and monitor learning over time such as a) agreement of player's actions with an ideal model over short spans of play and b) repeated performance on individual social practices over time (e.g. degree of success with each use of social practices such as greeting, trading, persuading, etc.)

2) metrics indicative of specific learning strategies (e.g. strategies such as systematic search/induction, hypothesis testing, confirmatory reasoning). Example metrics in this category include a) presence of specific actions sequences that are associated with learning strategies and b) amount of each strategy used overall (as indicated by the action sequence proxies)

The purpose of study 1 is to identify candidate in-game metrics and outcome measures, while the purpose of study 2 to is to evaluate the validity of the accumulated information across in-game metrics and outcome measures.

Study 1

This study addresses the first fundamental question, namely are game-based social simulators a viable method to collect evidence of human performances that reflects patterns and strategies of social learning?

Method. This study will take the form of a cognitive lab in which participants "think aloud" while working their way through the simulation to achieve specific performance and learning goals. This approach has been shown to be an effective way of identifying components of learning and their connections to performance. Specifically we will follow an elicitation approach similar to Green and Azevedo (2007). We will elicit verbal think-aloud protocols during participants' use of the simulation, and after completion of each major game challenge, participants will articulate their current understanding of effective ways to interact with NPCs to achieve each challenge and their beliefs about the underlying social and cultural values of the NPCs. Forty participants will work through the challenges while 'thinking aloud.' We will record all game telemetry (person-game interactions) and record their think-aloud verbalization and videotape their game play.

Coding and Scoring.
Green and Azevedo (2007) provide a coding scheme for think aloud protocols focused on aspects of learning, and including components of metacognitive evaluation and regulation. We will use a modified set of codes that also build upon the coding we have developed for learning strategies

about social practices and cultural values that have come out of our current work. We will also score the participants' articulated beliefs for each challenge with respect to their completeness and accuracy. We will identify key shifts in belief about effective interactions and underlying cultural values from one challenge to the next.

Analysis.

To address question 1, our goals for the analysis are to

1. Identify hypothetical social learning strategies from coglab verbal protocols
2. Identify candidate patterns in the telemetry that either a) co-occur with coded verbal indicators of strategies from the think aloud data or b) are directly associated with social learning strategies (e.g. systematic search can likely be identified from telemetry) and shifts in articulated belief.
3. Identify in-game performance metrics (variants of speed, accuracy including match to expert models for segments of game play)
4. Identify candidate outcome measures success (achievement of game challenges, overall speed and accuracy, accuracy of end-of-game articulated beliefs) and create a separate overall assessment for the social declarative-type knowledge players should learn by playing the game

The analysis will evaluate the degree to which social learning strategies articulated in the verbal protocols are associated with patterns in the telemetry. This will provide some evidence on the viability of using game play telemetry as a basis for making inferences about social learning strategies. We will also explore the extent to which different strategies are associated with shifts in articulated beliefs about the NPCs and their culture.

The primary analysis will use a mixed methods approach that focuses on identifying patterns of decisions in the game telemetry that are associated with participants' articulated changes in belief in the think aloud data and beliefs they articulate after each challenge. We will use qualitative analyses of the performance data to identify candidate patterns of game behavior, learning strategies, and learning trajectories – qualitatively different levels of sophistication in approaches to learning. We will then apply machine learning techniques including partially observable Markov decision processes (POMDPs; Kaelbling, Littman, & Cassandra, 1998) to develop preliminary detectors for these patterns based on the logged game interactions. Because the simulation engine inherently provides a model of social decision making, the player's behavior can be compared to ideal behavior according to the simulation engine. As a secondary analysis we will explore whether there are similarities between patterns of participants' decisions and the engine's algorithmic decisions.

Study 2

This study will begin to address the second foundational research question: Can the evidence from performance in social simulations be aggregated into a valid and reliable assessment of social learning? In order to answer this question we will run a study with a much larger sample to explore the statistical properties of the patterns in the telemetry and their relation to external measures. A psychometric model will be developed to aggregate information across the in-game

and outcome metrics and preliminary reliability and validity evidence will be analyzed. We will address:

a) the stability and reliability of the in-game metrics i.e. can we detect the sequences automatically? are they consistent across different scenarios? Do they correlate with game outcome metrics? how are the sequences distributed across challenges and players?
b) the external construct validity, i.e. the strength of association between the game measures and external measures of cognitive and social abilities along with demographic variables believed to correlate with 3C competency.

With respect to the relation among the measures we will evaluate the following hypotheses:

1. We expect the in-game performance measures to predict the out of game performance measures well.
2. We expect the metrics on the in-game learning strategies to predict the out of game learning measure (this hypothesis claims that different learning strategies will result in different learning outcomes).
3. We expect moderate but significant associations of in-game metrics to the external measures of cognitive and social abilities.

Method. To collect data to address external construct validity, candidate participants will complete a short battery of assessments that will provide estimates of overall cognitive ability and social competency along with a survey of relevant demographic information. For cognitive ability we will focus on assessments of fluid intelligence) and crystalized intelligence e.g. (e.g. Raven's Progressive Matrices as a measure of fluid intelligence and WAIS IV measures of fluid and crystalized intelligence). Candidate assessments for social competency include the WAIS-R Picture Arrangement and Comprehension assessments,). Four Factor Tests of Social Intelligence (O'Sullivan & Guilford, 1976), and the Social Skills Inventory (Riggio, 1989). One thousand participants will be selected that represent a range of performance across the cognitive and social measures. Selected participants will each play through six of challenges. Participants' game play telemetry will be saved, parsed, and analyzed.

Analysis. We will use the candidate patterns of performance, learning strategies, and learning trajectories identified in the study 1 and the prototype pattern detectors developed from psychometric modeling as a starting point for analysis of Study 2 data. With this larger data set, we will refine the pattern detectors developed for the social learning tasks. General patterns of performance will be identified using aggregate count statistics for single and n-gram actions taken during different challenges. However, more sophisticated decision-theoretic models (i.e. POMDPs) will be applied to the player actions, from which we expect to be able to identify evidence of both strategic approach (e.g. exhaustive search vs. targeted information seeking) and learning orientation (e.g. highly adaptive vs. resistant to belief change). We will validate these classifications using a smaller portion of human-coded performance records.

Each of these pattern detectors will serve to identify evidence of the social learning construct which we are measuring. We will then combine this evidence and evidence from the game outcomes using a more comprehensive psychometric model. Our initial approach will involve a 2PL item response theory (IRT) model with possible multidimensionality. As IRT models assume conditional independence of evidence nodes (item responses), we will test for problematic local item dependence using ability-banded response correlation matrices. If the

local item dependency in this model is found to be problematic, we will expand the model to a full Bayesian network which is able to model that dependency.

To answer question 2, our measures will be checked for reliability and validity using 1,000 participant data collection. Internal validity will be analyzed by examining the correlation between individual evidence nodes and the overall measure. External construct validity will be analyzed using the correlations between our measure, the external cognitive and social skills measures, and the demographic variables. Response process validity will be examined through hand coding a subset of the full records.

Future Work

If the results of these early studies are promising, follow on work will address additional factors that affect learning, transfer of learning to actual performance contexts, and scalability of the methodology to other constellations of social competence and learning. One area of interest is the degree to which cultural and interpersonal knowledge are subject to different kinds of cognitive bias. For example, those participants with strong prior cultural knowledge of a single may be more prone confirmation biases that will result in lower performance when faced with learning within situations that are affected by different cultural dimensions and norms.

References

Department of the Army. (2013). TRADOC Pam 525-8-2.

Graf, A., & Harland, L. K. (2005). Expatriate selection: Evaluating the discriminant, convergent, and predictive validity of five measures of interpersonal and intercultural competence. *Journal of Leadership & Organizational Studies*, *11*(2), 46-62.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101 (1), 99-134.

Kim, J. M., Hill Jr, R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., ... & Hart, J. (2009). BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education*, *19*(3), 289-308.

LaMar, M. M. (2014). *Models for understanding student thinking using data from complex computerized science tasks*. (Unpublished doctoral dissertation). University of California, Berkeley.

Lane, H. C., Hays, M. J., Core, M. G., & Auerbach, D. (2013). Learning intercultural communication skills with virtual humans: Feedback and fidelity. *Journal of Educational Psychology*, *105*(4), 1026.

McCoy, J., Treanor, M., Samuel, B., Reed, A.A., Mateas, M. & Wardrip-Fruin, N. (2014). Social Story Worlds with Comme il Faut. In IEEE Transactions on Computational intelligence and AI in Games. 6(2), 97-112.

Mislevy, R., Oranje, A., Bauer, M. I., von Davier, A., Hao, J., Corrigan, S., DiCerbo, K. E., Hoffman, E. & John, M. (2014). Psychometric considerations in game-based assessment. *GlassLab Report*.

Rafferty, A. N., LaMar, M. M., & Griffths, T. L. (2015). Inferring Learners' Knowledge from their Actions. Cognitive Science, 39 (3), 584-618.

Riedl, M. O., & Stern, A. (2006). Believable agents and intelligent scenario direction for social and cultural leadership training. In *Proceedings of the 15th Conference on Behavior Representation in Modeling and Simulation*.

Riggio, R. E., Messamer, J., & Throckmorton, B. (1991). Social and academic intelligence: Conceptually distinct but overlapping constructs. *Personality and individual Differences*, *12*(7), 695-702.

Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and social Psychology*, *51*(3), 649.

Ruben, B. D. (1976). Assessing communication competency for intercultural adaptation. *Group and Organization Studies*, *1*(3), 334-354.

Samuel, B., Reed, A. A., Maddaloni, P., Mateas, M., & Wardrip-Fruin, N. The Ensemble Engine: Next-Generation Social Physics. In *Proceedings of the Tenth International Conference on the Foundations of Digital Games (FDG 2015)* (pp. 22-25).

Shapiro, D., Tanenbaum, K., McCoy, J., LeBron, L., Reynolds, C., Stern, A., ... & Coon, W. (2015, May). Composing Social Interactions via Social Games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (pp. 573-580). International Foundation for Autonomous Agents and Multiagent Systems.

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. *Serious games: Mechanisms and effects*, *2*, 295-321.

Sodowsky, G. R., Taffe, R. C., Gutkin, T. B., & Wise, S. L. (1994). Development of the Multicultural Counseling Inventory: A self-report measure of multicultural competencies. *Journal of Counseling Psychology*, *4*1(2), 137.

Taylor, G., & Sims, E. (2009). Developing believable interactive cultural characters for cross-cultural training. Proceedings of HCI International (pp. 282-291).

Treanor, M., McCoy, J., & Sullivan, A. (2015, September). Social Play in Non-Player Character Dialog. In Proceedings of the Social Believability in Games Workshop co-located with the *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE 2015)*.

Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational psychologist*, *30*(4), 173-187.