Coevolution of Theory and Data Analytics of Digital Game-Based Learning

By

Mario Manuel Martinez Garza

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Learning, Teaching, and Diversity

May, 2016

Nashville, Tennessee

Approved:

Douglas B. Clark, Ph. D.

Melissa S. Gresalfi, Ph. D.

Rogers P. Hall, Ph. D.

Daniel T. Levin, Ph. D.

To my family – passionate teachers, lifelong learners, and players of games.

ACKNOWLEDGEMENTS

This research was supported, in part, with funding from the National Science Foundation and the United States Department of Education. I am deeply grateful to my advisor, Dr. Doug Clark, for his generous mentorship, unwavering support, and constant energy and enthusiasm for my research. Dr. Clark's passion and unique perspective on learning, technology, and design have enabled and inspired my own work, and will continue to do so for many years to come. And also to the members of my dissertation committee, Dr. Melissa Gresalfi, Dr. Rogers Hall, and Dr. Daniel Levin, for their selfless contributions of time, insight, and talent. Without your expertise and dedication, this work, and myself, would doubtless be the poorer. I am also thankful to the faculty of the Department of Teaching and Learning of Vanderbilt University, and in particular to Dr. Richard Lehrer, Dr. Barbara Stengel, and those already mentioned, for their inspired work as professors and researchers and the opportunities they provided me to learn from them. One can never truly thank one's teachers – except, perhaps, by honoring their example, which is itself the work of a lifetime. In more pragmatic terms, I am grateful to The *R* Project for Statistical Computing and its global network of contributors, and to the many anonymous experts who participate in the Stack Exchange community, for the knowledge and resources, freely given to the world, that made this work feasible. And finally, to my beloved wife Brittany, for absolutely everything else.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Digital games are an influential and ubiquitous presence in the lives of young learners. A 2008 study by the Pew Internet and American Life Project found that 97% of American teens aged 12-17 play digital games, and 50% of them report daily or nearly daily play (Lenhart et al., 2008). With increasing access to computers, consoles, and cell phones, young people encounter opportunities for gaming everywhere. Educational researchers are increasingly interested in the affordances of digital games as a medium for learning. Investigation into the use of games for learning has grown from a small niche area to a major focus of research over the past decade (e.g., Young et al, 2010; Gee, 2003, 2007). A growing body of evidence indicates that digital games can be powerful vehicles for elementary and secondary learning in the STEM (science, technology, engineering, and mathematics) disciplines. Numerous studies have linked classroom use of learning games with increased learning outcomes and improvement in students' conceptual understanding, engagement, and self-efficacy in these fields (Martinez-Garza, Clark, & Nelson, 2013; de Freitas, 2006).

Realizing the full potential of games for learning poses significant challenges. Digital games for learning are remarkably diverse as a medium; they are designed with vastly different affordances and constraints, they target a wide range of age groups and content areas, they are deployed in a variety of educational settings, and lend themselves to a range of quantitative and qualitative research methodologies. Because of this variety, educational games research supports a number of theoretical lenses to explain the related processes of game design, play, and learning through play. In addition to more traditional perspectives on media-based learning effects (Mayer, 2009), scholarship into games can now be grounded in terms of the formation of communities of practice (Gee, 2004/2007; Lave & Wenger, 1991) and the creation of islands of expertise (Crowley & Jacob, 2002), the use of games as cultural and ideational resources for identity construction (Holland, Lachicotte, Skinner & Cain, 1998; Nasir & Cooks, 2009) and regimes of competence (Wenger, 1999), and the opportunities afforded by games for participation in shared enterprise (Wenger, 1998) and to connect with authentic practices (Barab et al, 2007; see Brown, Collins & Duguid, 1989).

These and other theoretical frameworks are all useful and valid means of examining and explaining educational gaming, yet they tend to focus on different processes and entities. One aspect of games for learning that has recently garnered attention from researchers is game play data, i.e. the automatically-collected record of student activity within the game environment. This record has the potential to provide a wealth of insight about how students learn (Romero & Ventura, 2012). However, the exhaustive and detailed nature of the data is both a source of promise and of great challenges. The richness of game play data makes it difficult to find meaning and draw causal lines between game play actions and learning outcomes. A useful method for research that aims at using game play data is statistical computing, known variously as "educational data mining" or "learning analytics" (LA). These techniques are concerned with developing, researching, and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist (Romero & Ventura, 2012; Berland, Baker, & Blikstein, 2014).

One feature of learning analytics as a research methodology is its intimate connection with the process of developing and refining learning theory. The goal of LA is discovering latent structure within a set of data, but this structure (if it exists) is not automatically meaningful or actionable. Thus, LA methods require interpretation by human experts who can situate the results of data mining within the appropriate context, determine whether the results are useful and actionable, and provide the necessary causal explanations – actions which are inherently theory-laden. Conversely, the results of LA can potentially prompt refinement or revision of the assumptions and hypothesis of the theoretical lens, by testing the connections between the various concepts and relationships described in the theory. For these reasons, learning theory and learning analytics can be said to *coevolve*, that is, to refine and improve each other reciprocally, each aspect providing a necessary element for the growth and advancement of the other.

In this three-paper dissertation, I explore this process of coevolution between learning theory and data analytics. From the theoretical side, I investigate how a general theory of cognition (the two-system or dual-system model) can be applied to game-based environments. A base hypothesis in this theory is that certain patterns of action in the game-space indicate cognitive action aligned with the two-system dynamic which I call stances. From the methodological side, I applied techniques of statistical computing that allow the detection of these stances as they are reified in a physics learning game, by demonstrating an analysis of the collected actions of players that can be applied in other educational gaming contexts in a reliable and comprehensive fashion.

The first paper consists of a theory-agnostic approach to game-play data analysis using learning analytics (Martinez-Garza, Clark, & Nelson, 2013). It was published in the International Journal of Gaming and Computer Mediated Simulations. This paper is an early exploration of game play data viewed as a record of students' thinking without commitment to a particular theoretical lens. A central claim in this paper is that game play is a manifestation of the player's underlying mental processes, and since knowledge of how to play a specific digital game is not innate, one of these mental processes must involve learning. When a game's challenges are aligned with learning goals, then a person's actions in-game can be conceptualized as process data, indicating a person's ongoing effort to understand the concepts and relationships encapsulated by the game. Then, if the actions that comprise these efforts are logged, we can analyze these logs for evidence of learning.

This paper summarizes my efforts and preliminary findings in the analysis of game play data from SURGE Classic (Clark, Nelson, Chang, Martinez-Garza, Slack, & D'Angelo, 2011). That game was designed in such a way that players had to apply frequent inputs; thus, a single player's game play log could contain tens of thousands of actions. More problematically, we observed that because of the genre and overall pace of the game, these inputs were often impulsive, tentative, or reactive. This made it difficult to claim that individual actions were significant or specifically indicative of the types of thinking and learning SURGE was designed to encourage. However, it was still possible that latent patterns recovered from a large collection of these individual actions could indicate changes in the trends of players' actions, which we could then diagnose as evidence of learning. In this paper, I described two machine learning algorithms (sequential pattern mining and hidden Markov modeling) that showed the most potential for recovering interesting patterns and trends, along with discussion of challenges involved in using them. Significantly, this early exploration lead to the insight that these techniques are not only predicated on the properties of game play data, but are necessarily theory-laden: the researcher must make some strong a priori assumptions regarding the interpretation and significance of the record; the data cannot "speak for itself". This insight prompted the development of a theoretical framework that could support these a priori assumptions required by LA.

The second paper is focused on developing and warranting such a theoretical framework. In Paper 2, I set forward a learning framework called the Two-Stance Model, or 2SM. The 2SM proposes two epistemic stances students might use when playing a digital game: a "learner" stance and a "player" stance. These stances are conceptualized as collections of epistemic resources (see Hammer & Elby, 2003) associated with the cognitive processes described in the two-system model of cognition (Evans, 2008). Two-system models of cognition distinguish between effortless thought, or “intuition”, and deliberate purposeful “reasoning”. These modes of cognition are neutrally labeled as System 1 and System 2, respectively. The former is described as fast, automatic, and associative; the latter as slower, deliberate, and self-aware. In the 2SM framework, System 1 is associated with the “player” stance and System 2 with the “learner” stance. These two stances guide different patterns of behavior within a game. A user in the learning stance might purposefully investigate the game in search of information that confirms or disconfirms his or her understanding. A user in the player stance might engage in developing and adopting effective control strategies, selecting proper actions, and repeating actions most likely to lead to desired results. This paper was defended as my major area paper in March 2014, revised in June 2014, and presented it as a poster at the 11th International Conference of the Learning Sciences. A revised version was submitted to Educational Psychologist, where it received a decision of revise and resubmit. Part of the reviewers’ feedback focused on the reasonable observation that most, if not all, of the claims of the 2SM were extrapolated from earlier findings but not yet specifically proven. The empirical investigation of the main claim of the 2SM was the subject of Paper 3.

In this third paper, I applied LA techniques to investigate the patterns in students' play of *The Fuzzy Chronicles*, and how these patterns relate to learning outcomes with regards to

Newtonian kinematics. This paper focused on two research questions, and each was supported by its own analysis. The first research question examined whether students playing The Fuzzy Chronicles showed evidence of dichotomous fast/slow modes of solution. The 2SM theorized that slow modes of solution would correlate to higher learning gains. In order to discover the existence and features of these modes of solution, I used an affinity propagation clustering algorithm that revealed similarities within six clusters of data points. These clusters described typical moments in EPIGAME game play and were assigned to one of six descriptive codes. Then, using pattern mining techniques I found which sequences of codes are common across students of similar achievement levels. These sequences represent more extended episodes of a student's play that span more than one attempt. As predicted in the 2SM, students who use mainly fast iterative solution strategies achieved lower learning gains than students who preferred slow, elaborated solutions, or a more balanced mix of the two.

The second research question investigated the connection between conceptual understanding and student performance in conceptually-laden challenges. Each of these challenge is a situation on the game map where a student has to apply one or two maneuvers to advance past that situation. The challenges that were selected embodied situations were inertia and/or Newton's second law of motion were most relevant. I hypothesized that students would perform better and commit fewer errors traceable to conceptual understanding as play progressed. The finding was that students do generally improve their performance in these challenges, but that this improvement is strongly moderated by their prior knowledge of physics.

Both findings in this paper suggest that, while still very much unproven as a whole, the basic underpinnings of the 2SM pass muster. I found evidence of both fast, low-information play and slow, deliberative play. More importantly, these styles of play co-varied strongly with learning outcomes, as predicted by the 2SM. However, the results of the learning analytics also warrant a revision of the 2SM to better account for the role of prior knowledge in helping students organize their stances, and the issue of whether the “player” stance is a control-oriented epistemic stance or a coping strategy to deal with game situations students find too difficult. Neither of these considerations were part of the original framing of the 2SM, and yet they demonstrate the process of co-evolution; namely, that even when the 2SM was crucial to shaping the learning analytic methodology, the results of the LA process were necessary to refine the theory.

As a whole, these three papers represent my contribution to what I believe to be a promising new area of game-based learning research. The study of educational gaming will continue to benefit from richer, more extensive descriptions of game play than are feasible with traditional observational methods. A single game-play action that appeared uninterpretable when viewed in isolation, or as part of an undifferentiated aggregate, may reveal more about the learner when examined as part of a cluster of actions within a specific context, or an evolving sequence of actions, which in turn may be abstracted into something like a plan, a strategy, or a gameplay style. In these papers, I demonstrate how this analysis might be performed, the theoretical work required to give the analysis meaning and applicability, and the kinds of insights into students’ learning through play that are possible when learning theory and learning analytics are refined together and co-evolve.

References

Barab, S., Zuiker, S., Warren, S., Hickey, D., Ingram-Goble, A., Kwon, E. J., ... & Herring, S. C. (2007). Situationally embodied curriculum: Relating formalisms and contexts. *Science Education*, *91*(5), 750-782.

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: applications to constructionist research. *Technology, Knowledge and Learning*, *19*(1-2), 205–220. http://doi.org/10.1007/s10758-014-9223-7

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational researcher*, *18*(1), 32-42.

Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M. M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education*, *57*(3), 2178–2195. http://doi.org/16/j.compedu.2011.05.007

Crowley, K., and Jacobs, M. (2002). Islands of expertise and the development of family scientific literacy. In G. Leinhardt, K. Crowley, and K. Knutson (Eds.), *Learning conversations in museums* (pp. 333-356). Mahwah, NJ: Lawrence Erlbaum Associates.

de Freitas, S. (2006). Using games and simulations for supporting learning. *Learning, Media and Technology*, *31*(4), 343–358.

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. http://doi.org/10.1146/annurev.psych.59.103006.093629

Gee, J. P. (2007). *What Video Games Have to Teach Us About Learning and Literacy. Second Edition: Revised and Updated Edition* (2nd ed.). New York, NY: Palgrave Macmillan.

Hammer, D., & Elby, A. (2003). Tapping Epistemological Resources for Learning Physics. *The Journal of the Learning Sciences*, *12*(1), 53–90.

Holland, D., Lachicotte, W., Skinner, D., & Cain, C. (1998). *Agency and identity in cultural worlds*. Cambridge, MA: Harvard University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge University Press.

Lenhart, A., Kahne, J., Middaugh, E., Macgill, A., Evans, C., & Vitak, J. (2008). *Teens, video games, and civics* (p. 64). Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/Reports/2008/Teens-Video-Games-and-Civics.aspx

Martinez-Garza, M. M., Clark, D., & Nelson, B. (2013). Advances in Assessment of Students' Intuitive Understanding of Physics through Gameplay Data. *International Journal of*

Gaming and Computer-Mediated Simulations (IJGCMS), *5*(4), 1–16.

Mayer, R. E. (2009). *Multimedia learning*. New York: Cambridge University Press.

Nasir, N. I. S., & Cooks, J. (2009). Becoming a hurdler: How learning settings afford identities. *Anthropology & Education Quarterly*, *40*(1), 41-61.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27.

Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems thinker*, *9*(5), 2-3.

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., … Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research*, *82*(1), 61–89. http://doi.org/10.3102/0034654312436980

CHAPTER II

ADVANCES IN ASSESSMENT OF STUDENTS' INTUITIVE UNDERSTANDING OF PHYSICS THROUGH GAMEPLAY DATA

Introduction

The goal of this paper is to describe the development of novel approaches to assessment of learning in games. Often, researchers that design experiments around games for learning must rely on post hoc instruments to measure the progress that students make. This approach, while necessary and fruitful, does not leverage the potentially rich store of evidence that students provide about their own learning while they play. While the means exist to collect complete records of the actions and decisions that learners make while they play, no widely-accepted techniques or tools for making sense of this data stream currently exist. In this paper, we present our initial forays into analysis of game play using data-driven statistical and visualization techniques, and provide examples using data from the SURGE project (Clark, Nelson, Chang, Martinez-Garza, Slack, & D'Angelo, 2011). Furthermore, we provide a rationale and framework that for the use of these techniques and argue for their appropriateness and applicability in other contexts of educational games research.

Theoretical framework

The potential of video games to support science learning is generally agreed upon (Gee, 2007; Mayo, 2009; Squire et al., 2003), but the analysis and structuring of evidence for game-based learning remains a challenge. This, in turn, has supported a mixed view of the effectiveness of games as tools for learning (Foster & Mishra, 2008; O'Neil, Wainess, & Baker, 2005). We believe, however, that this conclusion may be premature. The past fifteen years have seen great advances both in the sophistication of game designs and also in the supporting technology; there simply has not been enough time for a commensurate evolution in appropriate research methods. One central methodological difficulty involves capturing and measuring game-induced learning, which tends to be strongly situated within the game context, in out-of-game contexts such as post-tests. More advanced game designs compound this problem by supporting complex player actions that are challenging for learners to summarize and express, difficult for instruments to reliably capture, and resistant to conventional analytical methods. In addition, the use of formal assessments alongside games can compromise a game's capacity for engagement and immersion, thus potentially reducing the efficacy of both the learning experience and the assessment.

The use of assessments of learning which reside outside a game used to measure learning that happens inside a game presents issues and vulnerabilities that merit careful consideration. Assessment is, after all, not a neutral activity. All assessments carry assumptions about the nature of learning, the nature of knowledge, and the purpose of assessment itself (Willis, 1993). The action of assessment places premiums on certain forms of knowing and understanding while de-emphasizing others. In the case of games for learning science, for example, an assessment may privilege declarative forms of knowledge, e.g. definitions and abstract principles, while the game itself might be more productive in reinforcing tacit knowledge or qualitative understanding of relationships. This insight becomes even more salient given the contrast between different types of games for learning: those in which the curriculum concepts are embedded in the game environment in a manner such that the game environment is structured mainly as context ("conceptually-embedded" games) and those in which the material to be learned is integrated into the core game-play mechanics with which the player is in constant interaction ("conceptually-integrated" games) (Clark & Martinez-Garza, 2012). It follows that these two kinds of games would favor different assessment strategies, given the differences in how they engage the learner, how they gauge success in the game, and how they represent knowledge. These nuances are not necessarily well captured by traditional assessments of learning, which traditionally favor summative declarations of concepts, articulated in discipline-specific forms and language (Sutton, 1996; Fang, Lamme & Pringle, 2010).

Other researchers have expressed similar views about the shortcomings of external assessments in capturing learning that happens in games and interactive media settings in general. De Jong & Van Joolingen (1998) argued that one of the difficulties that research has in achieving unequivocal findings in favor of the learning outcomes of unstructured learning environments lies in data interpretation. Quite often in educational games studies, researchers rely exclusively on outcome measures because there are simply no developed frameworks for interpreting and evaluating process data. This resonates with de Freitas (2004, 2005), who proposed that games for learning require more specialized frameworks and methods of evaluation given the gaps between the educational content and the context in which games are used. An important component to these frameworks is a more robust understanding of *what students actually do* when they are playing a video game, what competencies they present, and how these competencies change in response to increasing difficulties and challenges. In service of this goal, Shute has proposed an approach called stealth assessment (Shute, Rieber, & Van Eck, 2011; Shute, Ventura, Bauer, & Zapata-Rivera, 2009), which regards the rich sequences of actions produced by students who are playing a game as indicative of the skills that researchers care to assess, and “evidence for learning is thus provided by the students’ interactions with the game itself – the processes of play, which may be contrasted with the product(s) of an activity, as is the norm within educational settings” (Shute et al, 2009, p. 300). In this perspective, game play is a direct, faithful representation of the learning processes that are taking place, and thus may be less vulnerable to the contextual and design weaknesses of external assessments. Several questions arise. How can we make sense of this data? What phenomena does it encode? How can it be constructed into evidence of learning?

The nature of game play data

This paper explores what game play data might indicate and how it might be analyzed. We propose that game play is a manifestation of the player's underlying mental processes, and that like other such processes, these can be described and studied with the appropriate analytical lens. Game play is not random or purposeless; all good game designs contain compelling mechanisms that keep players cognitively engaged towards the completion of goals within the game. These mechanisms create challenging situations that constrain a player's actions but also focus them towards a goal, much like a problem-solving activity. When the game's challenges are aligned with its educational purposes, the player's purposeful experimentation, information-seeking, and problem-solving choices drive the actions they make in the game. Our goal is to track the actions that indicate that those decisions took place and then to reconstruct (inasmuch as possible) the latent learning activity of the player in service of the learning goal.

It may be argued that no assessment methodology that analyzes these actions can accurately reflect learning because not every action a player takes in a game is purposeful and reflective. It is certainly true that game play can support actions that are experimental, tentative, or counterproductive. Players may take wrong turns and choose incorrect options despite 'knowing better'. However, it is precisely this "fuzzy" nature that makes game play data so productive, and in many ways, it also makes it construable as process data.

Another aspect of this approach which may be seen as problematic is the matter of grain size. In our view, the most complete and detailed record of a learner's actions in a game will be the most productive in terms of detecting learning trends. However, it may be the case that individual actions (at the level of mouse-clicks, button-presses, view-point changes, etc.) do not directly code for any particular aspect of learning. It may indeed be a flawed approach to overload meaning into individual moments of gameplay, whose duration is often measured in fractions of a second. Our approach instead involves collecting longitudinal slices of game play and properly structuring and examining these slices at an appropriate time scale. Thus, no specific claim is made about any action that the player takes, but rather, we look for trends and changes in decisions and outcomes that we reconstruct as learning.

Productive Uses of Game Play data in other contexts

The value of game play data is well-understood in the realm of commercial game design. By recording, organizing, and analyzing the actions of players within a game environment, designers glean important insights into the strengths and weaknesses of the game. Game play data is especially useful in bridging the gap between design goals and actual use, as even the most straightforward design can have unforeseen limitations, difficulties, or possible distortions that do not become apparent until others outside the design process (i.e., testers or final users) engage with the game.

The emergence of networked technologies and server-based massively-multiplayer game environments have allowed designers to seamlessly collect large banks of game play data, with a corresponding push towards developing techniques for analyzing and presenting this data. An example of this was the debate surrounding the 'Valhalla' map of the multiplayer game *HALO 3*. During the testing process for this game, the community began to suspect that this map was unbalanced, with players defending the one of the two bases being at a disadvantage. The players who were advocating this position, however, could not produce compelling evidence. After creating a visualization of the game play data, the designers of *HALO 3* were able to confirm the systematic bias of the 'Valhalla' map and use the information conveyed by the visualization to rebalance the map in time for the final release (Thompson, 2007).

We would argue that this example falls short of accurately representing the potential for assessment based on game play data. Because the 'Valhalla' case, among others, focuses the lens of assessment on the game, not on the player, it does not provide answers to questions related to player abilities, capacities, or learning. However, we consider this example a good demonstration of the type of questions that can be answered by game data, and the kind of technology, practices, and techniques required to use this kind of data systematically as evidence. Namely, that gameplay data should be collected automatically and centrally; that the data should capture as much information as possible about salient events; that this information should be stored in indexable formats that are amiable to computational methods of statistical analysis.

Context: The SURGE Project

SURGE is a conceptually-integrated computer game designed to support students learning about force and motion. More specifically, our design efforts for *SURGE* have focused on helping students articulate their intuitive understandings about Newtonian mechanics. The SURGE project integrates research on conceptual change, cognitive processing-based design, and socio-cognitive scripting with design principles and mechanics of popular commercial video games. The game is designed to support students' articulation and connection of their evolving tacit intuitive understandings into larger explicit formalized structures, thus allowing knowledge transfer and application across broader contexts relevant to Newtonian mechanics.

Gameplay in SURGE is organized into short intervals (“levels”) that are designed so that concepts taught through gameplay build upon one another and gradually introduce the student to new ideas and ways of interacting with the world of the game. Each level involves specific navigation challenges, in the form of turns, starts and stops, that are presented to students in the form of a maze. To navigate a maze, students need to apply a sequence of moves (roughly 5-20 per level) that reflect the principles of Newtonian mechanics (impulse, inertia, vector addition, velocity, acceleration, etc.). Levels are intended to be played in 1-2 minutes; however, because players can control their speed of navigation, the actual duration of each played level is highly variable across students, levels, and attempts.

The game is presented as a space-based adventure, where students play as the character Surge, a smart and brave female alien, who must rescue cute creatures (called Fuzzies) from captivity. Students use the arrow keys on their computer keyboards to navigate Surge’s spherical spaceship around barriers and through corridors as safely and efficiently as possible. If Surge’s ship collides against the walls of the maze, it will turn a shade of red as a warning to the player; too many collisions and the ship will explode, and the player must restart the level. Overlaid on the computer screen are different read-outs of gameplay information for the student, including their current speed, the number of impulses they've used, the number of collisions with the walls, and their elapsed time. There are also buttons to reset or pause the level and to stabilize Surge’s ship if it begins to move out of control. A vector representation of students' velocity is also on the screen, showing their current speed and direction.

The data used for the present studies originates from a *SURGE* module that features an "impulse control system, where every time the student pushes an arrow key a fixed impulse is applied to Surge in the direction of the arrow key pressed. Students in the study were told to minimize their collisions, elapsed time, and number of impulses applied in a level in order to get a high score.

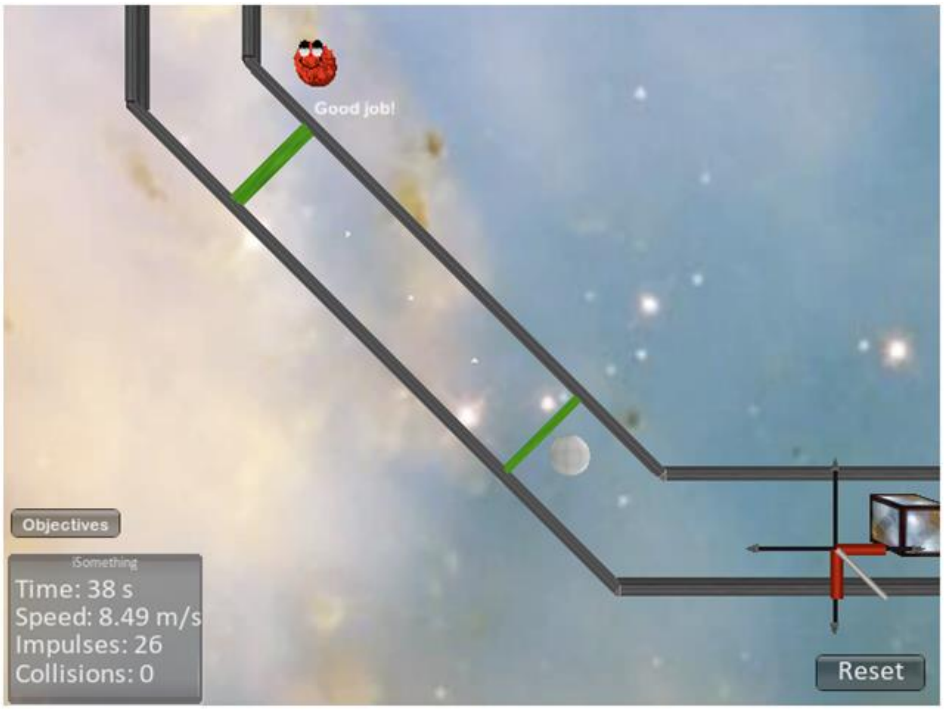

Figure 1. Screenshot from SURGE.

Experimental results with students have shown that a conceptually-integrated game such as *SURGE* is generally successful at producing learning gains (Clark, Nelson, Chang, Martinez-Garza, Slack, D'Angelo, 2011; Clark, Nelson, D'Angelo, Martinez-Garza, & Slack, 2010). However, the link between these gains and specific gameplay events was not clear. We observed that as students became more proficient at playing the game, they enacted through their play a more sophisticated intuitive understanding of momentum, but could not consistently express that new understanding in the framework of a physics test. We have suggested that learners may have difficulty articulating concepts gleaned from gameplay into scientific terms, and thus their ability to accurately represent their knowledge on a traditional post-play assessment may be compromised (Clark, Nelson, D'Angelo, Slack, & Martinez-Garza, 2010).

Challenges of Assessment of Learning in SURGE

To assess learning gains, our research with SURGE has focused primarily on conventional pre- and post-test designs. Our primary assessment in SURGE has been a multiple-choice conceptual test based on items from the Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992), one of the most widely recognized conceptually-focused physics tests of force and motion. In our studies with three 6-8th grade sciences classes, students who play SURGE make progress on the FCI-based test (Table 1), which is promising, considering the short duration of our intervention and that the FCI was developed for assessing learning across a full semester of undergraduate physics.

Table 1.
Summary of assessment results of SURGE experimental groups

Description of group	N	Gain (in SD)	*p* <	Assessment used
Taiwan middle school	180	0.38	0.008	Selections from FCI: kinematics and Newton's 1st Law
Diverse urban school in Southern US	71	0.31	0.114	As above
Title I urban middle school in Southern US with 19% IEP	69	0.48	0.020	Selected FCI kinematics items adapted to 6th grade reading level
Undergraduates in a calculus-based physics course*	155	1.27	0.001	Selected FCI kinematics items with additional vector representations

Note: Undergraduates experimental group included for comparison only.

However, this assessment strategy has proven suboptimal for several reasons. While students in the studies were highly engaged while playing SURGE, we observed that the shift from a "game-playing" frame to a "test-taking" one was jarring to many of the students, causing adverse reactions that may have reduced the quality of our data (as the players often skipped through the test to return to the game). Also, the pre-post approach provided no information about learning trajectories during the game to help us understand how aspects of the game contributed to learning.

The perceived deficits in our assessment strategy led us to consider other assessment approaches, specifically those that would allow us to minimize, or entirely avoid, the problems associated with our initial pre-post design in which the assessment is an on-screen rendering of a paper test. Along with further refinement of our conventional test to increase sensitivity to learning by including a broader range of item difficulties within the instrument, we launched efforts aimed at using gameplay data, generated by learners within the context our game, as an avenue of assessment that could potentially allow valid inferences about learning outcomes (Shute, Rieber, & Van Eck, in press).

Analyzing Game Play Data within SURGE

Having studied the various examples of analytic techniques based on game play data from the realm of commercial game design, we launched our own effort at using data of a similar nature for assessment of learning. Certain design decisions during our initial programming phases made our efforts to assess game play possible. The SURGE software tracks the user actions, salient events, and game states (such as the position and velocity of the player object multiple times per second) that are later automatically sent to and stored in the SURGE database. We found that with some recasting and normalization of the data (e.g. removing incomplete or duplicate records), this data could not only be examined as a whole, but also at multiple component and aggregate grain sizes that could be to used for various analyses.

As a first analytical pass, we created various visualizations from the gameplay data using the free graphing application, *ploticus*, to identify the most promising commonalities, patterns and sequences for examination. These visualizations gave us some concrete artifacts of the learner's play that afforded a valuable window of insight into the nature of students' decisions and concepts during play (Figures 3 & 4). Our initial examination of these visualizations prompted discussion about how to characterize students' actions while preserving validity. What follows is an example of the type of qualitative analysis that these kinds of visualizations can support.

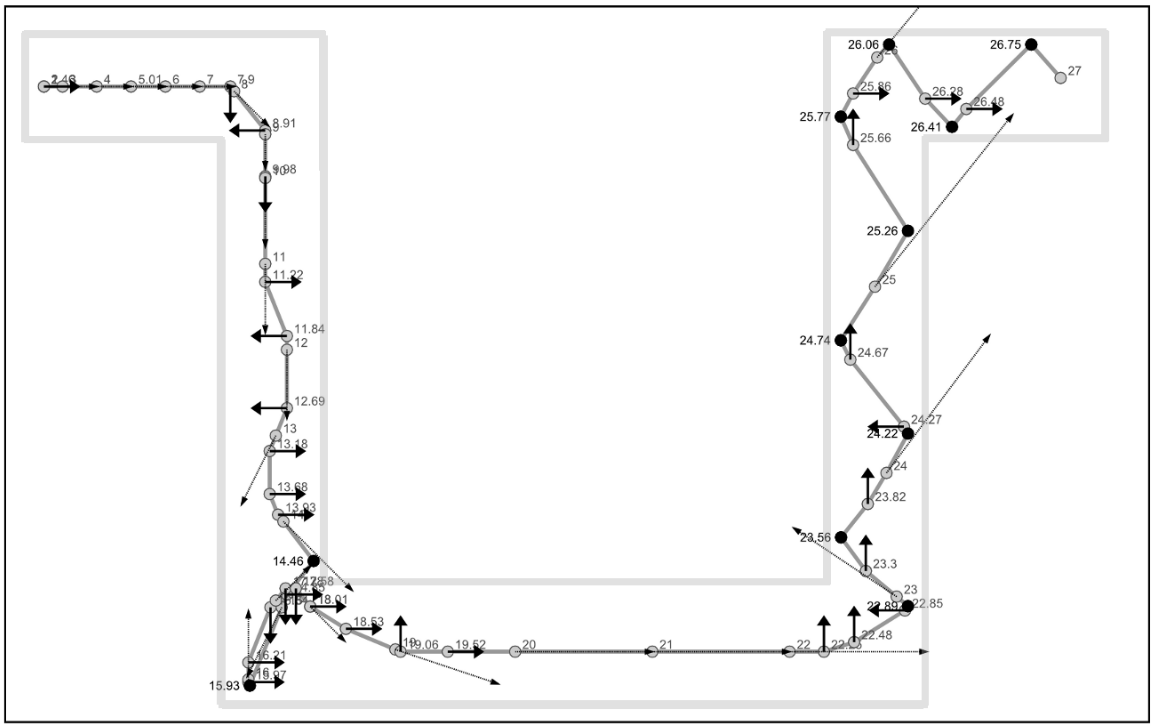

Figure 2. Visualization of one student's path through a level.

Figure 3 shows a visualization of one student's path through the first level of the SURGE game, specifically of the path they followed when they completed the level successfully for the first time. Grey circles represent the position of the player ship at N seconds after starting the level. Short black arrows represent player inputs. Thin grey arrows represent the direction of the ship at one second intervals. Black circles represent collisions of the ship. Note the "chain reaction" of collisions that occur at the right-hand region, starting at 23 seconds, from which the player never quite recovers. This seems to be the moment of departure, where the student's intuitive understanding is not up to the challenge provided by the game. We therefore examine that instant of gameplay more carefully (Figure 4).

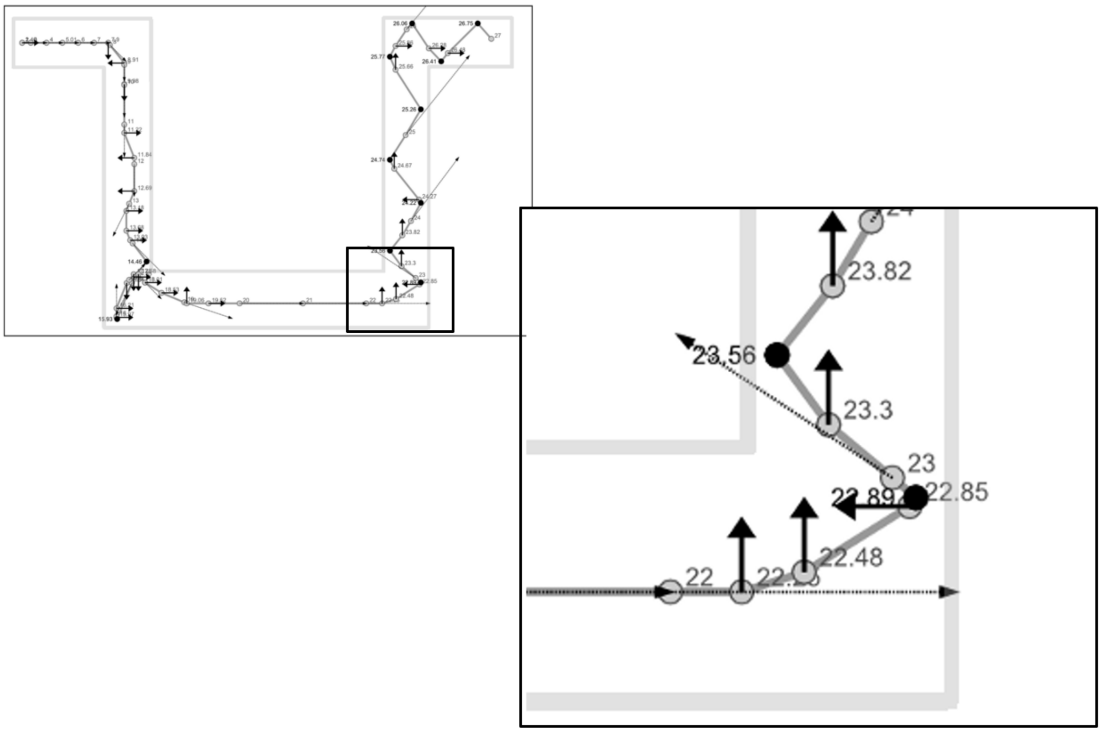

Figure 3. Detail of the "moment of departure" in the previous visualization.

A step-by-step analysis of the moves and events represented in this visualization allows us to create a reasonably detailed interpretation that may account for the player's chosen moves, and may allow insight into the player's intentions and understandings.

- At (21), the player is steadily moving right.
- After (22), the player apparently decides she must turn upwards.
- (22.25) An "UP" impulse is applied.
- (22.48) Another "UP" impulse. We can see that the player has not accounted for inertia, suggesting that their intuitive model may not be compatible with Newtonian ideas of inertia in this context. It is probable that the player expects Surge's ship to move "like a car" towards the direction it is pointed.

- (22.85) The player applies a "LEFT" impulse. This move is the correct one if the player wishes to arrest the rightward motion of Surge. Here it can be seen as a last-ditch correction to avoid colliding with the rightmost wall; although it is effective at partially correcting the course, this is a reactive move that may or may not evidence an understanding of inertia.
- (22.89) A collision, the natural consequence of the player's chosen combination of moves. A different combination, i.e. one that contained a "LEFT" move, would have prevented this collision.
- (23) The player object is moving fast upwards and to the left. Note the long dotted arrow indicating the current velocity.
- (22.3) The player applies another "UP" impulse, attempting to correct their direction. Again, the player does not account for inertia, and apparently expects the path to "right itself" in the way of a skidding car if enough "UP" impulses are applied without needing to account for horizontal inertia.
- (23.56) A second collision, caused by an ineffective adjustment in (22.3) … etc

In summary, the student does not apply a working model of inertia to inform her gameplay, and the result is a chain of collisions from which the player does not recover.

How common was this pattern? This type of analysis can be very generative and provides a comprehensive view of gameplay actions and decisions; the next logical step was to create a systematic classification scheme to allow comparison and aggregation. To investigate the prevalence of this result, we created a visualization that contained aggregate results from the entire data set of players for this level, not just a single player (Figure 5). This graph shows the locations at which players collided with the walls of level the first level of Surge, selecting for the first attempt in which each student successfully completed the level, i.e. the same data selection criteria as we used to generate the visualization in the previous case.

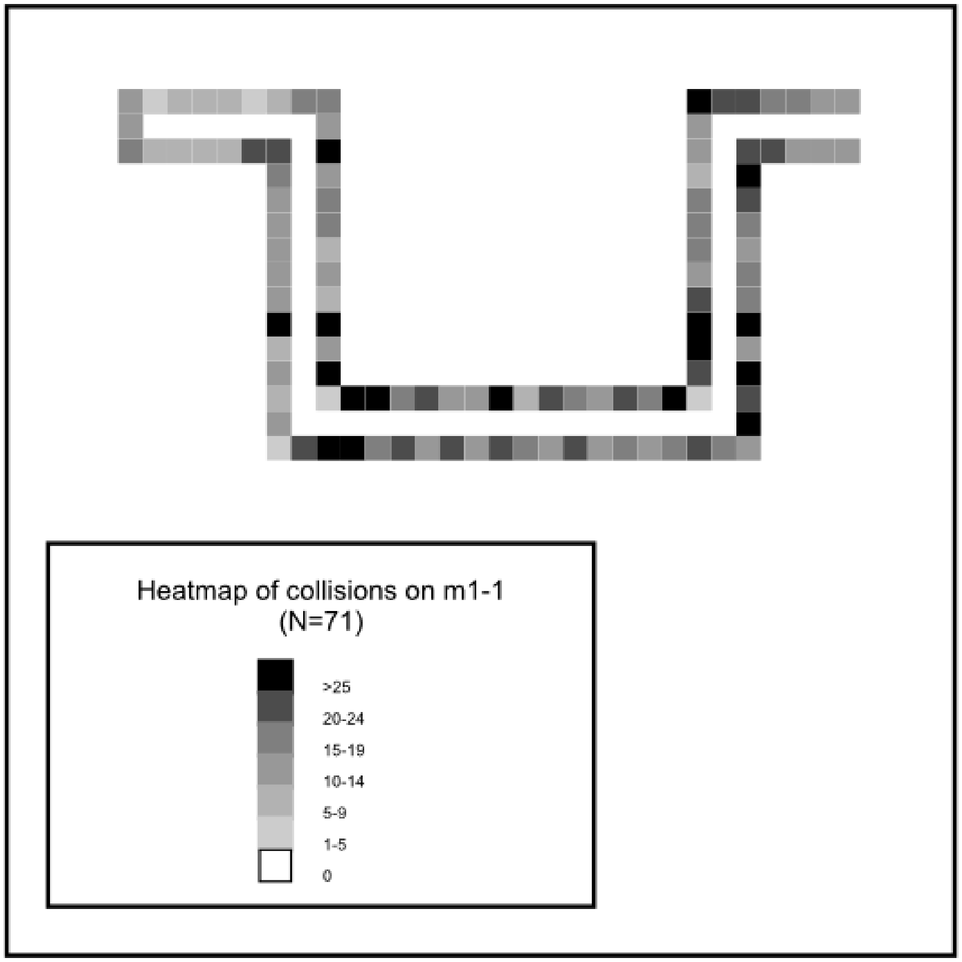

Figure 4. “Heat map” visualization of collisions.

In this graph, we can see that the “chain reaction” of collisions was a common phenomenon; students were not just making the same error, but were making it in the same moment of gameplay. This result encouraged us to seek further regularities and patterns of action and response, but Figure 5 also highlights limitations of this simple aggregate representation. One key limitation is that temporal and spatial variations between the paths of individual students occlude higher level similarities in patterns between students in the aggregate, e.g., while many students display the high level pattern described in the example in Figure 3, the exact points of collision and application of thrusts vary enough that an aggregate representation of the points of collision in the level would imply a great variation in strategy and decision-making from one student to another, whereas in reality, the detail discussed in the bullets about the events in Figures 3 and 4 may hold true for most cases. The converse may also be true; an aggregate representation could result in a clustering of game play actions that, in fact, resulted from very diverse forms of students' thinking. We thus needed an approach to aggregation that could capture and process the patterns of play in richer manner, but given the scope of the data collected, qualitative coding of all students seemed unworkable. In light of this, we are now focusing on creating and testing computational data analysis routines to aid in interpreting the actions of *SURGE* players in a manner that captured larger patterns in their game play.

Sequential Pattern Analysis

The structure of the *SURGE* data initially suggested that we could detect patterns of actions that were common across the dataset or within certain groups. A single student's gameplay log can be seen as a chain of sequences, where each sequence corresponds to a challenging point in one of the game's levels. It is during these instants where players must decide, based on their understanding, what is the most appropriate "move" to make in order to successfully navigate through that area of the level. Thus, it can be said that each chain carries a solution exercised by the player to a conceptual problem posed by the game. Given these premises, we initially concluded that a sequential pattern analysis might be appropriate for this type of data.

Sequential pattern analysis (SPA) is a database-driven analytical technique that allows the identification of coherent actions from a complex series of recorded events (Agrawal & Srikant, 1995), and has been applied successfully in educational software contexts (Nesbit, Xu, Winne, & Zhou, 2008; Zhou, Xu, Nesbit, & Winne, 2010). In essence, SPA consists of series of algorithms that, when given a sequenced input data, can find sub-sequences that occur in a proportion significantly greater than chance. These sub-sequences in *SURGE* involve common solutions for situations present in the game, which can then be pooled to produce response probabilities for the experimental group as a whole. Sequential pattern analyses require that researchers specify a certain set of sub-sequences that can be matched, in effect isolating them from any surrounding random actions. In effect, SPA can be thought of as a computerized coding process that is much more flexible and robust than simple automated pattern-matching. Figure 5 outlines the specific steps involved in applying SPA to SURGE gameplay data.

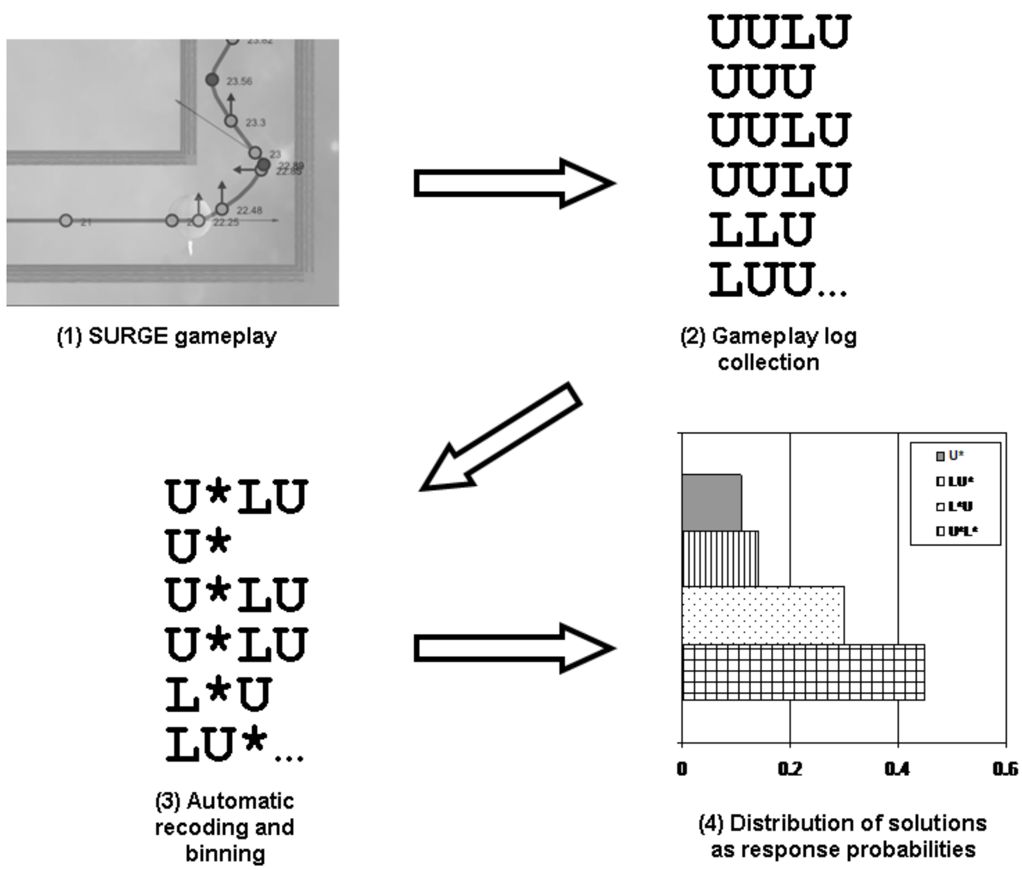

Figure 5. Overview of computerized data analysis of SURGE data using SPA.

For example, suppose a player in SURGE is facing the situation illustrated in the inset of Figure 3. The player object is moving in the “right” direction and is coming up on an “up” turn.

The moves that a player may take in this kind of game scenario consist of applications of force in the cardinal directions, up/down/left/right (henceforth, U/D/L/R). Three common response patterns that we could expect would be: UL, LU, U, and UU. The first two responses indicate an emergent understanding of inertia; the student has accounted for the upward motion of the player object, cancelling it with an L move. The latter two indicate incomplete understanding, as the learner takes no action to stop his upward motion, expecting the U move to "carry" the player object. We can then collect and bin the responses that are functionally equivalent, e.g. LLUU, LULU, LLU, LU, etc., and code for move efficacy value (see Figure 5, step 3), then calculate the aggregate response value for each gameplay event (step 4). If this analysis were applied across the full game for all of the students in a dataset, it would be possible to trace the probability of correct and incorrect response; an increase in correct patterns of response from the beginning to the end of the game could then be constructed as evidence of learning.

However, as we piloted SPA protocols it became apparent that, although SPA has the potential to be used to make sense of gameplay data in general, it would not be specifically appropriate as an analytic tool for *SURGE* data. The reasons are twofold. First, and most importantly, *SURGE* is a game of continuous action and the timing of moves is a crucial part of gameplay. Since SPA collapses sequences in the time dimension, no inference can be made from the different arrangement of moves in time, which, among other things, signals the difference between thoughtful planning and simple reaction. Second, because of the pooling of sub-sequences that is required, SPA would not provide information on individual trajectories of learning, and thus our ability to reach one of our longer-term assessment goals (namely, to scaffold learning via adaptive feedback processes akin to formative assessment) would be compromised. Furthermore, it was unclear from our preliminary experiments whether SPA algorithms would correctly account for cases of non-linear play (e.g., retrying or restarting a level), which to a certain degree became unavoidable as we built more and more game-like structures into *SURGE*.

Hidden Markov Modeling

In order to account for the challenges we encountered with SPA, we are now focusing our efforts toward a more sophisticated technique called Hidden Markov Modeling (HMM). HMM is a clustering methodology that shows promise in terms of its ability to properly account for the information carried by different patterns of gameplay sequences and its ability to operate in real time. Once limited in application to tasks such as signal processing (Rabiner, 1989), Hidden Markov Modeling has since been used in a wide range of areas, such as speech recognition and bioinformatics. In the field of education, HMM has been applied to the study of collaborative learning (e.g. Soller, Wiebe, & Lesgold, 2002; Soller & Lesgold, 2007), and for assessing and modeling student learning (e.g. Stevens, Johnson & Soller, 2005). The potential of HMM in analyzing gameplay data has not yet been extensively explored, although it has been successfully demonstrated in a similar context: the learning-by-teaching environment *Betty's Brain* (Jeong et al., 2008).

To clarify the justification for using HMM methodology for analyzing gameplay data, let us consider a hypothetical example. Suppose that a researcher is investigating the effect of a learning game. We imagine that at any given time point, a player has six available actions *A*, *B*, *C*, *D*, *E*, and *F*. These actions can correspond to different types of interactions, e.g. menu selections, placement of elements on the screen, travel to specific locations in the game world, etc. The gameplay data for this experiment would then consist of player/action/time triads. One approach to analyzing this type of data is to seed the game design with a specific action that codes directly for the target learning goal. In our example, let that action be *F*, which by itself, can only tell whether the learner succeeded or failed. This action *F* would usually be a final step in a play session, but it also might be somewhere in the middle in a failed or sub-optimal state. The logs for a particular player might or might not include *F*, along with many other actions *A* through *E* in sequences and proportions that may look like a random noise. A more traditional analytical technique would be to decide, solely based on the presence or absence of *F*, whether or not learning took place. However, this approach leaves several critical questions unaddressed. What specific actions or parts of the game design were most strongly connected to the learning outcome? In what sequence were they most productive?

To answer these questions, we would have to know two things: whether there was any regularity or structure in the sequences of gameplay, and whether or not this regularity correlated with the desired learning outcome. In terms of the former, it is helpful to remember that the actions of the learner are not random; they are organized towards a specific goal, which we assume to be the game's "victory condition". All of a player's actions are in service of this goal, but they do not proceed along a pre-determined sequence. Thus, there is some latent structure in the data that is provided by the game's goals, even when there may be much variability in the sequences of actions that players perform. In this regard, HMM analysis proves useful. The central assumption of HMM is that each sequence of responses is derived stochastically from one or more underlying states (Jeong et al., 2008). Each of these states is linked in a "Markov chain", namely, a discrete event-space where outcomes depend not on previous trajectories but only on the current states of the system. This is very useful for our purposes, since we cannot assume that a learner proceeds from action *A* to *B* to *C* and so forth until reaching the end-step *F*. Many trajectories are expected, but they will tend to cluster in some broad patterns or states. The transitions between these states, from one pattern of game action to another, can be interpreted as learning. HMM is furthermore a modeling technique that can portray a clear account of how changes occur.

A further advantage of HMM is that it can account for the differing pace at which learners play. A novice gamer might be very tentative in his play, while a thoughtful gamer might have a more deliberate style, and an expert gamer might be more inclined to take actions in quick, sudden bursts. Clearly, gameplay has a temporal texture which may modify, or mediate, the effect of certain sequence of actions. While several alternative cluster analysis methodologies can discover patterns in data in a static form, HMM can analyze data with a temporal dimension (i.e., data whose feature values change during the observation period) (Li & G. Biswas, 2002). Thus, we believe that HMM modeling is well-suited to applications within games for learning, where the dynamic features of students' play can vary not only from student to student but also over time as learners become proficient in the game and the operating principles behind it.

As in the case of SPA, HMM requires that the data be segmented into contexts (i.e., sections of game play that can be coded under the same criteria). In the case of SURGE, these contexts can be seen as corners or turns where the player must apply a combination of impulses to proceed. As with SPA, each of these contexts is encoded and binned to produce a response probability, but instead of comparing these responses by individual or by context, an algorithm is used to find the best fitting set of transitional (z) and output probabilities (a through d) that produce these probabilities from the various underlying states (Figure 7). As these transitional probabilities connect to underlying models that are assumed to evolve from one to another as part of a learning progression (expressed as z1 and z3 in Figure 7), then the emergence of strong values for these probabilities, combined with weak values for the converse (z2) could be interpreted as evidence of learning.

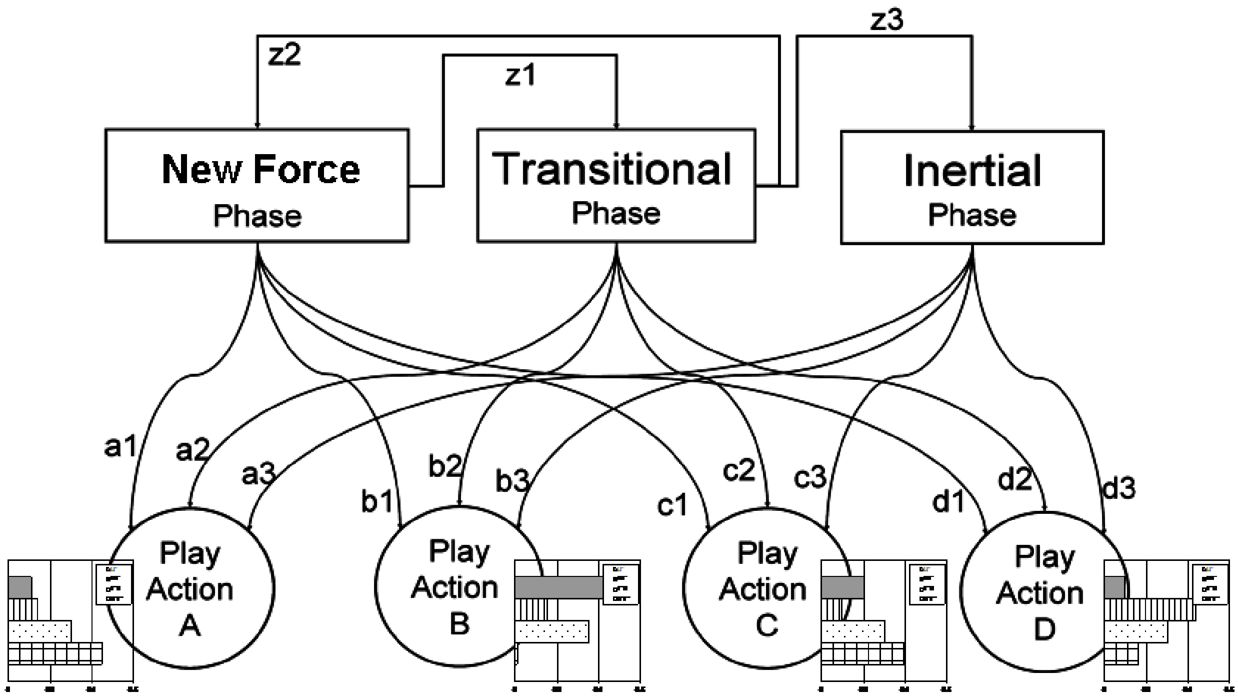

Figure 6. Schematic of the relationships between player actions and underlying states in HMM analysis.

This use of Hidden Markov modeling has so far presented several challenges. As with other unsupervised learning analyses, HMM works best with a comparatively large data set in order to enable initial training of the model, so it requires prior piloting and reliable data-gathering on a large scale. It is also computationally demanding and requires specialized technical know-how to implement because HMM software tools are generally not user-friendly. Furthermore, a typical HMM model requires that the researcher pre-determine the number of hidden states, which requires not only content knowledge but also a theory-guided stance about the form and nature of the conceptual change processes occurring in the learner. In the case of SURGE, these hidden states could be framed as forms of intuitive understanding concerning the operation of the laws of physics that inform the players' actions in each context. The validity of these forms and the connections between them defined *a priori* could be considered arguable.

Also, as in factor analysis, the researcher must also provide meaning to the states that the HMM model reveals. Thus, HMM-based analysis of game-based learning activities requires strong theoretical assumptions about the learning happening through these activities. In response to this difficulty, we are currently implementing microgenetic study protocols with *SURGE* players in order to refine the theoretical assumptions we will use to frame the HMM analysis of game play.

Concluding Thoughts and Avenues for Future Work

Research into the use of video and computer games for learning has moved beyond an initial exploratory stage into more systematic evidence-based investigations of how games can be designed to bolster learning and provide more meaningful assessment. To bolster the case in favor of using games in educational contexts, it becomes necessary to construct a multi-faceted body of evidence that can more clearly demonstrate the diverse forms of learning that well-designed games can promote. Our continuing efforts at developing these assessments in the context of the SURGE project are consistent with that goal.

Furthermore, our investigations suggest the conditions under which the analysis of gameplay data becomes valuable to a program of research. The particular methods of analysis we used may not be appropriate in all cases; however, the practices that made these analytics fruitful are probably available to all researchers. More specifically, our investigations highlight the value of (a) theorizing the significance of game play actions for the learning process, and (b) capturing these actions at a level of detail that is appropriate for the research questions being asked. In the case of *SURGE*, our hypothesis was that individual game play actions were interpretable both individually and longitudinally. This notion was framed by the same theoretical background that informed the entire *SURGE* design, i.e. literature on conceptual change that accounts for naïve or intuitive knowledge (e.g. Parnafes, 2007; Clark, 2006; and earlier work by Roschelle, 1991, and

White, 1983), which itself flows from cognitivist perspectives on science learning. These theoretical commitments, however, are not necessary for gameplay data analysis to be useful. The more central questions are whether the process of playing the game (which may be tentative, experimental, and somewhat haphazard) reinforces a pattern of purposeful behavior in the learner that can be observed with some regularity, and whether these patterns can be shown to have any effect on desirable learning outcomes. If anything, our experience with these statistical methods indicates that gameplay data can be an intriguing source of this kind of evidence, and that future refinement of the tools and techniques will enable educational gaming researchers to craft more comprehensive accounts of student learning using this form of data.

While any single approach to analyzing gameplay will likely portray game-based learning in an incomplete manner, we think that multi-method analyses of the behaviors and choices made by players during gameplay can provide novel perspectives that complement established analyses based on knowledge tests and clinical interviewing. While learners' increasingly effective strategies and problem-solving approaches in a physics game like *SURGE* can be demonstrated through HMM, for example, it is important to correlate these improvements to other measures, including external measures of curriculum-based learning. Such an assessment strategy, based on triangulation of evidence from multiple sources through multiple analytical approaches, will facilitate detailed inquiry into questions about the nature, strength, and persistence of students' learning while playing games.

References

Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. *Proceedings of the Eleventh International Conference on Data Engineering* (pp. 3-14). IEEE Computer Society.

Clark, D. B., & Martinez-Garza, M. M. (2012). Prediction and explanation as design mechanics in conceptually-integrated digital games to help players articulate the tacit understandings they build through gameplay. In C. Steinkuehler, K. Squire, & S. Barab (Eds.), *Games, learning, and society: Learning and meaning in the digital age*. Cambridge, Mass: Cambridge University Press.

Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M. M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education, 57*(3), 2178–2195. http://doi.org/16/j.compedu.2011.05.007

Clark, D. B., Nelson, B., D'Angelo, C., Martinez-Garza, M. M., & Slack, K. (2010, June). *SURGE: Intended and unintended science learning in games*. Poster session presented at the International Conference of the Learning Sciences, Chicago, IL.

Clark, D. B., Nelson, B., D'Angelo, C., Slack, K., & Martinez-Garza, M. M. (2010). SURGE: integrating Vygotsky's spontaneous and instructed concepts in a digital game? In *Proceedings of the 9th International Conference of the Learning Sciences - Volume 2* (pp. 384–385). Chicago, Illinois: International Society of the Learning Sciences.

Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction, 24* (4), 467–563.

De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of educational research*, *68*(2), 179-201.

Fang, Z., Lamme L., & Pringle, R (2010). *Language and literacy in inquiry-based science classrooms, Grades 3-8.* Thousand Oaks, CA: Corwin Press and Arlington, VA: National Science Teachers Association.

Foster, A., & Mishra, P. (2008). Games, claims, genres and learning. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education.* (Vol. 1, pp. 33–50). Hershey, PA: Information Science Reference.

Gee, J. P. (2007). *What Video Games Have to Teach Us About Learning and Literacy. Second Edition: Revised and Updated Edition* (2nd ed.). Palgrave Macmillan.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, *30*(3), 141-158.

Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008). Using hidden markov models to characterize student behaviors in learning-by-teaching environments.

In B. P. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems* (pp. 614–625). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-69132-7_64

Li, C., & Biswas, G. (2002). Applying the hidden Markov model methodology for unsupervised learning of temporal data. *International Journal of Knowledge Based Intelligent Engineering Systems*, *6*(3), 152–160.

Mayo, M. J. (2009). Video games: A route to large-scale STEM education? *Science, 323*(5910), 79-82. doi:10.1126/science.1166900

Nesbit, J. C., Xu, Y., Winne, P. H., & Zhou, M. (2008). Sequential pattern analysis software for educational event data. Presented at the International Conference on Methods and Techniques in Behavioral Research, Maastricht, The Netherlands.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: evidence from the computer games literature. *Curriculum Journal*, *16*(4), 455. doi:10.1080/09585170500384529

Parnafes, O. (2007). What does fast mean? Understanding the physical world through representations. *The Journal of the Learning Sciences. 16*(3). 415-450.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Soller, A., & Lesgold, A. (2007). Modeling the process of collaborative learning. In H. U. Hoppe, H. Ogata, & A. Soller (Eds.), The Role of Technology in CSCL (pp. 63–86). Springer US.

Soller, A., Wiebe, J., & Lesgold, A. (2002). A machine learning approach to assessing knowledge sharing during collaborative learning activities. In Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community (pp. 128–137). International Society of the Learning Sciences.

Shute, V. J., Rieber, L., & Van Eck, R. (2011). Games... and... learning. In R. Reiser & J. Dempsey (Eds.), *Trends and issues in instructional design and technology* (3rd ed., pp. 321-332). Upper Saddle River, NJ: Pearson Education, Inc.

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. In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Mahwah, NJ: Routledge, Taylor & Francis.

Squire, K., Jenkins, H., Holland, W., Miller, H., O'Driscoll, A., Tan, K. P., & Todd, K. (2003). Design principles of next-generation digital gaming for education. *Educational Technology*, *43*(5), 17-23.

Sutton, C. (1996). Beliefs about science and beliefs about language. *International Journal of*

Science, 18, 1-18.

Thompson, C. (2007, August). Halo 3: How Microsoft Labs invented a new science of play. *Wired*, *15*(09). Retrieved from http://www.wired.com/gaming/virtualworlds/magazine/15-09/ff_halo?currentPage=all

White, B. Y. (1983). Sources of difficulty in understanding Newtonian dynamics. *Cognitive Science, 7*(1), 41-65.

Willis, D. (1993). Learning and assessment: Exposing the inconsistencies of theory and practice. *Oxford Review of Education*, *19*(3), 383-402.

Zhou, M., Xu, Y., Nesbit, J. C., & Winne, P. H. (2010). Sequential pattern analysis of learning logs: Methodology and applications. *Handbook of Educational Data Mining*, 107.

CHAPTER III

TWO SYSTEMS, TWO STANCES: A NOVEL THEORETICAL FRAMEWORK FOR MODEL-BASED LEARNING IN DIGITAL GAMES

Introduction

Players of digital games present a remarkable duality. On one hand, players of games often seem to sit in absorption, interacting with a complex digital game in an automatic and nearly effortless way. Observing an individual person at play, it might appear at times that the person is doing little more than reacting to stimuli, rarely demonstrating anything that resembles thinking or learning. On the other hand, players are also deeply reflective about the games they play. This is most visible in the online spaces where communities of players coalesce. In these spaces, we find that games are objects of analysis, inquiry, commentary, interpretation, and re-interpretation. As a means of participation, these activities are in some ways as valid as playing the games themselves.

This duality becomes more salient, and problematic, in the case of educational games. When the stated goal of a game is that its players learn particular content or concepts, the prevalence of automatic and reactive forms of play make it difficult to argue that anything is actually being learned beyond the performance requirements of the game itself. Ideally, the goal is that students engage with educational games in a thoughtful and purposeful way, using these games as tools for organizing their insights and furthering their understanding of the concepts the game is intended to teach. This is not to say that no learning can happen through automatic forms of play; it is certainly true that digital games can support learning of, for example, facts and procedural skills. A more challenging goal is to support learning of the causal relationships and functional properties that feature prominently in science education. These concepts may exist in learners' minds in an intuitive form, and might thus be reinforced through intuitive play. However, the process of normalizing and organizing these intuitions into more formalized modes likely requires the learner to engage with digital games in a more deliberate, analytical manner.

How can we make sense of this duality? How can we reconcile the two modes of digital gaming, the automatic and the thoughtful, into a coherent framework that explains how and what people learn from these games? How can we promote forms of play and learning that are more closely aligned with the goals of education, particularly science education?

This paper explains what I call the Two Stance Model of game-based learning, which is my attempt to answer these questions from a cognitive perspective[1]. The Two Stance Model, or 2SM for short, envisions that players of digital games shuttle between two distinct epistemic stances: (1) a "learning" stance, which is directed toward making sense of the games' rules, the entities and relationships it portrays and (2) a "playing" stance that is geared toward optimizing in-game performance and continuing play. Furthermore, I conjecture that people develop two distinct forms of knowledge through interacting with a game: (1) an understanding of the network of entities and causal structure of the interactive model and (2) a store of practical knowledge of how to act effectively within the game. The former relates to the question “how do things work?” while the latter answers the question, “how do I win?” The two epistemic stances are collections of resources that are geared and optimized for dealing with these two separate forms of knowledge. In the case of the Learner stance, the resources are used to make sense of the game’s rules, systems, and the entities that are represented in the simulated space. This is largely an internal process that implies the construction of a *mental model*. However, this mental

[1] Cognitive perspectives are by no means the only productive ways to examine game-based learning. Sociocultural, situative, and embodied perspectives are also relevant and have extensive research histories. My choice to use a cognitive lens is based on the persuasiveness of the available scholarship in light of the overarching goals of the 2SM.

model is a hybrid; its contents and function are not limited to a single domain, and include elements of diverse origin. Among these elements are the particular qualities of the experience as intended by the designer, the person's interpretation of the underlying logic that gives the game systems internal consistency and verisimilitude (the *interactive model)* and, in the case of games for science learning, the formalized system (the *external model)* that accounts for similar phenomena in the real-world. To express this hybridity, I call this particular form of mental model a *simulacrum*, and envision it as the main resource around which a person constructs his or her Learner stance. On the other hand, the Player stance collects resources concerned with navigating the goals that the game proposes in harmony with the person's own individual goals of play. Knowledge about how to overcome the game's challenges is stored in the form of recipes for efficient action, or *heuristics*. These heuristics have several properties that make them useful and valuable. Because of their utility, they are the privileged form of knowledge found in gaming literature and in the transactions of knowledge that underpin gaming communities.

The overall goals of this paper are to present the 2SM framework, review the theoretical bases it builds upon, and highlight the salient questions of game-based learning that it helps to clarify. The opening section presents the goals of the framework and the needs it is intended to address. The second section outlines the principal claims of this paper: the proposed function and main constructs of the 2SM. Then, I review research warranting the proposed structure and functions of the 2SM. The 2SM is an instantiation of a general theory of human cognition, *the two-system theory*, in the domain of digital games; thus, a brief overview of this theory figures prominently in this section. Next, I discuss implications of the 2SM for the design and research of both educational and non-educational games. The final section discusses the potential contributions of the 2SM to cognitive perspectives on games for learning.

Goals, context, and the need for the 2SM

I will first outline the goals of the 2SM and the research needs it addresses. At first glance it might appear that there is no real need for the 2SM – it might not seem that educational games research has thus far lacked for theoretical perspectives of how and what people learn from games. Nearly all theoretical frameworks to date, however, have proven either unpersuasive or less than useful. Shaffer, Squire, Halverson and Gee (2004) noted that "most educational games to date had been produced in absence of any coherent theory of learning or underlying body of research." In the ten years since, researchers have made considerable efforts to answer this challenge and produce viable theories of learning as they apply to games, yet the fact remains that an overwhelming majority of research into educational games is not grounded in sound learning theory. For example, of the 365 peer-reviewed articles on educational games published between 2000 and 2009, Wu, Hsiao, Lin, and Huang (2011) found that only 77 of them used a specific learning theory as a part of their work. This startling disconnect between empirical research and theoretical explanations can hardly be due to researchers deliberately ignoring theory. Rather, it can be interpreted as a sign that existing learning theory is simply not aligned to the goals of game-based learning research.

One goal of the 2SM is to provide broadly-applicable theory. This means that the 2SM should have explanatory power in a wide range of game-based learning cases. This is an important consideration given the diverse nature of games and the diverse the ways in which people play them. Theories of learning that are fruitful in, for example, lab studies of individuals playing a puzzle-based game separately in a controlled environment might not be as relevant in a classroom or in geographically and temporally distributed Internet communities. Commitments to theories of learning influence not only the design of educational games, but also the lens through which play and learning are observed.

A commitment to a specific learning theory is not problematic in itself, yet any such theory must attend to how people *actually* play digital games. Some educational game theorists envision play as an activity delimited in time and space, with a given purpose and a given structure (e.g. Amory, 2007). This perspective is rapidly becoming obsolete. Gaming has become (or is in the process of becoming) a liminal activity, existing at and across the interface between formal and informal settings, social and individual enterprises, and regimented activity and free-form play. Thanks to mobile technology and networked computing, many players experience games today less as a hobby or pastime and more as an ever-present, always accessible "background music" to their lives, with few contextual or chronological boundaries. These cases are particularly challenging for research because few if any experimental controls can be applied to them. On the other hand, if more controls are applied in service of simplifying inquiry and clarifying evidence, the controlled activity will move further from the messy realities of gaming as people naturally experience them, to the point that "games played for the purposes of educational research" may be argued to be a phenomenon of its own. Thus, game-based learning theory should have *ecological validity*, and be responsive to gaming-related phenomena as they occur "in the wild".

The 2SM aims for ecological validity in two ways. First, the 2SM describes two general-purpose stances that would be observable in players of any game, not just an educational game. This is a crucial consideration given that educational and recreational games share a design language, to the extent that educational games often pass for leisure games, and vice versa. How accurate or persuasive would a specific theory be if it explained play and learning in one kind of game but not in another? Second, the 2SM explicitly acknowledges and seeks to account for the learner's motivation and agency, which strongly influence the two general-purpose stances. An oft unspoken assumption of educational games research is that students will always play to the best of their ability, for as long as the experiment or observation lasts. This assumption does not align well with the design language of digital games or with the enacted experiences of players. Players "in the wild" will simply cease to play if the game is too easy, too difficult, too monotonous, or is simply not providing the experiences they seek; players of educational games often have no such options. The 2SM sidesteps this by simply assuming that whoever does not want to play, will not play *well.* As we will discuss later, this desire to play well (i.e., to act in the game context in a way that satisfies a person's goals) is the main driver of learning. From the perspective of the 2SM, without a will to play well, the Player stance is disrupted frequently, the Learner stance is rarely invoked, and learning simply does not happen.

The last goal of the 2SM is to complement existing game-based learning theory, and thus support new research questions by providing a new interpretative lens. As a general-purpose framework, while the 2SM lacks the specificity of some existing theories, it does provide a bridge between the two major themes of game-based learning theory. The first theme relates to the way in which games are viewed, namely, as providers of educationally valuable experiences, much like school laboratories. A digital game is able to represent events and systems more diverse and detailed than what students would normally have access to. This capacity for complex representation and scaffolding is thought to afford learning opportunities that are novel and unheralded, and digital games are often portrayed as a special case of learning. Thus, this theme is more or less centered on guiding game design, as in e.g. Amory (2007), de Frietas & Oliver (2006), and Kiili (2005). With regard to this theme, the 2SM examines the interplay between the intentional, or designed, features of a learning game and the ways in which these features influence a person's epistemic stance. The second theme focuses inquiry on the gamer *as a learner*. This theme includes situative and socioconstructivist roots (e.g., Lave & Wenger, 1991; Brown, Collins, & Duguid, 1989; Greeno, 1993). In this view, the virtue of games lies in that they allow learning by doing, and what students can "do" in a video game is often more complex and more demanding than what they are asked to do in school. The learner playing a video game appears engaged, enthusiastic, competent, and capable of intense bouts of collaboration, problem solving, observation, and experimentation. These are positive states for learning regardless of the particular topic or curricular focus. At least part of the theory-building effort within this second theme is devoted not only to helping learners achieve these states for the duration of play, but also to developing modes of play that empower learners to build effective and durable identities as they participate meaningfully within communities that value specialized performance. To this socioconstructivist theme, the 2SM provides clarity as to how and why social forms of play occur, and how and why specialized knowledge communities coalesce. The narratives for collaboration and specialized knowledge communities are incomplete without more comprehensive theory about what is being transacted and what the measure of value is – the 2SM contributes to this through the construct of heuristics. These points of intersection for the socioconstructivist theme, as well as the points of intersection for the design theme, are developed further in the Implications section later in this paper.

Proposed Structure and Function of the 2SM

This section operationalizes the proposed structure and function of the 2SM. Here I lay the main claims of the 2SM, which in subsequent sections I will explore, expand, and warrant through review of research across multiple fields.

Scope of Inquiry

To put some helpful boundaries around these claims, I will define the scope of inquiry first, limiting the analysis to digital games designed around *interactive models*. Games of this kind exhibit certain structural and design characteristics which — while not universal — are quite prevalent and, I conjecture, extremely influential in shaping thinking and learning. Interactive models are computational models of real or hypothesized situations or phenomena that allow users to explore the implications of manipulating or modifying parameters in a purposeful way (see also Clark, Nelson, Sengupta, & D'Angelo, 2009). Reform perspectives on science education hold that exploratory interactions that allow students to test hypotheses and make inferences are valuable for science learning (Committee on Science Learning, 2007). This description generally fits real-world activities (e.g., laboratory experiments) as well as digital environments that provide accurate and interpretable responses to students' explorations (e.g., games and simulations).

However, a digital game is not simply an interactive model presented to students in raw form, such as spreadsheets or numeric display panels. Rather, games provide a set of interfaces, displays, activities, scaffolds, goals, and interactions that are intended to guide the learner, communicate the scope and degree of user control, and set useful boundaries around the scale and nature of what is being modeled; what I call the *designed structure of experience*, or DSE for short.

The designed structure of experience refers mainly to certain characteristics of the users' experience in these environments, specifically: (1) the phenomenological characteristics of the environment, (2) how adaptive the environment is to the user's goals, and (3) the quality of the feedback the user receives from the environment. These factors are denominated by time and frequency, so that we can imagine them as densities, i.e. phenomenological density, goal density, and feedback density. Thus, an environment that is phenomenologically dense is one in which the events displayed in the environment are clearly perceptible, reasonably varied, representationally consistent, unambiguous, salient, and frequent. An experience is said to have goal density when a person's goals relative to the experience can be imported, created, revised, and completed at a brisk pace. Finally, we can say that an experience has feedback density when it exhibits clear, logical, and frequent responses to whatever changes a user may introduce.

Digital games can be fairly described as experiences that are highly phenomenologically-dense, highly goal-dense, and highly feedback-dense, although it may be more accurate to say that games are intended – "designed" – to be this way, because the ultimate judge of the structure of experience is the person who enacts that experience. Thus, the gaming experience has a certain degree of subjectivity and can diverge from the designer's intent. For example, a digital game with a low goal density and low feedback density could be experienced as an animated film, regardless of what the designer intended. It becomes clear that a workable framework for learning with digital games must attend to the DSE – specifically, to phenomenology, goals, and feedback.

The designed structure of experience is largely independent of the interactive model. Yet, not all digital games (educational or leisure) can be fairly said to include such a model. In some cases, games rely on more simple ludic forms (e.g. those that challenge a player's dexterity, memory, reflexes, etc.) to provide some logic to the events and entities portrayed in the game. On the other hand, some games require the player to manipulate interactive models that are more sophisticated than those featured in simulations typically used in formal learning environments. These leisure games can focus on formal science concepts, for example, where the central challenge involves navigation or manipulation based on Newtonian relationships (e.g., *Kerbal Space Program*) or the exploration, identification, and exploitation of mineral and ecological resources (e.g., *Dwarf Fortress*). If we also consider interactive models that represent complex networks of entities and relationships which do not necessarily correspond to real-world phenomena, then the domain of digital games that forefront interactive models becomes even larger. That domain now includes not only science-based examples, but also games like *Civilization* or *SimCity* in which the player develops cities and empires in the context of elaborate networks of resource production and consumption that mimic those in the real world, or *EVE Online*, where players collectively participate in a laissez-faire economy fueled by the production, refinement, sale, and destruction of resources directly caused by players' interactions within the game.

The broad term "digital game" thus captures a wide range of possible environments due to the range of variations in the scope and detail of interactive models and DSEs. However, interactive models are still hypothesized as the "learnable core" of many learning environments. Yet, as we argued in Clark and Martinez-Garza (2012), interactive models are not the exclusive province of games intended for science learning; most (although not all) digital games played for leisure include interactive models at the heart of the challenges they pose, even though the user might not be focused toward interacting with this model directly.

Two Models, Two Stances

Before discussing the Two Stance Model proper, we need to make two additional distinctions. The first distinction refers to the three types of models we will encounter while exploring the 2SM, that differ in terms of where they reside and what they are presumed to contain. These models vary substantially in terms of their domain, function, and where they are said to reside. I will refer to most of these entities as “models” for simplicity and to accurately convey the terminology used by other researchers. However, I will attempt to define and differentiate these terms as much as possible to avoid confusion and enhance clarity. The resulting constructs are summarized and compared in Table 2. The second distinction relates to the state-of-mind or stance a person takes in interacting with these models.

Models in the 2SM

Models have a well-established history in both science and science education, and the term encompasses several different ontologies. The two most germane to this inquiry are what I call the “external model” and the “interactive model”. By ‘external model’, I mean the formal abstraction of the scientific phenomena of interest. These abstractions have both explicatory and predictive power and as such are frequent targets of science instruction (Clement, 2000; see also Lehrer & Schauble, 2005). By ‘interactive model’, I refer to programmed software instantiations of these formal abstractions. Since both external models and interactive models have pedagogical value, the ability to present them simultaneously in an engaging way is an important affordance for science education.

A person who interacts with a game constructs some mental analogue as a necessary step in understanding its inner workings. The existence of this analogue is widely recognized by psychology and educational research, and is most often called a mental model (see Mayer, 2005). An accurate, flexible mental model is hypothesized to form a crucial aspect of science expertise (Chi, Glaser, & Rees, 1981), and thus, refining such a model is a focal pursuit of both science education and games for science learning research. The exact nature of this mental model, however, is somewhat underspecified (Doyle & Ford, 1998). The mainline view is represented by Vosniadou and Brewer's (1994) definition of a mental model as "a mental representation [whose] structure is analog to the states of the world that it represents" (p.125).

Mental models are frequently featured in cognitivist explanations of the causal mechanisms behind science learning in digital games. In the literature on the use of games to support the goals of science education reviewed by Clark, Nelson, Sengupta, and D'Angelo (2009), 21 of the 83 papers cited contained some causal explanation for how games may help students learn science, and of those, 11 explicitly mentioned mental models. In these papers, both play and learning are envisioned to be driven by a unitary *mental model* that a learner forms in order to handle the challenges presented by the game. This mental model grows in sophistication and completeness through play and remains available to the student as a tool for problem solving. The explanations that researchers propose for digital learning vary somewhat, but they converge on a form best articulated by Moreno and Mayer (2000): "When students try hard to make sense of the presented material, they form a coherent mental model that enables them to apply what they learned to challenging new problem-solving situations (Mayer & Wittrock, 1996)" (p. 727).

In other words, the proposed mechanism for how students learn science from games is a two-step process. In the first step, students purposefully investigate the digital environment, and "try hard to make sense" of the entities, relationships, and regularities of the portrayed reality. The product of their effort is a "coherent mental model." In the second step, students evaluate situations and solve problems in some future moment using this very same mental model. Notably, the proposition that presenting students with an external model in the hopes that through engagement they will develop a parallel internal mental model also underlies much research on science learning in labs, inquiries, simulations, and other activities.

This explanation implies that the mental models that students create are relatively persistent, flexible, and context-independent. In other words, students form mental models that are fixed in long-term memory, that can be applied to help solve problems of various forms, and that can be transferred beyond the context in which they are formed. It is fundamentally true that people use mental models for all kinds of cognitive processes, including inference, judgment, and prediction (Johnson-Laird, 1983), and digital games should be no exception. It is also likely true that these mental models originate in inferences made from repeated experience (Gigerenzer, Hoffrage, & Kleinbölting, 1991; see also Hasher & Zacks, 1979).

Some researchers propose additional properties of the mental models that students develop in the course of interacting with a digital game. For example, Rosenbaum, Klopfer and Perry (2007) equate increased understanding of a system with a more sophisticated mental model. Marino, Basham, and Beecher (2011) claim that video games promote mental models that have "coherence", using the term in the sense used by McNamara and Shapiro (2005), i.e. that mental models are well-structured representations built from a combination of the person's prior knowledge and the relevant conceptual elements from the game. The audiovisual affordances of the learning environment that aid the formation of mental models are also noted by Clark and Jorde (2004); Taylor, Pountney, and Baskett (2008); Jones, Minogue, Tretter, Negishi, and Taylor (2006); and Moreno and Mayer (2005). The mental model that a student forms is also envisioned by various authors as a tool for understanding and testing scientific theories (e.g., Li, 2010; Bekebrede & Mayer, 2006; Anderson & Barnett, 2011).

The underlying assumption is that external models and mental models are somehow parallel, either in their structure or their domain. Yet in the case of a digital game, what exactly is being represented? One possible answer is that the domain of this mental model is external reality, the “real world”. So, for example, a player of *Angry Birds* forms a mental model about game-specific objects in parabolic trajectories but also to real world entities that exhibit similar properties, e.g. fly balls (as in baseball). If interactive models found in digital games exhibit this sort of “transitive property”, and if the resulting mental models have both explicatory and predictive capacity (as external models do), then it is clear why they would be valuable as pedagogical tools.

However, if the domain of the mental model is not external reality, or not only external reality, then things grow somewhat more muddled. Although some research on games for learning seems to assume that playing games informs mental models that map seamlessly onto external reality, this proposition seems difficult to defend in absolute terms. First, bridging between in-game and real-world entities requires players to engage in a process of abstraction (or “high-road transfer”, Cobb, 2004) that is generally difficult and not often observed (see Detterman, 1993). Second, not all interactive models are wholly accurate representations of the “real world”. Interactive models may be sophisticated, yet need only be "sufficiently representative of the system to yield the desired information" (Apostel, 1961, p. 126) to function as the core of a game. So players of games would also have to negotiate the occasional mismatch between in-game representation and the information from their own senses (e.g., an interactive physics model might not include friction whereas the real world does).

A more tenable proposition might be that players of digital games form mental models that only partially extend beyond the game to describe external reality. So what other uses might the mental model have? The systems control literature, a domain perhaps more akin to gaming than science education, offers some guidance here. In this literature, the purpose of mental models is to generate descriptions of the system's purpose, form explanations of the system's functioning and the observed system states, and make predictions of future system states (Rouse & Morris, 1986). Any mental model that a person forms in response to a digital game has at least some control orientation; in other words, the mental model must provide some faculty that allows a player to exercise some measure of control over the game. Crucially, a person playing a digital game does not necessarily craft a mental model that accounts for the entire capabilities of the game's interactive model, because in fact, such a comprehensive model is not required. To be perceived as effective, a mental model only needs to provide the user with a sense that he or she understands the game and is in control of it. Yet the mental model cannot only have a control orientation, because not all of a game's phenomena flow from the player's control. Aside from game events that are guided by the underlying interactive model, games also comprise a large variety of elements that are intended to structure the experience. These include, for instance, elements regarding the game's interface, its characters and landscapes, narrative events, and interactions between game entities that exhibit regularity but are not subject to a player's input. These elements must necessarily form part of the person's mental model, even when they do not contribute to a control orientation.

Thus, I argue that the mental models that arise in a player's thinking when playing digital games are not exactly the same as the mental models that are hypothesized to influence science learning. Rather, the mental analogues that influence game play have a great deal of hybridity. They are in some measure informed both by external reality and by the figured world of the game, even when those two stand in conflict. They are also more control-oriented and more limited in scope, and thus may be less comprehensive, less accurate, and less consistent than the mental models as described by some researchers. The mental analogues are also strongly influenced by a person's previous game-playing experiences. To represent this departure from a mainline view of mental models, and to express this hybrid structure, I henceforth refer to the mental analogue as a *simulacrum* rather than *mental model*. Table 2 clarifies the distinctions between the model-related constructs from the perspective of the 2SM.

Table 2.
Taxonomy of model constructs

	Target domain	Goal	Applicability	Uses
External model	Physical reality	Explanation and prediction	Context-free	Describing phenomena and predicting possible future states
Interactive model	Physical reality	Representational fidelity	Context-unbound	Pedagogical demonstration Supporting exploration
Mental model	Physical reality	Understanding	Context-unbound	Problem-solving
Simulacrum	Hybrid	Risk-free experimentation	Context-bound	Testing and revising heuristics

Stances in the 2SM

Another necessary clarification for the 2SM involves stances. For the purposes of the 2SM, the term “stance” refers to the state-of-mind or stance a person takes with regard to his or her simulacrum, or the game itself. When creating, refining, or applying a simulacrum, a person might have two distinct goals. The first is to understand the formal structure of the interactive model and the affordances of the causal relationships it represents. The second is to use the simulacrum as a laboratory where actions can be planned and evaluated in terms of their effectiveness at creating a desired state. These two sets of goals imply qualitatively different forms of thinking. A user in the learning stance might purposefully investigate the game in search of information that confirms or disconfirms his or her understanding. A user in the control stance might engage in developing and adopting control strategies, selecting proper actions, and querying the game to determine whether or not these actions lead to desired results (for further description of the “control stance”, see Veldhuyzen & Stassen, 1977). To distinguish between these two stances, I envision the person seeking to further understand the interactive model as engaging in a "learner stance", and the person in the control stance who is actively working toward a goal as engaging in a "player stance".

These stances are epistemological in function; they are formed by collections of resources that a person uses to decide how best to think about the game as they experience it. This definition follows from Hammer and Elby (2003), who conceptualize “naïve epistemologies” as collections of resources, each activated in appropriate and familiar contexts (see also Elby & Hammer, 2001). Hammer and Elby do not provide an exhaustive taxonomy of resources, but what these resources all have in common is their relation to context. According to Hammer and Elby, resources are activated in a way that is *context-sensitive*; in the process of thinking epistemologically, students select from the resources they have the ones that appear to be appropriate and productive in the current context. If the stances of the 2SM have this same texture, then people who play games assemble their stances based on (a) what their own personal trajectories as gamers have furnished them by way of resources, and (b) what the current context offers in terms of cues.

It follows that the Player and Learner stances vary across people and games. For instance, players familiar with a given genre will have more developed resources than players that are less familiar; similarly, physicists may have qualitatively different simulacra available to them while playing a physics game than physics novices do. This is an important consideration that will no doubt warrant future research, yet for the moment it does not present a significant challenge. The 2SM only describes the form of the stances generally; it is more concerned with describing the stances' functioning and how they interact to influence play and learning. The broad characteristics of the stances are given in Table 3.

Table 3.
Characteristics of the Player and Learner stances

	Processes	Goals	Disrupted by
Player stance	Application of execution rules, evaluation of rule effectiveness after the fact	Achieve desired psychological states, maintain agency	Boredom, frustration
Learner stance	Definition and refinement of execution rules, testing their effectiveness	Signal understanding of the interactive model, bolster agency and self-efficacy	Interactive model that is inscrutable, inconsistent, abstruse.

In the 2SM framework, the Player stance contains, as one of many component resources, the person's motivation to engage. Ryan et al. (2006) found that motivation to engage with video games is strongly predicated on subjective experiences of autonomy, competence, and relatedness. These three constructs are described as basic human needs in self-determination theory (Deci & Ryan, 1985). While the basic needs explanation is more general than the games-specific explanation (and not as easily linked to instruction or design), the basic needs explanation more fully and convincingly accounts for why people choose to play a particular game as well as why they choose to play any games at all. Several researchers of games, for example, have investigated why individuals choose to play one game over another (e.g. Bartle, 1996; Sherry, Lucas, Greenberg, & Lachlan, 2006; Yee, 2006), and the general finding is that people choose their games based on an individual preference for certain psychological states (e.g., challenge, competition, social interaction, or a combination thereof). Regardless of whether he or she is seeking to get a good grade, win, perfect a moment of performance, connect with fellow players, or simply learn about the underlying formal abstraction of the scientific phenomenon, the player is driven by the expectation and realization of experiences that trigger subjective feelings of autonomy, competence, and relatedness.

Insofar as the person is able to find these experiences, then he or she will seek to continue to engage with the game. Yet in order to continue to engage (and access more of these experiences), the player must at times probe the complexities of the game's underlying interactive model, learn them, and learn from them. Learning is thus not merely a residual by-product of engagement, but a necessary activity for free, effective, and purposeful action. The 2SM conceptualizes a person's ability to pursue free, effective, and purposeful action within the game environment as agency. In other words, people who are interacting with a game gain or maintain agency when they feel that they can direct the game towards their personal goals. Conversely, when a person does not feel that he or she can affect the game in a way that feels meaningful, we say that he or she has lost agency. One of the two stances – the Player stance – is specifically oriented toward maintaining this sense of agency, by collecting and organizing practical knowledge on how personal goals may be achieved. Thus, the desire to continue engagement, maintain agency, and advance personal goals provides the impetus for continued play and learning.

Putting it all together: The proposed 2SM in action

The constructs described thus far are, for the most part, closely related to well-established concepts in the research literature. What follows is a description of the hypothetical mechanism by which these constructs interact during game play, and how these interactions may influence thinking and learning within educational games. This description is speculative; I offer it to illustrate how the individual components of the 2SM, derived from non-adjacent fields of inquiry, fit together into a more coherent whole. Many of the specifics of this description are supported by the literature (see Warrants section, following); some other specifics will require future research.

Let us assume that a person begins play with a small initial store of motivation to engage but no knowledge of the game's goals or its interactive model. The person's first instinct is to become situated within the environment, find the useful interfaces, and test the affordances of the environment with tentative actions. At this stage, previous experience playing similar games becomes important; if the person recognizes this particular game as a variant of a genre he or she has played before (and the game's DSE supports this recognition), then the person may cue all of his or her existing knowledge as part of process of becoming situated in the game environment. Sometime during this process, and depending on the designed structure of experience, a goal will be suggested or will suggest itself to the player's thinking, immediately triggering a self-query, "how do I achieve this goal?" The self-query shifts the person towards a Learning stance, and a simulacrum is constructed in response to the query. This simulacrum may be partial, inaccurate, or inconsistent, but at this stage, its only requirement is that it suggests one or more steps that might bring the state of the interactive model closer to the goal state. These steps are rendered as heuristics ("When this, Do that") and relayed to the interactive model through whatever controls or interfaces the game allows.

The interactive model processes the player's actions and outputs the appropriate response. The person is now in a position to evaluate the effect of the executed steps in terms of their effectiveness at modifying the state of the interactive model towards the goal state. Actions that prove effective are reinforced and actions that have a negative effect are rephrased as avoidance steps ("Don't do that"). Actions of negligible, ambiguous, or indeterminable effectiveness are discarded. With repeated reinforcement, the player will begin linking the effectiveness of an action with the circumstances that are present whenever that action is invoked. Thus, effective rules, both execution and avoidance, are matched to the context cues from the virtual environment and stored as heuristics, i.e. "If this, do that." These heuristics are easy to remember, quick to access, and require nearly no cognitive effort to execute.

Whenever players find themselves in a situation that is covered by one of their heuristics, they will in most cases default to that heuristic. In other cases, i.e. when the current conditions cannot be matched with the conditional part of any heuristic, players have two options. The first option is to attempt a coping strategy that allows play to continue even if the state of the game is not advanced toward the person’s goal in any obvious way. This strategy does not necessarily cue or reinforce a heuristic (thus, does not improve the player's apparent skill), but it allows the person to remain in the "player" stance. The second option is to shift to a Learning stance and re-examine the simulacrum and use it to find new possible actions. The revised simulacrum is qualitatively similar to the one last used, since the reinforcement of effective actions also has the effect of reinforcing the simulacrum that suggested that action. In this manner, a person's simulacrum can evolve cumulatively and iteratively, but only if (a) the person is prompted by the absence of productive heuristics and (b) the person reconsiders the simulacrum for the purpose of generating new heuristics. If the player is never without a heuristic to apply, the simulacrum can be disregarded as the person simply defaults to the available heuristics.

By the end of the interaction, a person has gathered three forms of knowledge about the game: (1) a collection of observations about the particular conditions that the game presents or can present, (2) a set of heuristics, i.e. rules of action whose activation criteria match up to these conditions, and (3) a simulacrum, or hybrid mental model, comprising a network of entities and causal relationships involved in the interactive model and working theories of how these entities and relationships influence the game's structure of experience.

The person can reflect and communicate differently about these three forms of knowledge. Observations and depictions of states and conditions within the game can be readily described verbally, with the aid of physical gestures or other visual aids. Once a semantic domain (see Gee, 2007) has been established, communicating heuristics is equally simple, either verbally as an if-then statement, or through a demonstration. The person's simulacrum, however, is much more difficult to communicate — language has fewer tools for expressing networked causal relationships. More often than not, these relationships are rendered piecemeal as if-then statements, in a manner resembling p-prims (diSessa, 1993). Simulacra are heavily dependent on the person's specific play experience and trajectory, so that two different people playing the same game might develop two distinct mental depictions of the game's structures. Conversely, people playing the same game (assuming they are equally focused on optimal performance) will converge on similar heuristics. Also, we must keep in mind that simulacra are mainly reconsidered if and when the person encounters a situation for which he or she has no available heuristics. With increasing expertise and experience, and the concomitant development of useful heuristics to guide play, it is likely that simulacra are no longer cued at all. These three factors (difficulty of description, differentiation, and increasing dormancy) combine to make heuristics far more available to people, both individually and as a community, than simulacra. This may also potentially incentivize the learning heuristics during play (rather than the characteristics of the interactive model), since heuristics are more visible and accessible hallmarks of playing skill.

Warrants for the Proposed Structure and Mechanisms

This section reviews and synthesizes research from across fields to warrant the proposed structures and mechanisms of the 2SM. The section first reviews the two-system theory of cognition upon which it builds and then reviews research supporting the structures and mechanisms of the 2SM distinguishing it from the two-system theory of cognition.

The Two-System Theory of Cognition

The 2SM seeks to explain the fact that players of digital games seem to be equally capable of quick decisions made with minimal information as well as slow deliberations that include a lot of data. This duality is not, however, unique to digital games: it is a feature of human cognition writ large. Digital gaming is simply one realm in which this feature has notable and, to a certain extent, problematic effects. Thus, general theories of human cognition that shed light on the aforementioned duality are especially relevant. A review of the literature shows that both elements of this duality in reasoning and choice have their own lines of research, yet the framework that harmonizes them is of comparatively recent mint.

Contemporary scholarship has produced two broad perspectives for explaining human reasoning: classical rationality and bounded rationality. Classical rationality builds on Aristotelian and Hegelian notions of how the mind operates (see Chater & Oaksford, 1998; Smith, Langston, & Nisbett, 1992; Stenning & Van Lambalgen, 2008); it is also called *unbounded rationality,* as it describes logic-based processes of reasoning with little or no regard for how time-consuming or information-intensive these processes might be. The argument implicit in classical rationality is that, to the degree that people are capable of making optimal choices, the mind must be capable of whatever computations or information processes are required. In contrast, the bounded rationality approach suggests that it may be more appropriate to acknowledge the peculiar properties and limits of the mind and environment; thus optimal performance is not a requirement or assumption of bounded rationality approaches (Gigerenzer & Goldstein, 1996; Gigerenzer & Selten, 2001). Bounded rationality therefore not only assumes that people are limited in computational power, time, and knowledge, but also that each environment varies in affordances for making information available.

Simon (1956) proposed what is arguably the predominant theory of bounded rationality, which he called "satisficing". In this approach, people's minds reason in a way that makes mostly correct choices, but within the constraints of their limited abilities to search for and process information. These satisficing mechanisms might have biological origins that are evolutionarily ancient (e.g., Stanovich, 1999). They require little information or processing power, and they bear nearly none of the usual hallmarks of analytical reasoning, yet they produce results that are usually adequate for the situation at hand. A theoretical description of these mechanisms was given by Valiant (2013), who proposed a class of computations he calls "ecorithms", whereby an organism can act effectively within a system that the organism does not fully perceive nor understand. These "ecological algorithms" are computationally simple rules that, for example, allow an organism to learn non-trivial classes of concepts in a limited number of steps (see Valiant, 1984), making heuristics-based learning behavior possible even for relatively simple organisms.

If satisficing and ecorithms can provide workable solutions to the challenges faced by relatively simple organisms, then these strategies must be available *a fortiori* to more cognitively sophisticated organisms, such as human beings. Yet when applied to humans, these explanations fail to account for the fact that people can and do, in some circumstances at least, make choices in a way that could be described as logic-based, optimizing, and information-rich (Conlisk, 1996;

Smith et al., 1992). Critics of bounded rationality also observe that many problem domains allow for both optimizing and satisficing approaches, such as business management (Odhnoff, 1965) or aircraft design (Brown, 1990). So although unbounded rationality seems unworkable and satisficing seems reductionist, it is clear there is a place in theories of cognition for both quick processes that work with limited information (satisficing and ecorithms) as well as more resource-intensive, rule-following, information-rich processes (unbounded rationality).

Thus a synthetic approach, called the dual-process theory of cognition, aims to explain the simultaneous existence of, and the interplay between, the process of mind best described as "intuition", or effortless thought, and the more deliberate, purposeful activity usually called "reasoning" (Chaiken, 1980; Epstein, 1994; Nisbett, Peng, Choi, & Norenzayan, 2001; Sloman, 1996; Stanovich, 1999). These modes of cognition are neutrally labeled in the literature as System 1 and System 2, respectively. The former is described as fast, automatic, associative, emotional, and opaque; the latter as slower, controlled, serial, and self-aware (see Evans, 2008; Kahneman, 2003).

Many of the cognitive functions included in System 2 can be described using existing language and constructs from classical rationality (e.g. mental logic, deductive reasoning, mental modeling), some of which date to antiquity. In contrast, the study of System 1 has required more novel constructs. These constructs have to be robust enough to sustain the broad range of functions that System 1 is theorized to perform, yet to be simple enough to fit the general description of System 1 as fast, efficient, associative, instinctive, and automatic (see Gigerenzer & Goldstein, 1996). One construct that has shown much promise is heuristics. The extensive role of heuristics in human reasoning was the focus of the work of Kahneman and Tversky (Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1973; Tversky & Kahneman, 1974), who proposed that the processes of human judgment are not just extremely-simplified rational (i.e. System 1) mechanisms. In Kahneman and Tversky's view, judgment relies on simple, efficient mental shortcuts, namely heuristics. While heuristics also appear in the description of satisficing, these satisficing heuristics are generally assumed to be a product of the structure of the environment and not necessarily useful outside a specific context. In contrast, Kahneman and Tversky (1973) propose a limited, stable set of general-purpose heuristics that underlie many of the judgments that people are called upon to make as part of their everyday thinking.

Frederick (2002) further distinguishes between the “judgmental heuristics” of Kahneman and Tversky and what he calls "choice heuristics". Judgmental heuristics are invoked mostly with concrete stimuli that evoke an immediate impression, and result from cognitive processes that are rapid and not entirely controllable, and thus are associated with System 1. Choice heuristics, on the other hand, act on abstract stimuli that evoke no immediate impression, and are oriented towards arriving at some type of analytic solution. In other words, a person's choice heuristics are invoked in response to situations in which System 1 processes are not appropriate to the task at hand, and thus choice heuristics serve to manage System 2 processes (a function analogous to some aspects of meta-cognition). Crucially, choice heuristics are intentional; they are seen and experienced by people who use them as simplifying strategies. Thus, a decision maker is seen as one who modifies his or her heuristics strategically, selecting from known strategies the set that is most appropriate for the current task or context, based on the quantity and quality of information available (see Payne, Bettman, & Johnson, 1993). These strategies themselves are subject to the limits of bounded rationality and are often subject to satisficing, e.g. a person will consider their strategy as "good enough", given the limits in time, effort, and available information. Similar processing assumptions are made by Evans (2006) in his heuristic-analytic theory of reasoning. This theory proposes that System 1 processes cue default judgments that are (usually very casually) endorsed by System 2; however, purposeful deliberation may be applied to inhibit the biased response and formulate a reasoned response in its place.

There are two main strands of the dual-processing theory; they differ in the hypothesized relationship between Systems 1 and 2. In one strand, (e.g., Sloman, 1996), both forms of processing are active in parallel, and in the other strand (e.g., Evans, 2006), they act sequentially and selectively depending on context. Research in both strands agrees that System 1 processing is more common in everyday tasks than System 2. The difference in effort required by these two systems indicates that the processes involved in System 1 are similar to basic performance-oriented computations that the mind has evolved to make. The biological origins of System 1, which are postulated to be shared with other animals, is a recurring theme in two-system theories of cognition (Evans, 2008). In parallel-processing theories of cognition, the preference for System 1 processes is explained as a strategy to minimize cognitive effort, i.e. the "cognitive miser" of Fiske and Taylor (1991). In sequential-processing theories, System 1 is seen as the default mode of cognition, with System 2 acting in a more supervisory, inhibitory, and/or interventionist role (see Stanovich, 1999).

In summary, the two-system theory of cognition seeks to explain the fact that people seem to be equally capable of both quick decisions made with minimal information and also slow deliberations that include a lot of data. The evidence supporting the two-system theory of cognition is extensive and persuasive (e.g. Evans, 2003; Sloman, 1996). Furthermore, the two-system theory explains this duality in reasoning without resorting to Bayesian notions of information processing, i.e. unbounded rationality, while still allowing for some analytical mental processes. This account supports the view that people have extensive resources for making choices, judging, and acting in a low-effort, low-information way, as well as through more sophisticated reasoning through the manipulation of complex mental models.

Expanding the Two-System Theory

While the two-system theory of cognition is still undergoing extensive refinement and conceptual clarification (Evans & Stanovich, 2013), even in simple form it provides a useful structure for understanding how students reason with interactive models and what they may learn from them. The 2SM expands the two-system theory with two additional features, namely the simulacrum-interactive model dichotomy and the Player/Learner stance dichotomy. I will discuss existing warrants for these two dichotomies.

Evidence for the simulacrum-interactive model dichotomy. The simulacrum / interactive model dichotomy has been suggested by other researchers who study the use of models for instruction. Rieber (1996) distinguishes between the "target system" (the system of interest) and the "mental model" (which represents the user's current understanding or working theory about the target system). These mental models form "the basis for the user's decision-making and action when confronted with problems in the target system" (Rieber, 1996, p. 44; cf. Carroll, Olson, Anderson, & National Research Council Committee on Human Factors, 1987). This pragmatic orientation of mental models is echoed throughout the systems control literature, where mental models are seen mainly as internal mechanisms for forming control strategies or selecting appropriate courses of action (e.g., Rasmussen, 1983; Veldhuyzen & Stassen, 1977). This resonates with the 2SM’s construct definition of simulacrum, namely, in that they form and function around the need to exert and maintain control of a system.

The characterization of simulacra as incomplete and unstable follows closely from Norman's (1983) oft-cited observations on mental models. Norman proposes that "the models that people bring to bear on a task are not precise, elegant models" (p. 8), but rather, they are minimalistic and full of idiosyncratic quirks. Also, mental models are variously described in the literature at best as "analogical, incomplete, and sometimes very fragmentary" (Farooq & Dominick, 1988, p. 487) and at worst as "messy, sloppy, [...] indistinct knowledge structures" (Norman, 1983, p. 14). Simulacra, as mental models, inherit much of this “messiness”, and include a further source of “idiosyncratic quirkiness” in the form personal goals and motivation.

A relevant example is found in Steingold and Johnson-Laird (2002), who studied the role of mental models in strategic thinking. Participants were asked to make an optimal choice in a simple two-player game where only two options were available to each player. Steingold and Johnson-Laird found that people regularly form mental models based only on their own options and not their opponents' options. This bias could be quite prevalent in digital games, such that people's simulacra might simply not account for the mechanisms and relationships that are not under their control. Another source of the "quirkiness" of simulacra might arise from what Zhang and Norman (1994) call "the representational effect", i.e. the potential for dramatically different cognitive behaviors to arise from different yet isomorphic representations of structures or relationships. Most games provide a variety of choices at any given moment, which may lead to many possible outcomes that, while not necessarily different in terms of representation, might still result in differences in individual users' mental models due to the representational effect.

Evidence for the learner and player stance dichotomy. In terms of learner and player stances, Schwartz and Black (1996) presented compelling evidence that people will use modeling as part of a solution strategy during problem solving *only when an applicable rule is* *not available to them*. Schwartz and Black conducted a series of experiments where the central task involved solving problems related to chains of gears. These problems could be solved either by using a parity rule that study participants could induce from the problems themselves, or by using a mental or a hybrid mental-gestural model of the behavior of chained gears. The problems were sequenced in such ways that rules could be more easily induced in some cases than in others, for example, by presenting several problems consecutively that could be solved using the same parity rule. Schwartz and Black associated quick response time with rule use: the absence of gestures as well as increased answer speed indicated that participants were reasoning without a dynamic model. Furthermore, the researchers found that most of the participants' errors occurred when the parameters of the problem were changed in such a way that an induced rule would not apply. In their discussion, Schwartz and Black identified three situations in which people used models: when they confronted a novel problem, when they needed to generalize rules, and when their rules failed (p. 493). These findings support the Player/Learner dichotomy in the sense of a heuristic “player” mode versus a model-building “learner” mode.

Similar findings appear in Gijlers and de Jong (2013), who investigated the use of concept maps to aid learning in an inquiry activity based on a kinematics simulation. Gijlers and de Jong coded the utterances of dyads and examined and classified them according to their transformative (i.e. directly yielding of knowledge) or social dimensions. One social dimension code was "integration-oriented consensus-building.” Gijlers and de Jong found that the relative prevalence of these types of utterances correlated with improved learning outcomes along three measures. Interestingly, integration-oriented consensus-building talk frequently appears in the form of heuristics (e.g., "if mass increases, we will go slower," "velocity has something to do with acceleration," or "the distance covered also depends on the initial velocity"). These type of utterances were often bracketed in the excerpts provided in Giljers and de Jong by talk related to experimentation and elicitation, suggesting that the dyads were transitioning between heuristic and systematic approaches.

Finally, Parnafes and diSessa (2004) found evidence of students coordinating two forms of reasoning around a physics simulation. These forms of reasoning were described as "model-based" and "constraint-based", following a distinction made in the artificial intelligence literature. Model-based reasoning was described as principled and integrated, and used mainly to examine plans and alternative actions – a description closely matching the Learner stance; constraint-based reasoning was seen as heuristic and simplifying, and focused on finding and using means-ends strategies – in other words, a control-oriented goal-seeking form of thinking, which possibly coincides with the Player stance.

Schwartz and Black (1996), Gijlers and de Jong (2013), and Parnafes and diSessa (2004) thus converge on the notion that the peculiar task demands of systems simulations produce clusters of behaviors oriented towards processing or control (i.e., the Learner and Player stances). These clusters of behaviors adapt to changes in a person's moment-to-moment needs regarding their continued participation in the activity.

Evidence from game-based learning research and gaming communities. Two studies in particular shed some light on the 2SM in terms of social spaces and gaming communities.

Steinkuehler & Duncan (2008) studied the Internet forums where players post and communicate about the massively multiplayer online game *World of Warcraft.* Steinkuehler & Duncan coded specifically for scientific habits of mind, defined as a combination of scientific discursive practices, reasoning, and tacit epistemologies. They found that "model-based reasoning" (i.e. using some form of model to understand the system being considered) accounted for 11% of the forum posts in the sample corpus. In contrast, 58% of the posts exhibited "systems based reasoning". This code includes reasoning in terms of inputs and outputs, a close parallel to both constraint-based reasoning and the 2SM construct of heuristics. Furthermore, 30% of the posts in the corpus exhibited an "absolutist epistemology" (from Khun, 1992). According to the authors, this stance "might serve someone well when operating in a virtual world where there really is a single algorithm (or set of algorithms) underlying a specific phenomenon and success is only a matter of finding them" (p. 539). This is exactly the attitude a person would have when acting from a very strong Player stance.

While Steinkuehler and Duncan's findings are certainly suggestive, it is important to keep in mind that their data comes from forum participation which, generally speaking, selects for players who might prefer participating in the community via an open-ended exploratory argument, as in Khun's "evaluative epistemology". Thus, it may be that means-end reasoning and absolutist epistemologies are most likely far more prevalent among players of *World of Warcraft* than the ratios Steinkuehler and Duncan report.

Schrader and McCreery (2008) investigated expertise as a function of game-related behaviors in a massively multiplayer online game. The authors conducted a factor analysis on a questionnaire of game behaviors. They found that the question "I learned what I needed to become an expert from my own trial and error" clustered with behaviors associated with game performance, e.g. "I understand the underlying game mechanics", and "I frequently try to think of new ways to react to in-game situations", but this correlation was by far the weakest in that factor. Conversely, the same question (endogenous attribution of game performance) correlated negatively with two other items that indicated exogenous attribution, e.g. "I reached my level of expertise due to another player's guidance." Another factor collected four knowledge-seeking behaviors (although the authors label this factor as “Technology Competence”, as in, knowing how to make the knowledge readily available within the game interface), asking for example, if players have other programs running to check information, or use multiple resources to solve in-game challenges. This factor, along with the game-performance factor (that included the endogenous attribution item) and another relating to status, was positively correlated with game expertise. On the other hand, exogenous attribution of game knowledge correlated negatively with game expertise. In other words, players who consider themselves experts frequently make use of outside sources yet tend to think that they learned the game by themselves; people who don’t see themselves as expert players don’t seek community-created knowledge and attribute their knowledge to the mentorship of other players.

Schrader and McCreery do not develop this theme, but the nature of their data and their analysis supports some conjectures. First, players of games feel they gain status (i.e. relatedness and public recognition of their competence, cf. Deci & Ryan, 1984) when they become skilled, and that the hallmark of a skilled player is seeking and using condensed sources of information. This indicates that performance and goal-seeking orientations, i.e. the Player Stance, inherently drive players towards gaining and using effective knowledge, wherever it is available. As to the origins of their own effective knowledge, more skilled players do not attribute their skill to the mentorship of other players, yet they do not unanimously attribute their skill to their own trial and error experiences. This makes sense if the process of growing more skilled involved shuttling in and out of the use of trial and error as a source of understanding (i.e. a Learner stance) and included tapping into other sources of knowledge.

Implications for Design and Learning

Thus far, I have conceptualized how learning from games happens from a 2SM perspective and presented evidence from diverse research programs that support the constructs and processes here proposed. I will now discuss the implications of 2SM as it impacts the design of learning environments and the range of learning outcomes that are possible.

Goals for design

From a 2SM perspective, the success of a game as a pedagogical tool depends mainly on two factors: (1) whether or not the student's simulacrum remains accessible after it loses its value as a tool to guide effective play, and (2) whether the simulacrum and heuristics that the student generates during play are useful in the target domain. The goals of the designer are therefore (1) to build enough support, feedback, and reinforcement into the learning environment so that the learner simulacrum is strengthened throughout play, and (2) to structure game mechanics such that the simulacrum and likely heuristics for optimal play support understanding or problem-solving in the target domain. The negotiation of these priorities might shift according to how a particular domain's notion of expert knowledge balances between abstract yet flexible understanding of a system's workings and quick execution of procedures that are known to be effective. In the case of the former, the simulacrum is the privileged form of knowledge; the goal becomes to enable the simulacrum to enter the student's long-term memory as the kind of context-free causal/relational cognitive structure that more resembles what some scholars view as expert knowledge (cf. Chi, Feltovich, & Glaser, 1981). On the other hand, if heuristics are the preferred mode of knowledge for a particular domain, then it is likely the development of expertise via games more closely resembles skill acquisition (cf. Anderson, 1987).

Alignment between external model, interactive model, and simulacrum

An ongoing challenge in game-based learning research is the matter of “alignment”, namely, ensuring that what students actually learn from an educational game coincides with both the designers’ intent and an externally-validated curriculum (Kebritchi, 2010; Squire & Jenkins, 2003). Mental model-based explanations of student learning generally frame this alignment as a matter of creating accurate facsimiles of the formal abstractions of scientific phenomena, encoding them into the interactive model, and providing the necessary tools and scaffolds so students can best “make sense of” the interactive model. Accurate interactive models, along with useful scaffolds to make the model more clear and apparent, are hypothesized to result in more effective mental models.

The 2SM shifts the issue of alignment from accuracy in the sense of fidelity of representation to accuracy in the sense of proximity between the task demands of the game and those of the curriculum topic of interest. A good example of this latter form of proximity is *Dragon Box*, a puzzle game intended to help students learn concepts of algebra. Designed by Jean-Baptiste Huynh and Patrick Marchal, *Dragon Box* has garnered international attention for its effective treatment of what is typically a difficult range of ideas for younger students to grasp. While no peer-reviewed studies are currently available, there is abundant anecdotal evidence available that pre-algebraic students are able to solve Algebra 1 problems after playing Dragon Box for as little as 42 minutes (Shapiro, 2013).

From the 2SM perspective, *Dragon Box* succeeds, at least in part, because the simulacrum that students form while playing the game is closely associated to the form that algebra problems tend to take, e.g. isolating the variable, or balancing the equation. Not only are the steps and processes similar in both the game and the real-world application, but so are the goals. This similarity in goal structure is rather unique. Researchers have noted that games could portray not only the knowledge base of science and the material methods of science inquiry, but also its purposes, priorities, and objectives, e.g. Shaffer (2005); Barab, Gresalfi, & Ingram-Goble (2010). One example of this enriched portrayal of science can be found in the *Quest Atlantis* games, or in the *FoldIt* protein research simulation/puzzle game. However, the disconnection between the goal structure of *Quest Atlantis* and that of the typical science unit is inevitably felt by students, and thus the mental structures that students form to successfully navigate both forms of learning are perhaps not interoperable. *FoldIt* and *Dragon Box*, on the other hand, hew so closely to the goal structures of the discipline that alignment is tight; students may see the game and the respective disciplinary practices as being integrated procedural knowledge. Thus, from the 2SM perspective, *Dragon Box* is a prime example of a game that successfully manages the "alignment" challenge and structures students' thinking in a manner closely tied to the learning domain targeted by the game.

Helping students form robust mental models

Helping students form robust mental models that can be applied across contexts has also historically been a strong focus of inquiry. Research from the past three decades provides substantial insight into the design of games to support students' ability to engage with them, understand them, and use them as tools for thinking. In fact, this goal has arguably been the overall design imperative in research on educational games, in line with existing principles of instructional design and multimedia learning (e.g. R. C. Clark & Mayer, 2011). Yet behind these principles lies the baseline assumption that once learners form their mental models, these models remain available to students at some later time. The 2SM framework problematizes that assumption by suggesting that simulacra, like all System 2 processes, are pre-empted by System

1 processes such as rules and heuristics whenever these are available (Schwartz & Black, 1996). In fact, given that System 1 mechanisms are the preferred mechanism for everyday thought, it is perhaps incidental that students retain any knowledge at all resembling a mental model that is useful in new contexts.

Thus, the challenge for designers of educational games is to help students store and recall their simulacra, which may run counter to the students' own cognitive biases against using them. The question then becomes, how can designers disrupt a person's natural tendency to discard simulacra? As discussed in this paper, a person is most likely to stop attending to his or her simulacrum (a) once effective rules or heuristics have been derived from the simulacrum and (b) once the heuristics cover all possible situations the player cares to affect. With regard to the first condition, it is probably not feasible to prevent students from forming heuristics. People have strong incentive to maintain agency, and they express this agency by directing the interactive model towards a desired state. Effective control-oriented rules are therefore constantly being created, selectively matched to available data, and evaluated for effect (in a matter reminiscent of the description of production systems in Neches, Langley, & Klahr, 1987). These cognitive processes are associative and prone to automatization. It is therefore unlikely that the designer can interrupt them without damaging people's agency and ability to continue to engage.

The second approach to preserving simulacra is potentially more promising. If a game is designed in such a way that it is constantly offering new goals as well as elements and constraints, then at no time do a person's heuristics provide good-enough play actions for all possible situations. On the contrary, the person must return to his or her simulacrum (from Player stance to Learner stance) to revise and refine it to include the new structures. If the person does not expand his or her simulacrum, then the simulacrum will lose its usefulness at predicting states in the interactive model and evaluating the effectiveness of the potential actions. In this situation, the person cannot retain agency, and his or her ability to engage effectively quickly decays. Thus, whenever the game introduces new elements and these new elements are different enough that the person cannot cope with them using existing heuristics, the person shifts from Player stance back into the Learner stance, the simulacrum is re-invoked, and the process of modeling begins again (so person can eventually shift back to Player stance). If this chain of events happens frequently enough, the person will reinforce the mental model, rather than the set of heuristics. This may, in turn, result in greater availability of the mental model for problem-solving in other contexts.

Dragon Box appears to effectively transition users between Player stance and Learner stance with a very extensive and gradual level progression. *Dragon Box* contains 200 levels, each building on the prior levels, and almost all of them add a new quirk, wrinkle, or complexity. As new symbols and new rules are introduced, students must constantly adjust their simulacra to keep pace. This prevents students from settling on effort-reducing strategies such as heuristics; however, the heuristics that do form are closely aligned the tasks and subtasks found in algebra problems in any case.

A similar example of the effect of constantly evolving challenges can be found in the recreational game *Dwarf Fortress*, a construction and management simulation akin to *SimCity* or *Railroad Tycoon*. What makes *Dwarf Fortress* different is the depth and detail of its modeled world, whose principles operate with a regularity and complexity far beyond most digital games. The game includes systems to simulate basic economic activities like farming, fishing, hunting, and a broad variety of crafts, such as smithing, masonry, and brewing. Each of these activities is supported by simulations of resource growth and propagation (i.e., seeds that grow into plants that bear fruit, fish and wild game that reproduce, and predators that compete with the dwarves for the same food resources), and the behavior and interactions of materials (e.g. wood burns, iron melts, bones decompose, and water that flows into magma produces obsidian and steam).

The complexity and depth of *Dwarf Fortress*' interactive model means that the learning curve for the game is uncommonly steep. Furthermore, there are specific in-game events that, when game conditions are met, trigger an explosion of systemic complexity. For example, when the fortress reaches a certain population, the "dwarven economy" will activate, introducing issues of monetary policy, wealth disparity and allocation of labor, forcing players to mint coinage, set tax rates and deal with the supply and demand of goods. Thus, players remain constantly off-balance: just when the current difficulty fades as the challenges are mastered, new and more complex goals appear, exposing new functionalities of the interactive model. From the 2SM perspective, simulacra are constantly undergoing revision and are never quite discarded, and conversely, the heuristics that people form are (at best) very general guidelines to smooth play, because no general-purpose or always-applicable rules are possible.

Dwarf Fortress exposes a tension, however, between the equally desirable goals of (a) proficiency with the concepts, entities, and relationships of the interactive model and (b) learners' sense of self-efficacy and motivation. There is something fundamentally off-putting about complex games like *Dwarf Fortress*. The sense of disorientation that they produce, of not knowing exactly all that is going on, is arguably not an optimal starting point for learning. The reality is that educational games depend largely on their motivational affordances for their buy-in. Teachers and educators are receptive to games because students tend to find them engaging. It thus seems counterproductive to make games complex and disorienting. On the other hand, the depth and responsiveness of the interactive model encased in *Dwarf Fortress* is closest to the letter and spirit of the justification for using interactive models in the first place, viz. to allow users to investigate the causes of phenomena, and explore the implications of manipulating certain parameters (p. 10, this paper). This tension is perhaps not one that can be resolved, but it can be studied and negotiated through skillful design.

Attending to social texture

Designing educational games from the perspective of the 2SM also requires the designer to attend to the social texture of the context in which the game is played. The two stances are epistemological constructs, and as such, refer mainly to private cognition. However, the boundaries of these stances are not impermeable, and can inform and be informed by social interaction. My conjecture is that most such interactions will center on the circulation of heuristics. The heuristics construct, as described in the 2SM, is a form of knowledge that is easy to transmit, easy to decode, easy to remember, and highly portable. As such, a person's heuristics can become social capital and commodities among player communities. Effective heuristics are prized by players because these heuristics are often the difference between progress and frustration, between satisfying and unsatisfying play. The person who describes his or her heuristics to others gains social currency, a form of prestige that comes from being the person who "figured it out". This effect can be observed both in classrooms, where *ad hoc* co-operative play and helping behaviors are common, as well as in distributed online spaces, where game knowledge is freely shared in a form of *potlatch* or gift economy. The skillful *use* and application of this knowledge is also encouraged when it directly improves a person's level of play, since this improvement is a source of prestige (as shown by Schrader & McCreery, 2008).

The heuristics component of the 2SM supports transactivity in this regard. While simulacra are unwieldy to communicate and share, heuristics are easy to express both in written and spoken form, and the phenomenological regularity of the common game experience means that heuristics are difficult to misunderstand or misapply. Heuristics thus support online social interactions around the game, often in the form of prescriptions for expert play (e.g. FAQs[2], "walkthroughs", "cheats", or tips), and provide impetus for the affinity spaces that organize these interactions around learning and mastery (i.e. "big-G games", as discussed by Gee, 2008). These transactions are framed largely in terms of how to play games optimally, with comparatively little emphasis on formal analysis of how the game works.

Cognitive explanations of game-based learning often rely on interpreting the interactions of the learner with the learning environment rather than taking a broader situated view encompassing collaboration and community. This is not an insurmountable limitation in the case of games that are structured around one-to-one user-to-computer interactions. Yet educational games are frequently used in classrooms without clearly defined user boundaries; it is very difficult to determine the degree to which a mental analogue arises from individual cognitive effort as opposed to participation alongside other students in a collaborative enterprise. In fact, even in games designed for solo play, *ad hoc* collaborations between students are more likely the norm and not the exception (e.g., the student-directed collaborative task pursuit in single-player games described by Sharritt, 2008, and Nilsson and Jakobsson, 2011). Other naturalistic settings, such as massively-multiplayer online games, also share this very thin border between private and shared cognition.

[2] For "frequently asked questions", a genre of guide document in which a community of more expert players collect information for the benefit of more novice players in an effort to limit redundancy. This genre is described extensively by Gee (2004), along with its print analogue, the "strategy guide". The fact that these documents exist indicates that different players have consistent enough experiences so that many questions become "frequently asked".

The 2SM provides a plausible structure for these collaborations. Essentially, students in classrooms use a sort of *distributed* Learner stance. When individuals induce effective rules through the processes described above, they can make them available to others as needed. If people become stuck in the game, they have the option of consulting their peers instead of querying their own simulacra. If a rule of effective action is available, more often than not a more proficient player will share it, although help-seeking and help-giving behavior based on the exchange of heuristics is likely mediated by the norms and sociocultural practices that operate in that particular classroom. If a rule of effective action is not available or forthcoming, then students can continue playing on their own or collaborate in a shared Learner stance until a rule is found. These *ad ho*c collaborations are made possible by the portability and context-unbound nature of heuristics. If heuristics did not have these qualities, help-seeking and help-giving would involve higher cost in time and effort. *Ad hoc* collaborations would thus be far rarer.

Connecting educational and leisure games

One final issue concerns the applicability of the 2SM (or any learning theory) to both educational and leisure games. It may be argued that games for learning require different theories than leisure games in order to account for the added demands of learning. Yet to a large extent, the 2SM treats educational and leisure games as one and the same. The fact that certain games are conceptualized as helping to teach specific content is more an artifact of their design and intended application than any departure from the medium as a whole (see Gee, 2003). From a 2SM perspective, there are only two main differences. First, the elective versus compulsory nature of out-of-school versus in-school gaming probably has some bearing on the resources available to the Player stance. Second, classrooms generally present lower barriers to collaboration compared to gaming “in the wild” due to the direct physical colocation of the participants, emphasizing the importance of attending to the sociocognitive structure of the game-playing community. In other respects, games are games; they use the same structures of experience, the same design language, the same technologies, the same genre conventions, and the same semantics. Thus, I propose that we might prefer a unitary theory like the 2SM, one that allows for taxonomies and contexts yet helps explain the thinking and learning of all players of all kinds of games.

Conclusions

This paper began by focusing on a perplexing duality; players of digital games appear to both act automatically and reflect deeply. I have proposed that this duality, as problematic as it might seem if we assume that learning only happens during moments of analysis and reflection, is perhaps not specific to digital games, but rather a feature of human cognition in general. As such, there is significant and persuasive scholarship that demonstrates how this duality in our thinking and learning capacities works, and I have endeavored to synthesize and summarize this research here.

The Two Stance Model represents the first attempt to reify this general theory of cognition explicitly into the realm of game-based learning. The 2SM, as I have argued here, has a number of promising features that may perhaps shed some light on the persistent challenge of designing a game that helps students learn in such a way that their improved performance in-game has some bearing on their proficiency out-of-game. Among these features are (a) improved explanatory power regarding intrapersonal variation in learning from games; (b) more complete theory regarding individual needs, goals, and agency; (c) a more extensive account of collaboration and community; and (d) improved perspective on knowledge-rich interactions in online affinity spaces. These affordances harmonize well with existing theories of learning, as befits a synthetic approach like the 2SM. Clearly, there are issues regarding game-based learning that the 2SM can probably not address. In these cases, I have tried to limit my claims by deferring to more applicable theory in matters of scope (as in the initial sections of this paper) or by indicating where the 2SM might share a point of contact and focus with general theories of learning, as I have done later in my argument.

While the 2SM is conjectural, and therefore destined to undergo revision and refinement, the research literature provides promising indicia that support its general premises. The work that remains is to craft specific investigations to demonstrate the 2SM empirically. Meeting this challenge will require a mix of observational and quantitative methodologies including sophisticated pattern-finding analytics that support linking game behaviors to epistemological stances to learning outcomes. Fortunately, this combined methodological approach can leverage a burgeoning foundation of tools in the field of game-based learning, as scholars recognize and seek to account for the richness and complexity of game-based phenomena. Enriched multiple-perspective research strategies, supported by both existing theory as well as new frameworks like the 2SM, promise to support more sophisticated student understanding, help learners build more powerful identities, and advance our understanding of the rapidly-evolving world of digital gaming.

References

Amory, A. (2007). Game object model version II: a theoretical framework for educational game development. *Educational Technology Research and Development*, *55*(1), 51-77.

Anderson, J. R. (1987) Skill acquisition: Compilation of weak-method problem situations. *Psychological Review 94* (2), 192-210.

Anderson, J., & Barnett, M. (2011). Using Video Games to Support Pre-Service Elementary Teachers Learning of Basic Physics Principles. *Journal of Science Education and Technology*, *20*(4), 347–362.

Apostel, L. (1961). Towards the Formal Study of Models in the Non-Formal Sciences. In *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences* (pp. 1–37). Springer Netherlands. Retrieved from http://link.springer.com/chapter/10.1007/978-94-010-3667-2_1

Barab, S.A., Gresalfi, M.S., & Ingram-Goble, A. (2010). Transformational play: Using games to position person, content, and context. *Educational Researcher*, *39*(7), 525-536.

Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD Research*, *1*(1), 19.

Bekebrede, G., & Mayer, I. (2006). Build your seaport in a game and learn about complex systems. *J. of Design Research*, *5*(2), 273 – 298. doi:10.1504/JDR.2006.011366

Brown, C. B. (1990). Optimizing and satisficing. *Structural Safety*, *7*(2–4), 155–163. doi:10.1016/0167-4730(90)90065-W

Carroll, J. M., Olson, J. R., Anderson, N. S., & National Research Council (U.S.) Committee on Human Factors. (1987). *Mental models in human-computer interaction: Research issues about what the user of software knows*. Washington, D. C.: National Academy Press.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. doi:10.1037/0022-3514.39.5.752

Chater, N., & Oaksford, M. (1998). *Rational models of cognition*. Oxford; New York: Oxford University Press.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Chi, M. T., Glaser, R., & Rees, E. (1981). *Expertise in problem solving*. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.

Clark, D. B., & Jorde, D. (2004). Helping students revise disruptive experientially supported ideas about thermodynamics: Computer visualizations and tactile models. *Journal of*

Research in Science Teaching, *41*(1), 1–23.

Clark, D. B., & Martinez-Garza, M. (2012). Prediction and explanation as design mechanics in conceptually-integrated digital games to help players articulate the tacit understandings they build through gameplay. In C. Steinkuhler, K. Squire, & S. Barab (Eds.), *Games, Learning, And Society: Learning And Meaning In The Digital Age* (pp. 279-305). Cambridge: Cambridge University Press.

Clark, D. B., Nelson, B., Sengupta, P., & D'Angelo, C. (2009). Rethinking science learning through digital games and simulations: Genres, examples, and evidence. In *Learning science: Computer games, simulations, and education workshop sponsored by the National Academy of Sciences, Washington, DC*.

Clark, D. B., Tanner-Smith, E., & Killingsworth, S. (submitted). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research.*

Clark, R. C., & Mayer, R. E. (2011). E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning (3rd rev. ed.). Chichester, UK: Jossey Bass Wiley.

Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, *22*(9), 1041-1053.

Committee on Science Learning, K. through E. G. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8*. (R. A. Duschl, H. A. Schweingruber, & A. W. Shouse, Eds.). The National Academies Press. Retrieved from http://www.nap.edu/openbook.php?record_id=11625

Conlisk, J. (1996). Why Bounded Rationality? *Journal of Economic Literature*, *34*(2), 669–700.

De Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?. *Computers & Education*, *46*(3), 249-264.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Publishing Company.

Detterman, D. (1993). The case for prosecution: Transfer as an epiphenomenon. In D. Detterman & R. Sternberg (Eds.), *Transfer on Trial: Intelligence, Cognition and Instruction* (pp. 1-24). Norwood, NJ: Ablex Publishing.

diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, *10*(2-3), 105–225.

Doyle, J. K., Ford, D. N. (1998) Mental model concepts for system dynamics research. *System Dynamics Review 14*(1), 3-29.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American*

Psychologist, *49*(8), 709–724. doi:10.1037/0003-066X.49.8.709

Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. doi:10.1016/j.tics.2003.08.012

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378–395. doi:10.3758/BF03193858

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. doi:10.1177/1745691612460685

Farooq, M. U., & Dominick, W. D. (1988). A survey of formal tools and models for developing user interfaces. *International Journal of Man-Machine Studies*, *29*(5), 479–496. doi:10.1016/S0020-7373(88)80007-5

Fiske, S. T., & Taylor, S. E. (1991). *Social Cognition* (2nd. ed.). New York, NY: McGraw-Hill.

Frederick, S. (2002). Automated choice heuristics. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 548–558). New York, NY, US: Cambridge University Press.

Gee, J. P. (2007). What Video Games Have to Teach Us About Learning and Literacy. Second Edition: Revised and Updated Edition (2nd ed.). New York, NY: Palgrave Macmillan.

Gee, J. P. (2008). Learning and games. In K. Salen (Ed.), *The Ecology of Games: Connecting Youth, Games and Learning* (pp. 21–40). Cambridge, MA: MIT Press.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*(4), 650.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded Rationality: The Adaptive Toolbox*. MIT Press.

Gijlers, H., & de Jong, T. (2013). Using Concept Maps to Facilitate Collaborative Simulation-Based Inquiry Learning. *Journal of the Learning Sciences*, *22*(3), 340–374. doi:10.1080/10508406.2012.748664

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388.

Honey, M. A., & Hilton, M. (Eds.). (2010). *Learning science through computer games and simulations.* National Research Council. Washington, DC: National Academy Press.

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness (Vol. 6). Cambridge, MA: Harvard University Press.

Jones, M. G., Minogue, J., Tretter, T. R., Negishi, A., & Taylor, R. (2006). Haptic augmentation of science instruction: Does touch matter? *Science Education*, *90*(1), 111–123.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*, *93*(5), 1449–1475.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge; New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. doi:10.1037/h0034747

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and higher education*, *8*(1), 13-24.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. doi:10.1037/0033-2909.108.3.480

Lehrer, R., & Schauble, L. (2005). Cultivating model-based reasoning in science education. In R.K. Sawyer (Ed.) *The Cambridge Handbook of the Learning Sciences* (pp. 371-388). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511816833.023

Li, Q. (2010). Digital game building: learning in a participatory culture. *Educational Research*, *52*(4), 427–443.

Marino, M. T., Basham, J. D., & Beecher, C. C. (2011). Using video games as an alternative science assessment for students with disabilities and at-risk learners. *Sci Scope*, *34*(5), 36–41.

Martinez-Garza, M., Clark, D. B., & Nelson, B. C. (2013). Digital games and the US National Research Council's science proficiency goals. *Studies in Science Education*, *49*(2), 170–208. doi:10.1080/03057267.2013.839372

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.) *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge, UK: Cambridge University Press.

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & Calfee (Eds.), *Handbook of educational psychology* (pp. 47–62). Mahwah, NJ: Lawrence Erlbaum Associates.

McNamara, D. S., & Shapiro, A. M. (2005). Multimedia and hypermedia solutions for promoting metacognitive engagement, coherence, and learning. *Journal of Educational Computing Research*, *33*(1), 1–29.

Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology*, *92*(4), 724–733.

Moreno, R., & Mayer, R. E. (2005). Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game. *Journal of Educational Psychology*, *97*(1), 117-128.

Neches, R., Langley, P., & Klahr, D. (1987). Learning, development, and production systems. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development* (pp. 1–53). Cambridge, MA, USA: The MIT Press.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310. doi:10.1037/0033-295X.108.2.291

Nilsson, E. M., & Jakobsson, A. (2011). Simulated sustainable societies: students' reflections on creating future cities in computer games. *Journal of Science Education and Technology*, *20*(1), 33-50.

Norman, D. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7–14). Hillsdale, NJ: Lawrence Erlbaum Associates.

National Research Council (Singer, S. Hilton, M.L., & Schweingruber, H.A., Eds.) (2005). America's lab report: Investigations in high school science. Washington, DC: National Academies Press.

Odhnoff, J. (1965). On the Techniques of Optimizing and Satisficing. *The Swedish Journal of Economics*, *67*(1), 24–39. doi:10.2307/3439096

Parnafes, O., & diSessa, A. A. (2004). Relations between types of reasoning and computational representations. *International Journal of Computers for Mathematical Learning*, *9*(3), 251–280.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker* (Vol. xiii). New York, NY, US: Cambridge University Press.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *Systems, Man and Cybernetics, IEEE Transactions on*, (3), 257–266.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development*, *44*(2), 43–58.

Ritterfeld, U., Cody, M. J., & Vorderer, P. (2009). *Serious games: Mechanisms and effects*. New York, NY: Routledge.

Rosenbaum, E., Klopfer, E., & Perry, J. (2007). On Location Learning: Authentic Applied Science with Networked Augmented Realities. *Journal of Science Education and Technology*, *16*(1), 31–45. doi:10.1007/s10956-006-9036-0

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*(3), 349–363. doi:10.1037/0033-2909.100.3.349

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 344–360.

Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K., & Irvin, P. S. (2011). Student learning in science simulations: Design features that promote learning gains. *Journal of Research in Science Teaching*, *48*(9), 1050–1078. doi:10.1002/tea.20437

Schrader, P. G., & McCreery, M. (2008). The acquisition of skill and expertise in massively multiplayer online games. *Educational Technology Research and Development*, *56*(5-6), 557-574.

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science*, *20*(4), 457–497.

Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan, 87*(2), 104-111.

Shapiro, J. (2013, 07 01). *It only takes about 42 minutes to learn algebra with video games*. Retrieved from http://www.forbes.com/sites/jordanshapiro/2013/07/01/it-only-takes-about-42-minutes-to-learn-algebra-with-video-games/

Sharritt, M. J. (2008). Forms of Learning in Collaborative Video Game Play.*Research & Practice in Technology Enhanced Learning*, *3*(2), pp. 97-138.

Sherry, J. L., Lucas, K., Greenberg, B. S., & Lachlan, K. (2006). Video game uses and gratifications as predictors of use and game preference. In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 213–224). Mahwah, NJ: Lawrence Erlbaum Associates.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138. doi:10.1037/h0042769

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. doi:10.1037/0033-2909.119.1.3

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The Case for Rules in Reasoning. *Cognitive Science*, *16*(1), 1–40. doi:10.1207/s15516709cog1601_1

Squire, K. D., & Jenkins, H. (2003). Harnessing the power of games in education. *Insight*, *3*(1), 5–33.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Steingold, E., & Johnson-Laird, P. N. (2002). Naive strategic thinking. In Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society, Fairfax, VA (pp. 845–849).

Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology, 17*(6), 530-543.

Stenning, K., & Van Lambalgen, M. (2008). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.

Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy*, *69*(3), 213–225.

Taylor, M. J., Pountney, D. C., & Baskett, M. (2008). Using animation to support the teaching of computer game development techniques. *Computers & Education*, *50*(4), 1258–1268. doi:10.1016/j.compedu.2006.12.006

Tobias, S., & Fletcher, J. D. (2007). What Research Has to Say about Designing Computer Games for Learning. *Educational Technology*, *47*(5), 20–29.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*(4157), 1124–1131. doi:10.1126/science.185.4157.1124

Valiant, L. (1984). A theory of the learnable. *Communications of the ACM*, *27*(11), 1134–1142. doi:10.1145/1968.1972

Valiant, L. (2013). Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World. New York, NY: Basic Books.

Veldhuyzen, W., & Stassen, H. G. (1977). The internal model concept: An application to modeling human control of large ships. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *19*(4), 367–380.

Williams, M. D., Hollan, J. D., & Stevens, A. L. (1983). Human reasoning about a simple physical system. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 131–154). Hillsdale, NJ: Lawrence Erlbaum Associates.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, *9*(6), 772–775.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., … Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research*, *82*(1), 61–89. doi:10.3102/0034654312436980

Wu, W. H., Hsiao, H. C., Wu, P. L., Lin, C. H., & Huang, S. H. (2012). Investigating the learning-theory foundations of game-based learning: a meta-analysis. *Journal of Computer Assisted Learning*, *28*(3), 265-279.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, *18*(1), 87–122.

CHAPTER IV

INVESTIGATING EPISTEMIC STANCES IN GAME PLAY WITH DATA MINING

Introduction

Digital games are potentially powerful vehicles for learning (Gee, 2007; Prensky, 2006; Mayo, 2009; Shaffer, Squire, Halverson, & Gee, 2005; Rieber, 1996; Squire et al., 2003) and numerous empirical studies have linked classroom use of educational games to increased learning outcomes in science (e.g., Annetta, Minogue, Holmes, & Cheng, 2009; Dieterle, 2009; Neulight, Kafai, Kao, Foley, & Galas, 2007; Squire, Barnett, Grant, & Higginbotham, 2004). Several reviews have concluded that game-based learning offers numerous theoretical and practical affordances that can help foster students' conceptual understanding, engagement, and self-efficacy (Aldrich, 2003; Cassell & Jenkins, 1998; Kafai, Heeter, Denner, & Sun, 2008; Kirriemuir & Mcfarlane, 2004; Martinez-Garza, Clark, & Nelson, 2012; Munz, Schumm, Wiesebrock, & Allgower, 2007). Clark, Tanner-Smith, and Killingsworth (2015) find favorable support for the use of educational games overall, but particularly in cases where games are augmented through application of sound learning theory.

While the general question of whether games can provide productive contexts for learning is approaching consensus, *how* and *why* games work is a more open question. A large number of constructs receive attention as potentially important for game-based learning (Linehan, Kirman, Lawson, & Chan, 2011; Dondlinger, 2007), including constructs as varied as fun, feedback, engagement, flow, problem solving, narrative, etc. Several scholars have proposed design principles to optimally leverage some or all of these constructs (e.g. Annetta, 2010; Kelle, Klempke, & Specht, 2011; Tobias & Fletcher, 2007; Plass, Homer, & Kinzer, 2014). Also, educational games claim a broad spectrum of possible learning outcomes (Martinez-Garza, Clark, & Nelson, 2013b) which, when combined with the vast range of gaming genres, gaming populations, and technology platforms educational researchers have available, creates a vast and constantly changing space of inquiry that resists generalized claims. Furthermore, digital games also present unique assessment challenges. Since games often incorporate novel student activities for which there are no well-established existing measurement methods, measures often need to be developed along with the game in an iterative fashion (Harpstead, Myers, & Aleven, 2013). Thus, some scholars have called for increased methodological rigor and emphasis on usable (i.e. generalizable) knowledge in educational games research (Dede, 2011; Foster & Mishra, 2008).

Regardless of the variation in theoretical framing, methods, or learning outcomes, the common denominator of all game-based learning research is the act of learners' play. Thus, a general claim of game-based learning research can be phrased as "if a learner plays this particular game, he or she will learn this particular thing." Warranting that claim requires identifying who exactly is the player and what exactly is the game, justifying and defining the educational goal, and analyzing the evidence to determine if the goal has been reached - indeed, a significant portion of research is devoted to these ends. I would include one additional part: unpacking what exactly constitutes "play", i.e. what choices the player has available, what informs those choices, and what feedback the game offers in response. Much inquiry into game-based learning is directed towards explicating issues that influence and structure educational gaming, e.g. design considerations, materials and curricula to support educational games, detection of learning outcomes, etc., although not so much play itself. Generally speaking, the act of play as the central driver of learning is somewhat under-examined in the educational gaming literature. Among the possible reasons for this lack of focus are (1) the general difficulty of observing, encoding, and analyzing play systematically, and (2) the limitations of general theoretical frameworks that might help operationalize play in meaningful actionable ways.

Previous educational research efforts that analyzed digital game play at the individual level have relied primarily on observational methods (e.g., Annetta, Minogue, Holmes, & Chang, 2009; Hou, 2012; Sengupta, Krinks, & Clark, 2015). Observational studies that aim for thick description (Geertz, 1973) of gamers at play explicate this richness and often succeed in building strong cases for learning (e.g. Squire, DeVane, & Durga, 2008). However, investigations of play that use a learner's *in situ* performance as an indicator of the learning are generally limited in scope and scale by the costs and demands of observation and coding. A possible way to address this limitation involves the use of log file data. Learners' actions within the game environment, when recorded and compiled, can potentially produce a rich and detailed account that can be productively analyzed using methods of statistical computing (Martinez-Garza, Clark, & Nelson, 2012). These statistical computing methods, variously known as *learning analytics* (LA), or *educational data mining* (EDM), could be used not only for assessment of learning (as we proposed in Clark, Martinez-Garza, Biswas, Luecht & Sengupta, 2012) but also to find underlying structure and regularity in learners' play that would in turn inform meaningful generalizations about what constitutes learning through play in a game environment. Using a combination of log file data and learning analytics, educational games scholarship could potentially transcend these limitations without abandoning deep qualitative analysis (Berland, Baker, & Blikstein, 2014).

Goal and Structure of this Paper

This paper has two goals. The first goal is to investigate the basic claims of the proposed Two-System Framework of Game-Based Learning (Martinez-Garza & Clark, revise and resubmit), a cognitive perspective that may serve as part of a general-use explanatory framework for educational gaming. The second goal is to explore and demonstrate the use of automatically-collected log files of student play as evidence through educational data mining techniques. These techniques have drawn interest from researchers seeking a more nuanced understanding of student action within digital environments. The data mining techniques featured in this paper could potentially find general use, and this paper aims at offering a demonstration of plausible methods and processes that are suited for the specific challenges of game play data.

The sections immediately following lay out the necessary groundwork for addressing these goals. The context for this research is an educational game intended to help middle school students develop a better understanding of Newtonian kinematics. Among its other functionalities, this particular game stores all student actions and collects them in a central database. The Conceptual Framework section describes this game, titled *The Fuzzy Chronicles,* in some detail. Then, a summary of the Two-System Framework (or 2SM) is presented, followed by specific discussion on the implications of the 2SM in the context of *The Fuzzy Chronicles.* A brief overview of current research that makes use of log files from digital educational environments as evidence rounds out the Conceptual Framework section.

Afterward, I articulate the first goal more specifically as two research questions. An overview of the studies and participants, definition of the Player and Evidence models, and the treatment protocol for EPIGAME data is then provided. Each question is investigated in its own section, with separate Results and Discussion subsections. In the Conclusions, I outline some of the opportunities and difficulties of using educational data mining on digital game play logs, future directions for this kind of research, and also propose improved design factors for educational games that might better promote students' behaviors during play to more closely align with those behaviors found linked to positive learning outcomes.

Conceptual Framework

Overview of the Game Environment: *The Fuzzy Chronicles*

For this study, I used the educational game titled *The Fuzzy Chronicles*, codenamed EPIGAME. *The Fuzzy Chronicles* is the third iteration of the SURGE line of digital games intended to help students advance their understanding of Newtonian kinematics. EPIGAME was designed principally by Douglas B. Clark and developed in collaboration with the University of California at Berkeley’s WISE project and Filament Games with grants from the Institute of Education Sciences at the US Department of Education and the National Science Foundation.

The Fuzzy Chronicles (hereafter, EPIGAME) takes the form of a series of puzzles presented as a science fiction adventure. Students play as the space navigator Surge, who must find and rescue space capsules piloted by Fuzzies, adorable but somewhat hapless creatures who are stranded in space. In order to accomplish these rescues, the player must navigate Surge’s spaceship through a two-dimensional spatial grid (see Figure 7 and Figure 8) by tracing a *Trajectory* to the stationary Fuzzy, then placing *Actions* at *Waypoints* along that Trajectory. Most Actions take the form of *Boosts* that propel Surge’s ship in one the four cardinal directions with an amount of force that the player chooses. Game play is divided into *Levels*, each comprising a separate navigational and/or rescue challenge. All Levels have a *Start Point* and an *End Gate,* and may also optionally contain obstacles, such as impenetrable *Nebulas* and *Radiation*, as well as *Velocity Gates* and *Mass Gates* that impede Surge’s progress. These Gates signal an attribute of Surge's capsule (i.e., a specific velocity or mass) that the player is required to match before the Gate will open. Colliding with a Nebula, a Radiation field, or a Gate causes the destruction of the Surge capsule and any rescued Fuzzies, thus failing the Level.

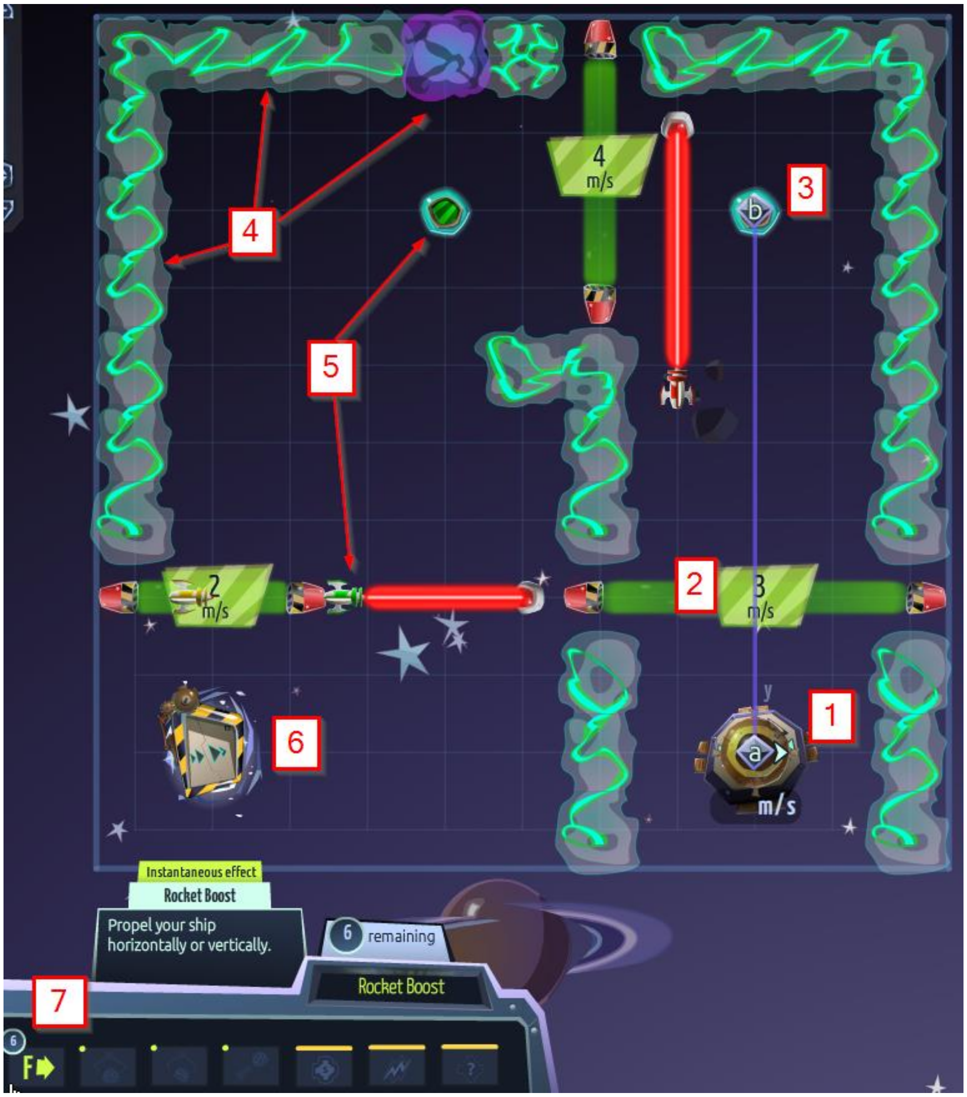

Figure 7. Anatomy of an EPIGAME level. (1) Start Point (2) Velocity Gate (3) Laser Deactivator (or "Button") (4) Nebula (5) Matching Button and Laser (note green color of both) (6) End Gate (7) available Actions.

The interactive structure of EPIGAME has two phases - a *Planning Phase* and an *Action Phase*. In the Planning Phase, players decide their Trajectory and place their Actions appropriately. The player signals the end of the Planning Phase by hitting the *Run Lever,* thus starting the Action Phase. In this phase, Surge's capsule follows the player's plan, which may result either in a successful navigation to the End Gate and the rescue of any stranded Fuzzies or the destruction of Surge's capsule. If successful, the player moves on to the next Level. If the player is not successful and Surge's capsule is destroyed, he or she is returned to the Planning Phase in order to change the planned Trajectory and/or add or remove Actions before triggering a new Action Phase. Together, a Planning Phase and its resultant Action Phase are called an *Attempt* (which may be successful or unsuccessful).

Figure 8. An Attempt in process. The player is setting direction (8) and force (9) parameters on an Action. The player has set a Trajectory (10) through several Waypoints (a-e). To begin the Attempt, the player presses the Launch Lever (11).

In theory, a player may complete the game having needed only one Attempt (i.e. one Planning Phase and one Action Phase) per Level. In reality, players often require multiple Attempts before they successfully advance. In a given Level, the player is free to construct a plan for the entire Trajectory for the level and place all necessary actions before first activating of the Run Lever. Alternatively, players may also choose to segment the Trajectory and place only a few Actions at a time, thereby solving the level incrementally, i.e., draw part of a trajectory, place a few actions, activate the Run Lever, see what happens, and adjust and extend the trajectory and actions iteratively through multiple cycles of attempts. The game neither suggests nor encourages either approach, so a player may select whichever method he or she finds more suitable.

A full game of EPIGAME as designed for this study consists of 32 Levels of generally increasing complexity. Each subsequent level more often than not requires more Actions than the previous ones, contains more challenges and obstacles, and demands more effort by the player to plan and strategize for success. Because of this, it is likely that any player of EPIGAME will find at least one level that requires multiple Attempts in order to succeed. Some levels, particularly near the end of the game, allow only a very limited margin of error. Therefore, progress in the game requires to player to be persistent at times, take several different approaches when faced with apparently insurmountable levels of difficulty, and explore and experiment with different combinations of Actions to find a correct solution for each Level.

The Two-System Framework of Game-based Learning

A goal of this paper is to investigate a theory of game-based learning called the Two-Stance Model framework, or 2SM (Martinez-Garza & Clark, revise and resubmit). The 2SM framework seeks to support a more sophisticated understanding of how and what people learn from digital games. It was motivated by the contrast between recent scholarship that finds uneven evidence that people learn much from digital games (Young et al., 2012) and the observation that players inhabit rich ecologies of knowledge about the games they play (Gee, 2007) that include often-impressive feats of cognition.

Many digital games can be accurately described as software models of scientific phenomena encased within game-like structures that are intended to increase player engagement. In the case of educational games, the intention is that students develop understanding of the principles that underlie these phenomena through the thoughtful and purposeful exploration of their scientific models. The premise of the 2SM framework is that players of educational games do not necessarily form accurate mental analogues of the software models that drive the phenomena they experience in-game (i.e. the encased "simulation"); rather, they create a second-order model (as in, a model of a model of a phenomenon) that is oriented towards explaining the functioning of the encased simulation, predicting its future states, and allowing the user to feel that he or she understands the simulation or game, and has some measure of control over it.

These two stances can be conceptualized further using features from the two-system model of cognition (Evans, 2008). Two-system models of cognition distinguish between effortless thought, or "intuition", and deliberate purposeful "reasoning". These modes of cognition are neutrally labeled as System 1 and System 2, respectively. The former is described as fast, automatic, associative, emotional, and opaque; the latter as slower, controlled, serial and self-aware. In the 2SM framework, System 1 is associated with the "player" stance and System 2 with the "learner" stance.

Players might have two distinct goals when interacting with a game's encased simulation. The first involves develop their second-order model to better understand the simulation and use it as a laboratory the objects and relationships within the simulation can be investigated. The second goal involves executing various game actions to manipulate the simulation to create a desired state (i.e., winning). These two sets of goals imply different forms of thinking about the information being presented by the digital game. Our hypotheses are that (a) the first goal prioritizes or incentivizes an inquiry stance oriented towards purposeful and systematic investigation of the operating principles of the encased simulation; and that (b) the second goal prioritizes or incentivizes a heuristic-driven problem-solving stance oriented towards efficiently achieving the player's goals. A user in the inquiry (or "learner") stance might probe the simulation for information that confirms their understanding. A user in the problem-solving (or "player") stance might only engage in exploratory actions and observing whether these actions lead to positive results.

Starting from the two-system model of cognition, we proposed the following mechanistic explanation for how people play and learn from digital games. A person begins play, and a goal will be suggested to the player's thinking, immediately triggering a self-query, "how do I achieve this goal?" The self-query shifts the person towards the learning stance, and in response to the query a second-order model is constructed. This model's functional requirement is that it suggest actions that would bring the state of the game closer to what the person has identified as a goal state. These actions are rendered as execution steps ("Do that") and enacted in the simulation through the game's interface. Actions that prove effective are reinforced and actions that have a negative effect are rephrased as avoidance steps ("Don't do that"). With repeated reinforcement, effective rules are matched to the context cues from the environment and stored as conditionals, i.e. "If this, do that." These conditionals are easy to remember, quick to access, and require nearly no cognitive effort to execute: they fit the functional definition of heuristics.

Whenever the player finds herself in a situation that is covered by a stored rule, she will in most cases default to doing what that rule stipulates. In other cases, the player must shift to a learning stance, reinstate the second-order model and use it to find new possible actions. If the player is never without a rule to apply, the model is most likely deactivated — the person defaults to System 1-style processing, i.e. fast, effortless, intuitive heuristics. Thus, through play, a person gathers three forms of knowledge about the game: (a) the conditions that the game presents, (b) a set of heuristics, i.e. rules of action whose activation criteria match these conditions, and (c) a second-order mental model, an idiosyncratic explanation of how the game produces the observed conditions. In the case of educational gaming, these three forms of knowledge combine to form part of the learning benefit that students may develop from playing the game.

The 2SM is a novel application of two-system theory of reasoning to educational games. There are suggestive findings from adjacent programs of research that examine forms of reasoning within and around digital learning environments that hint at its validity (e.g. Parnafes & Disessa, 2004; Gijlers & de Jong, 2013). One of the goals of this paper is to explore the fundamental claims of the 2SM, namely that traces of students' System 1 and System 2 reasoning can be observed during play, and that preference for one stance over another has a significant effect on learning. These possible effects are explored in more detail in the following section.

Implications of the 2SM for Learning

In the 2SM, stances are defined as collections of resources (see Hammer & Elby, 2003). The framework stipulates that the two stances that can be associated with cognitive processes described in the two-system theory of cognition (Sloman, 1996; Stanovich, 1999; Kahneman, 2003; Evans, 2008). Thus, a stance or collection of resources organized around System 1 would be optimized for processing speed and effortless thought, while a stance organized around System 2 would be primed for information use and deliberative reasoning. Stances, like resources, are cued around task demands; certain tasks, e.g. driving a car, are structured in a way that they discourage analytic reasoning, while others, like academic writing, are less amenable to quick, associative thinking. That said, human beings are biased in general towards System 1 reasoning as an effort-saving and time-saving strategy (Reyna & Ellis, 1994).

The question then becomes, which of the two stances is most conducive to learning? Intuitively, it would seem that the effortful, analytic processes described as System 2 that drive the Learner stance would be preferred over faster, less deliberate thinking. This would be particularly true in the case of games that are *conceptually integrated* (Clark & Martinez-Garza, 2012) because such games are designed in such a way that thinking about game rules and challenges closely parallels thinking about science concepts and relationships. However, it is unlikely that an educational game can sustain System 2-type processing over long periods. First, players will tend to find ways to save time and effort when negotiating cognitively-demanding challenges, i.e. the "cognitive miser" of Fiske and Taylor (1991). Second, players facing a game they consider *too* challenging may simply disengage, thus negating any educational benefit the game might offer. Thus, a “happy medium” may be more desirable in which players both (a) reflect deeply about concepts and ideas represented in the game and (b) put their understanding into practice in motivating and interesting ways.

As many educational games, EPIGAME is intended to invite learners to think and reason about the concepts and relationships the game portrays and not to merely passively experience them. Players of EPIGAME encounter obstacles and situations of increasing difficulty that are designed not only to provide opportunities for learning but also to adapt to players' increasing knowledge and proficiency over the course of the game. Ideally, students encounter game levels whose difficulty matches but does not significantly exceed their own skill - this alignment keeps interest and engagement high even in the face of ostensibly higher cognitive demands (cf. "flow" in Csikszentmihalyi, 1991). This adaptation is not perfect: students may encounter game levels that are too difficult or too easy. The goal is ultimately not to shield students from difficulty but to provide enough scaffolding and feedback so that the *perceived* difficulty remains manageable.

We propose that a player's response to perceived difficulty is what cues the stances. Which stance gets cued may depend largely on each player's developing understanding of the concepts and relationships underlying the game. Early in the game, the perceived difficulty may be influenced by the learner's prior experience with similar games, or familiarity with the game's targeted concepts and relationships. Thus, the player's prior knowledge of the game, or the principles behind the game's encased simulation may also be a significant factor that cues and organizes the stances. For instance, players with low prior knowledge might prefer a slower, more methodical approach, while players who feel confident in their understanding might play faster, and with less tentativeness, because they may have a more detailed and functional internal model. Later in the game, once all players have had similar opportunities to engage with the game's challenges, these differences might not be so stark, or they may disappear altogether. Therefore, it becomes important to examine the learners' game play to ascertain how the game's varying set of structures and experiences influence players' learning.

Learning Analytics in Educational Gaming

Digital environments that promote learning should prompt a change in student behavior within that environment. If an educational game is designed in such a way that that students are able to apply what they learn in the context of the game, then these changes in behavior should be reflected not only in external measures of learning, but in play itself. If so, then these changes are potentially recoverable and traceable from log data *post hoc*. However, even comparatively simple games allow for a broad range of player interactions, all of which leave their varied and distinct traces. Changes in student behaviors that signal learning can therefore be easily lost in the vastness and complexity of the available data. Methods based on learning analytics (LA) can provide researchers with tools to classify, predict, and discover latent structural regularities even in data sets as voluminous and idiosyncratic as game play logs (Berland et al., 2014). LA techniques not only can help us characterize and describe learning behavior, but they can also deploy Markov-type approaches (e.g. Bayesian knowledge tracing and performance factors analysis) to provide some insight into latent student knowledge. Interestingly, these Markov-type models could be used for prediction, and not just description; for example, to guide adaptive scaffolding and feedback. That said, while more research is required for these applications to achieve their full promise, significant on-going work is already exploring and refining the use of learning analytics on data logs from educational environments.

The use of in-game performance data as evidence of learning outcomes has been proposed by Shute (Shute & Ventura, 2013) and others. Shute and colleagues propose that a learner's actions within the game environment can used as a form of assessment when evaluated against an evidence model, as per the evidence centered design (ECD) assessment framework (Mislevy, Almond, & Lukas, 2003). Under this framework, evidence models are preceded by activity models, which are contextualized and tailored to the particular affordances and constraints of the learning environment. One implementation of ECD which seems particularly suited to educational games, called "stealth assessment", aims to collect evidence model data is directly from the learning environment, bypassing the need for overt knowledge testing that may detract from the play experience. Using this methodology, Shute and Ventura have measured both learning of specific knowledge, e.g. as qualitative physics (Ventura, Shute, & Small, 2014), and also broad cognitive skills and traits, such as persistence (Ventura, Shute, & Zhao, 2013), and 21st-century skills (Shute, 2011).

Activity models can become highly complex, especially in the case of games where many different interactions are possible. This complexity often leads to a large number of observable variables, which in turn complicates the task of formalizing them into an evidence model. Thus, researchers have found value in machine-learning (ML) techniques of computational statistics that can make finding patterns and relationships between large numbers of variables more tractable. Examples of educational games where researchers have used ML techniques to analyze student performance data along an ECD paradigm are the investigation of systems thinking in *SimCityEDU* (R. J. Mislevy et al., 2014) and inquiry skills in *Mission Biotech* (Lamb, Annetta, Vallett, & Sadler, 2014). ECD models that are focused on content-specific outcomes that apply ML techniques are also feasible, e.g. the investigation of student learning of biological processes of stem cells in *Progenitor X* (Halverson & Owen, 2014); of fraction arithmetic in *Save Patch* (Kerr & Chung, 2012); and of Newtonian mechanics in *Impulse* (Rowe, Asbell-Clarke, & Baker, 2015). There are several more exemplars of ML techniques being used to characterize learners' performance in digital environments, although these focus either on learning environments that are simulation-based (rather than game-like) or they do not align exactly with an ECD paradigm. Researchers have successfully applied ML techniques, for example, to describe (a) students' science inquiry activity in *Science Assistments* (Gobert, Sao Pedro, Baker, Toto, & Montalvo, 2012) and in *Virtual Performance Assessments* (Baker & Clarke-Midura, 2013; Clarke-Midura &

Dede, 2010); (b) students' developing engineering thinking in *Nephrotex* (Chesler et al., 2015); and (c) students' understanding of genetics in *BioLogica* (Buckley et al., 2004).

Research Questions

The groundwork laid thus far has discussed the 2SM as a theoretical perspective for examining game play and discussed learning analytics as an approach for analyzing game play through data logs. The next step is to articulate the specific hypotheses and the kinds of evidence that might support them. As mentioned in the Goals section, this paper has two research questions, which we expand upon in greater detail in the following paragraphs.

Question 1: Can the Two Stances of the 2SM, as Specified by the Framework, be detected in Game Play Data?

The first question is intended to test a cornerstone claim of the 2SM, while also evaluating whether the 2SM is a useful lens for interpreting game play data as recorded in *The Fuzzy Chronicles*. The hypothesis is that game play logs exhibit an underlying interpretable structure when features relevant to the 2SM are selected and analyzed. Alternatively, in the case of the null hypothesis, there would be no such structure, or it would not be easily interpretable, or the structures revealed do not correlate significantly with learning outcomes. Such a result would indicate that gameplay is more like a stochastic process, or idiosyncratic, or that players are using purely reactive or arational processes rather than those grounded in cognitive models of performance.

Question 2: How Do Changes in Students' Functional Understanding of the Game Relate to Performance on a Test of Conceptual Understanding?

The second question refers to the feasibility of directly assessing students' emergent understanding of the concepts of Newtonian kinematics represented in *The Fuzzy Chronicles* based on their solutions to small, localized challenges. Each manoeuvre the students are asked to make in EPIGAME (starting and stopping, changing directions, keeping to a set velocity, picking up or throwing an object, etc.) are designed to reify a relevant concept or cognitive resource. By identifying and analyzing students' actions with regard to challenges *of the same type*, both within a student and over time, or between students, we can better understand how these challenges focus thought and learning for individual players. Since EPIGAME is intended to be a conceptually-integrated game (Clark & Martinez-Garza, 2012), the hypothesis is that improved performance in these conceptually-laden challenges indicates greater understanding of the underlying principles of Newtonian kinematics. If the null hypothesis were true, variations in student performance would not correlate significantly with learning outcomes.

Methods

Studies and Participants

To investigate the research questions, two experimental runs using EPIGAME were performed in the months of March and April 2015. The first run was used to address possible confounds, as well as pilot the gameplay data "pipeline", i.e. the entire process of collecting, collating, testing, and analyzing EPIGAME logs. We report on study 1, the pilot study, only briefly as foundation and comparison for study 2. The second study, which is the focus of the current manuscript, deployed the full data analytic process to investigate both research questions. The two studies used the same EPIGAME version, the same assessments, and had roughly the same duration.

Study 1 (pilot study). The participants were 86 ninth grade students from a public high school in Middle Tennessee. In this study, the students were divided into four groups, each randomly assigned into a Solomon four-group design (Solomon, 1949) (Figure 9). The two non-treatment groups participated in their normal classroom curriculum on the topic of force and motion, while the treatment groups only played the game for three 90-minute sessions. Approximately 20 minutes were reserved at the beginning and end of the entire study for a 21-item multiple-choice test intended to assess the students' conceptual and qualitative understanding of Newton's First and Second Law. Two of the groups (one treatment, one non-treatment) completed pre-tests; all four groups completed post-tests 5 days after the experiment began.

The 4-group Solomon experimental design was used in order to obtain a test of the internal validity of the post-hoc effect sizes, and test for interactions between the pre-test and the intervention. My initial conjecture (in line with the 2SM) was that high pre-test score (indicating high prior conceptual understanding of physics) would enable players to form more advanced play strategies. The use of these strategies would then be reflected in post-test gains. However, students might also be primed by the relationships and situations that appear in the pre-test, and post-test gains might correspond not to differences in game play or in prior knowledge, but in a testing effect. Thus, the goal of Study 1 was (1) to determine whether the version of EPIGAME was effective as a learning experience, (2) to investigate any possible testing effects, and (3) to prototype the data collection protocol and some of the analytical techniques. The statistical treatment of the four-group design that allows this disentanglement can be found in Braver and Braver (1988).

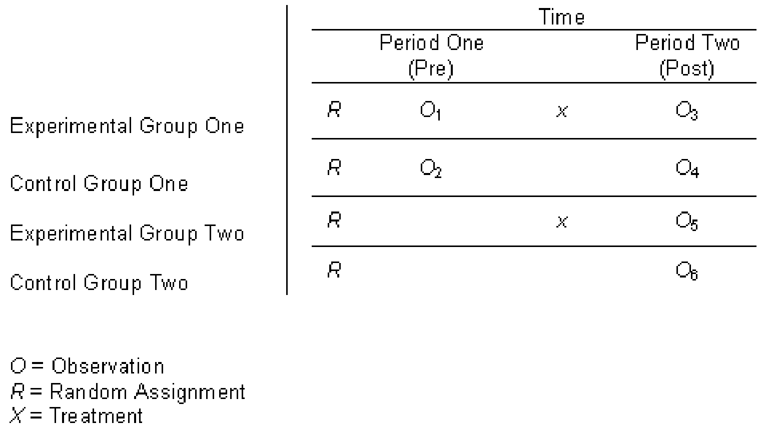

Figure 9. The Solomon 4-group design. Graphic from Braver & Braver (1988).

Two-way within-subjects ANOVA (Table 4) performed on the assessment data showed that students in Study 1 made significant pre-post gains (F = 10.61, df = 104, p < 0.01), with no strong evidence in favor of testing effects (F = 1.11, df = 104, p = 0.29), or interactions between pre-test scores and treatment (F = 0.36, df = 104, p = 0.55). This represents strong evidence that whatever knowledge students are bringing into game play was not gleaned from the pre-test, nor did the pre-test prime students as to which relationships or interactions were important and thus biasing performance in the post-test.

Table 4.
Two-way within-subjects analysis of variance for Study 1

Effect	DF_n	DF_d	F	p	ges
pretest	1	104	1.1161	0.29	0.010618460
treatment	1	104	10.614	0.002 **	0.092608905
pretest:treatment	1	104	0.3552	0.55	0.003404688

Study 2 (research study). Study 1 helped to discard two competing hypothesis: that EPIGAME was not effective as a learning tool (and thus any patterns or changes in game play could not affect learning) and that pre-testing rather than gameplay was the source of any observed pre- to post-test gains. The remaining hypothesis, i.e. that differences in game play were the source of pre- to post-test gains, was the focus of Study 2. In this second study, 123 seventh grade students from a public middle school in Middle Tennessee used the EPIGAME software as part of their normal classroom instruction for five consecutive class periods of 45 minutes each.

As in the prior study, each student had his or her own computer and was specifically instructed to avoid sharing information. The blanket policy was to provide encouragement or hints in lieu of direct assistance, but help was provided to students who appeared intractably stuck, were having technical issues, or had urgent questions about the game interface. As in Study 1, approximately 20 minutes were reserved at the beginning and end of the intervention for a 21-item test of conceptual understanding of force in motion. In this study, all students who were present at the first and last day of the intervention were asked to complete the assessment.

Thus, students who participated in each of the two studies generated two forms of data: pre-post assessment data and game play data. The pre-post assessment data was anonymized and students with missing pre or post test scores were dropped from the study. In the case of students with complete pre and post scores, a unique ID was generated for each; that unique ID was used to link the assessment data with the game play data.

Of the 123 students who participated in the study, 104 provided both pre- and post-tests. A matched-pairs t-test showed a statistically significant increase in test performance ($t = 11.702$, df = 103, $p < 0.0001$) (Figure 10). The value of Cohen's d suggests a large effect size ($d = 1.62$).

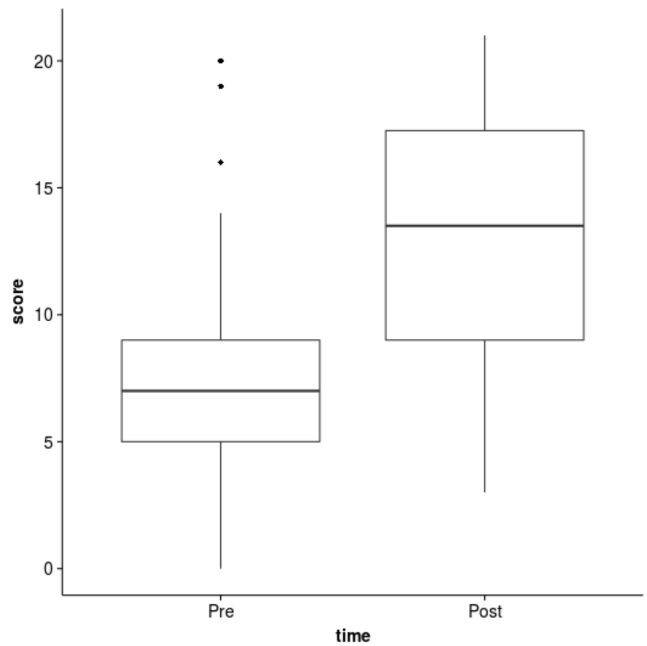

Figure 10. Boxplot of pre- and post-test results for Study 2.

Emulating the Evidence-Centered Design Approach

In the Learning Analytics in Educational Gaming section (above), a significant portion of the research reviewed that used learning analytics to make sense of students' process or log data used an evidence-centered design (ECD) framework for assessment as well. ECD offers several notable advantages for this form of research, *viz.*:

1. the Student Model serves to constrain the number of latent variables that the ML algorithm must infer, aiding in model fit
2. the Evidence Model provides identification rules and ready-made coding schemes, boosting the interpretability of the final model
3. the Task Model pre-selects observed variables likely to be significant, obviating the need for dimensionality-reducing steps, e.g. a Principal Components Analysis to help reduce the number of observed variables to a tractable number.

Considering these advantages, it is clear that learning analytics and ECD processes are well-suited for each other. Unfortunately, it is likely unworkable to apply the ECD framework retrospectively, as the products of ECD are intended to address the specific purposes of that particular assessment (Mislevy, Almond, & Lukas, 2003). Thus, the goal would be to emulate some useful features of ECD, i.e. the Student Model and the Evidence Model. The Student Model can be operationalized in terms of the hypothesized dynamics of the 2SM. The Evidence Model would then map these dynamics into the observable variables. The end result will not be nearly as robust as the full ECD evidentiary argument, but will at least qualify as a cognitive model of task performance, i.e. an illustration of the thinking processes underlying the knowledge and skills students apply *in vivo* when solving educational tasks in a specific domain (Leighton & Gierl, 2007, p. 10).

An important feature of learning analytics and machine learning methods is that they generally do not aim to produce results that have inherent meaning. Unlike statistical treatments of parametric data (e.g. pre/post test results), in which a statistically-significant result indicates a change in the participants' behavior along a measured construct, machine learning and data-mining algorithms generate, at most, descriptions of likely patterns and structures present in the data. It is up to the analyst to interpret what those patterns and structures mean, and evaluate whether or not they support the proposition being researched (Vellido, Martin-Guerroro, & Lisboa, 2012).

Ideally, the interpretation of patterns and structures revealed by learning analytics are supported by robust theory. That is, that features discovered in the data align with existing constructs and relevant explanations for the learning phenomena being studied. In this case, the proposed interpretive lens is provided by the 2SM. Under the 2SM framework, students use collections of resources (or Stances) that organize around the cognitive processes that are optimized for fast ("Player") or slow ("Learner") processing. Thus, the first task is to theorize how these stances would manifest as students play EPIGAME; in other words, how the "fast" and "slow" resources would affect game play. Evoking the evidence centered design paradigm, we will call this operationalization the "Student Model". The second task is then to create an Evidence Model, that is, to deduce how the actions and strategies defined in the Student Model will appear in the gameplay data logs. The goal of the Evidence Model is to select, from all the information contained in the logs, which pieces of data are most likely to characterize the operations defined in the Student Model.

The student model.

The trial-based dichotomous pass/fail task structure of EPIGAME suggests two general strategies for arriving at a solution, one mainly using "fast" processing, and the other using "slow" processing. These strategies (or modes) are:

1) Additive-Iterative Mode, when a student solves a level through a step-by-step iterative accumulation of actions, each checked for efficacy in a separate Attempt.
2) Solve-and-Debug Mode, in which an entire solution is drafted whole-cloth, then corrected only if and as necessary.

While both of these approaches imply that the learner is *thinking*, they differ in what students are thinking *with*, and what they are thinking *toward.* A player using the Additive-Iterative Mode does not necessarily have to have a working knowledge of the game's concepts and relationships in mind; all he or she requires is that EPIGAME provide an unambiguous signal that each added action is a step towards a solution (which EPIGAME provides, in the way of visually-clear animations, e.g., of Surge's capsule exploding or of the Exit Gate being activated). The Additive-Iterative Mode can be thought of as related to Parnafes and diSessa's (2004) "constraint-based thinking." On the other hand, a Solve-and-Debug approach necessitates that the player have a vision of a solution. Armed with a good working knowledge of the rules of operation, a player might feel more capable with taking more actions within each trial because he or she has a reasonable expectation that those actions will be effective. The Solve-and-Debug Mode can be thought of as related to Parnafes and diSessa's (2004) "model-based thinking."

The evidence model.

The two strategies described above represent the best estimate of the forms of play that players are most likely to use. While these forms of play sound very different mechanistically, it is useful to think of them as opposite ends along a continuum. On one end of this continuum, the Solve-and-Debug Mode is slow to plan, is more likely to be correct, and if it is not, it may require only small, effective fixes. On the other end, the Additive-Iterative Mode is fast, less likely to be correct (since a player using this mode may not always define a full solution), and the iterative fixes are more error-prone. Thus, the differences between these two approaches may be captured with only a few contrasting parameters (Table 5).

Table 5.
Forms of solution and their likely parameters

Parameters	Additive-Iterative	Solve-and-Debug
Response time	Low	High
Error rate	High	Low
Actions per attempt	Low	High

The first and third parameters, *Response time* and *Actions per attempt*, are straightforward and directly observable in the data. A longer Response Time indicates slower, more deliberate processing; shorter Response Time corresponds to quick decision-making. Similarly, the number of Actions per attempt is likely specific to each Mode: more Actions taken in the same attempt implies a more elaborate, thought-out plan, while fewer Actions might indicate iterations or corrections.

The second parameter, *Error rate*, will have to be computed from other variables. Broadly speaking, the difference in Error rate between the two Modes represents the willingness of players to accept failed Attempts. Failure during an Attempt is more or less required in the Additive-Iterative Mode, since a player may consider failure as a "partial success" if it creates a baseline upon which he or she can iterate. A student using this Mode may also create a partial solution with some set of parameters he or she knows, and guess at the remaining parameters, counting on the fact that the game will provide actionable feedback. On the other hand, failed Attempts when using the Solve-and-Debug mode are more likely to be unintentional or unforeseen mistakes, rather than intentional probes or guesses. Players using the Solve-and-Debug mode seek to avoid error rather than accept it as inevitable. Thus, the Error rate parameter should incorporate information on how often students fail a level repeatedly, as this continued error would indicate unsuccessful guessing and/or low-information processes such as exhaustive testing of all the available actions.

Treatment of the EPIGAME logs

The data analysis of EPIGAME logs from Study 2 proceeded in four phases:

1. data normalization and integrity checks,
2. variable selection and dimensionality reduction,
3. clustering of student gameplay data and sequence mining, for Question 1, and
4. contextual feature mapping, for Question 2.

Phase 1: The initial corpus of gameplay, recovered directly from the classroom WISE server, was composed of 16,239 records. Each record was comprised of one particular student's attempt to solve one particular level. The particular build of *The Fuzzy Chronicles* used in this study had 32 levels; thus each student produced an average of 132 attempts, approximately 4 attempts per level. Each record comprised a JSON object detailing the specific parameters of the attempt the student performed, i.e. where on the map an action was placed, how much time the student took to plan their actions, which values the student chose for each parameter of each action, etc. The dataset contained approximately 1.1 million of these gameplay parameters.

I then extracted a set of variables to help characterize each attempt. Broadly speaking, I extracted two kinds of variables: observed and derived variables. *Observed* variables were characteristics of gameplay directly recorded by the EPIGAME software (for example, planning time). *Derived* variables were those discovered through logical tests or comparisons performed on observed variables, akin to a coding scheme. A total of 23 observed and derived variables were defined, each capturing an element or aspect of gameplay (see Appendix A for a more complete description of these variables). These 23 variables were selected on the basis of their ability to describe differentially the parameters for the forms of solution described in Table 5.

Phase 2: Generally, when using LA techniques it is most desirable to have a dataset with the smallest, most meaningful set of variables possible. Datasets with large numbers of variables are computationally very expensive to process, and such data is vulnerable to a variety of phenomena that distort results and complicate these types of analyses. In order to select only the most meaningful variables, I performed a Principal Components Analysis (PCA) on the dataset (16,239 attempts x 23 variables) using the *FactoMineR* software for *R* (Husson, Josse, Le, and Mazet, 2007). The PCA returned 3 components with eigenvalues greater than 1, with a total of

72.1% variance explained by those three components. The full results of the PCA are included in Table 6 (below). The variables associated with the components were:

1) Component 1:

   a) *tl.Modifys*, a count of how many modifications a student made to the parameters of placed Actions, e.g. changing a Boost from 10N to 20N increases *tl.Modify* by 1.

   b) *tj.Adds*, a count of how many Waypoints were added to the Trajectory.

   c) *planningTime.log*, the observed time students spend planning and placing elements, in seconds, logarithmically transformed to e.g. amplify the difference between planning times of 5 and 8 seconds but deemphasize the difference between 47 and 50 seconds.

   d) *eff.actions.added*, a derived variable counting how many new Actions were executed effectively on a given attempt compared to the previous attempt.

2) Component 2:

   a) *par*. A model-based effectiveness score, derived from a Markov-chain model of the combined series of outcomes of all the students' plays of each level. Each student generated a chain of Attempts of length $n_{\max}$ with 6 possible end states[3] x for each attempt n per level. The Markov-chain model simulates a memoryless random process following the observed transition probabilities in end states x among all the student-generated chains for that level. The model is then used to calculate the posterior probability of a Success state occurring randomly at the end of an (n-th + 1) Attempt given the state at the end of the n-th Attempt. These probabilities can range from [0,1], with 0 (i.e., no chance of success on the next attempt) being indicative of random play, and 1 (certainty of success in 1 more attempt) indicating expert play. In other words, the *par* metric asks, "if

[3] See Appendix A for description of the six end states as recorded in EPIGAME data logs.

this student were playing totally randomly, given that his or her last attempt ended in State *x*, what is the probability that he or she will find the Success Gate through sheer chance in one more attempt?" An important property of this metric is that it penalizes very long chains of Attempts and rewards navigating to the Success Gate on the first Attempt. The *par* score was later transformed into *par.sqrt* via a square-root transformation to make the probabilities more legible.

b) *par.delta.sqrt*. The change in value of the *par.sqrt* metric from attempt *n*-1 to *n* for the current level and student.

3) Component 3:

a) *is.abort* , an observed variable that tests whether or not the student manually aborted the attempt using the Abort button.

b) *fail.same*, a derived variable that tests, if an Attempt *n* was failed, whether or not a student failed that Attempt at the same place in the map as the (*n*-1)th Attempt *and* whether both Attempts failed for the same reason. A TRUE value indicates a consecutive unsuccessful attempt by a student to navigate past a specific obstacle on the map.

Table 6.
Results of the Principal Components Analysis

	Component				
	1	2	3	4	5
tl.Modifys	0.4746	0.0015	0.0000	0.0001	0.0014
tj.Adds	0.4173	0.0078	0.0000	0.0001	0.0000
eff.actions.added	0.1787	0.0000	0.0014	0.2471	0.0000
par.sqrt	0.0345	0.4694	0.0000	0.0000	0.0005
par.delta.sqrt	0.0001	0.7498	0.0000	0.0002	0.0005
planningTime.log	0.3297	0.0028	0.0000	0.0274	0.0000
fail.same	0.0854	0.0048	0.1082	0.0000	0.0890
is.abort	0.0000	0.0014	0.6434	0.0000	0.0220

Note: values are given as squared cosines

Further analysis revealed that since *par.sqrt* and *par.delta.sqrt* were linear combinations of each other, *par.sqrt* could be discarded in favor of *par.delta.sqrt*, which has the higher squared cosine for Component 2. At this point, further treatment of the data followed the line of inquiry specific to each research question. Relevant details can be found in their respective sections below.

Research Question 1: Can the Two Stances of the 2SM be Detected in Game Play Data?

The main claim of the 2SM is that the stances organize around fast- and slow-processing mechanisms; thus, it is reasonable to look for play strategies that embody fast and slow play. After the dimensionality reduction process (above), we are left with a manageable number of variables which are nonetheless theoretically significant and useful in describing these strategies. To explore Research Question 1, we next apply LA techniques exploring the variables in terms of clustering and then in terms of sequence mining. We then discuss the implications of the findings in terms Research Question 1 and the proposed 2SM framework.

Clustering

The next step in the analysis is to examine the dataset to determine whether students' play has some latent order or structure that can be brought into focus using our theoretically-relevant variables. To find this possible structure, I will use *clustering*, i.e. an unsupervised classification method. The goal of a clustering algorithm is to find the groups of observations whose features are more similar within-group than with regard to the data at large. Since this technique is unsupervised, I do not provide a pre-determined classification scheme for the software to "learn"; the rationale being that if a clustering algorithm returns a reasonably-interpretable set of clusters, and these clusters were created by interactions between theoretically-significant variables, then that is a solid indication that the theory describes latent structures of the data.

With the final list of seven variables already selected, I proceeded to create a similarity matrix using Gower's coefficient to account for the mixed data types. Then, I performed affinity propagation clustering with the resulting similarity matrix. Affinity propagation (AP) clustering is a clustering (i.e. unsupervised classification) method that takes as input measures of similarity between pairs of data points and simultaneously considers all data points as potential exemplars. Real-valued messages are exchanged between data points until a high-quality set of exemplars and corresponding clusters gradually emerges (Frey & Dueck, 2007). This method was selected as preferable to the more conventional k-means/k-mediods method because of its ability to produce a set of meaningful exemplars for each cluster – a vital consideration given the need to later interpret the characteristics of each cluster.

The AP clustering algorithm converged on a set of 145 “proto-clusters” after 260 iterations. These “proto-clusters” were then collapsed using an agglomerative method akin to hierarchical clustering. The resulting cluster dendrogram is given in Figure 11.

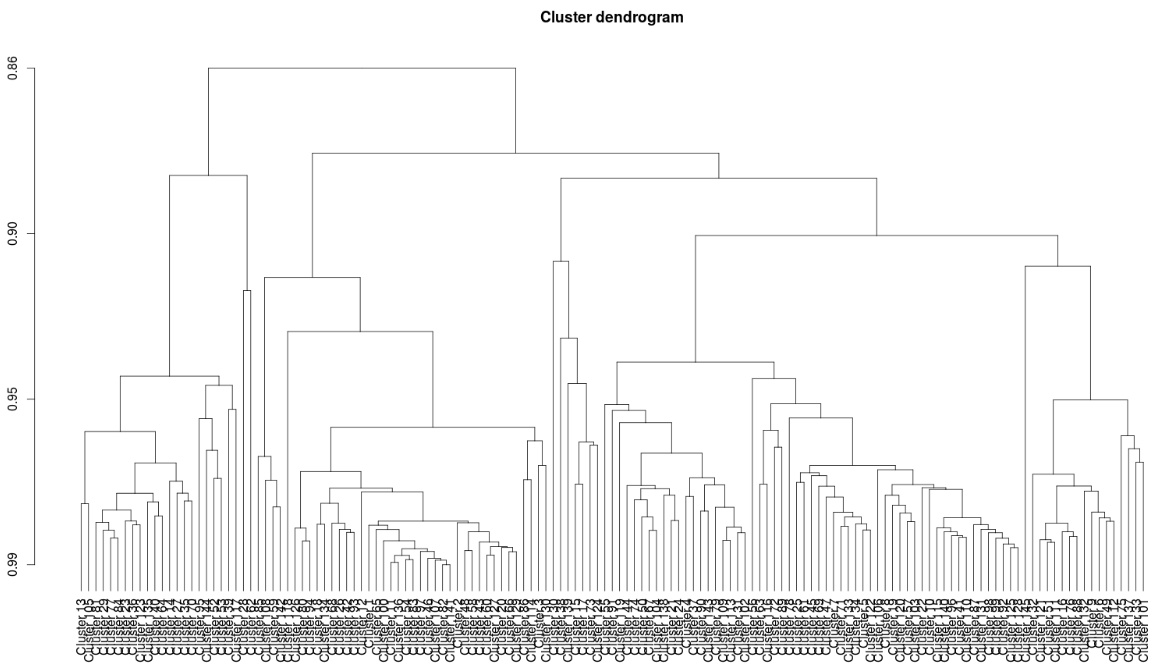

Figure 11. Cluster dendrogram of the AP clustering result.

Visual inspection of the cluster dendrogram suggested that a “cut” at 0.905 altitude would reduce the number of clusters to a manageable six. This clustering solution was codenamed *part.6.* The “goodness of fit” of an AP clustering solution is difficult to ascertain via standard methods (e.g. Rand coefficients) because AP clustering does not necessarily aim to produce compact clusters – rather to maximize the “representativeness” of the chosen exemplars. In order to determine the adequacy of the *part.6* solution, I created a heatmap from the similarity matrix (Figure 12).

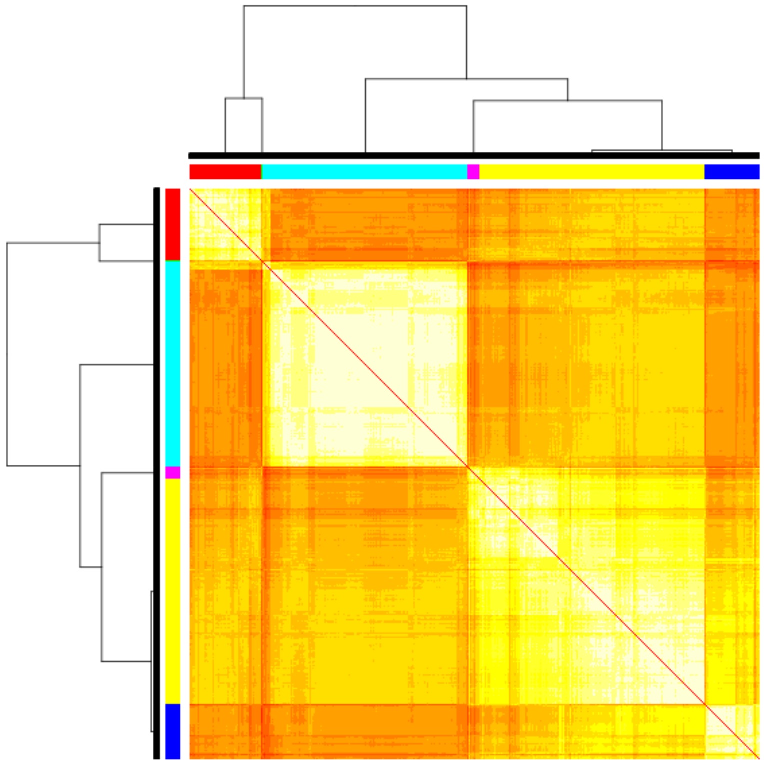

Figure 12. Heatmap of the similarity matrix, along with the dendrogram of the part.6 clustering solution.

The heatmap revealed 3 well-delimited and cohesive clusters along the diagonal, as well as one large cluster with some internal structure, and two additional smaller clusters. I iterated on the *part.6* solution several times in an attempt to resolve Cluster 2 (corresponding to the yellow region) into 3+ smaller clusters as suggested by the heatmap, but no satisfactory solution was found that preserved the other clusters, and thus the *part.6* solution prevailed. The distribution of Attempts across the six clusters of the *part.6* solution are given in Figure 13.

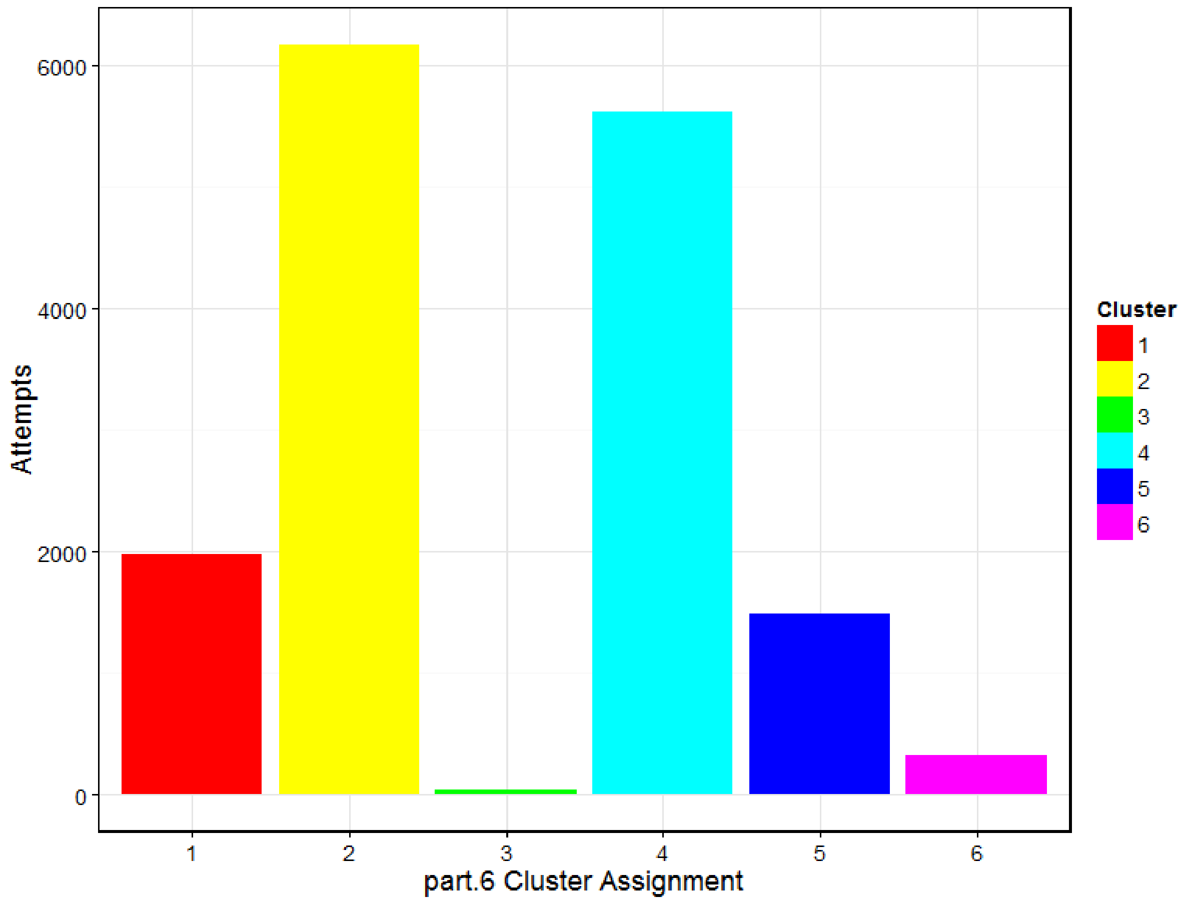

Figure 13. Histogram of the distribution of Attempts across part.6 clusters.

Before proceeding to the sequence mining, I studied the properties of the *part.6* clustering. As noted above, the preliminary variable reduction through PCA left us with only 7 theoretically-significant variables out of the original 26. The *part.6* solution represents a mathematical arrangement of students' attempts that have some similar structure in terms of these 7 variables. Figure 14, below, shows a generalized pairs plot (Emerson et al., 2012) that helps visualize how the structure of each cluster responds to each of the featured variables.

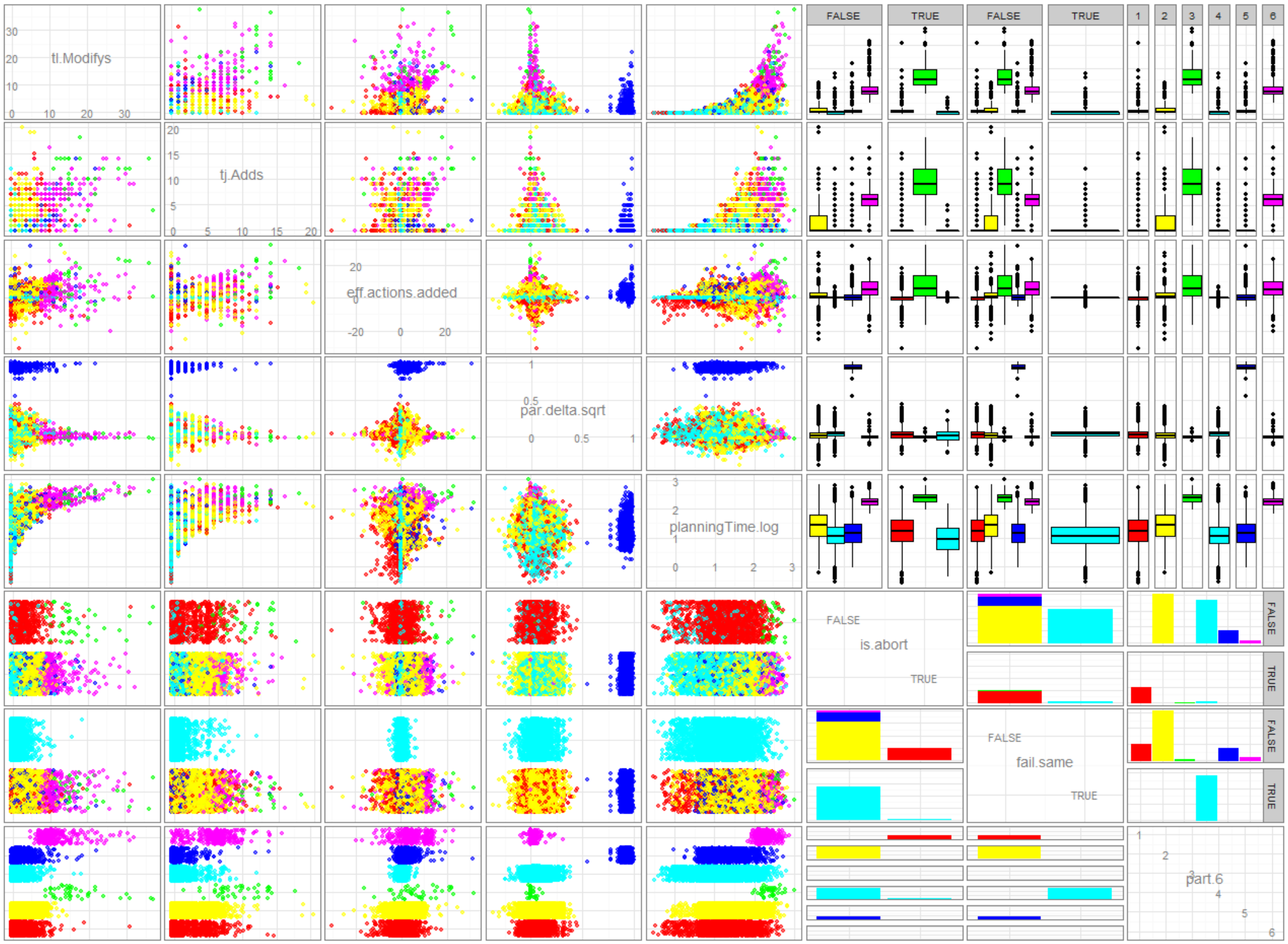

Figure 14. Generalized pairs plot (Emerson, et al., 2012) of the 7 theoretically-significant variables with the highest eigenvalues, plotted against each other, and classified according to the part.6 solution (rightmost column).

From each of the clusters I visually examined the exemplar chosen by the AP clustering algorithm, the two nearest neighbors to the exemplar, and two random members of that cluster. The 5 members of each cluster were compared, both by themselves and in context of the sequence of level attempts in which they occurred. Based on this analysis, I labeled the clusters qualitatively according to a general description of the students' actions therein:

- Cluster 1 (in red): ***ABORTS***. Students recognize that the level is going to fail and press the "Abort the Mission" button to preserve momentum of play rather than allow the simulation to end on its own.
- Cluster 2 (in yellow): ***TINKER***. Students add a few actions, advance a little further along in the level, and fail (but not in the same place in the map as the previous Attempt).
- Cluster 3 (in green): ***LONG ABORTS***. Very long planning episodes (> 100 seconds) that end in Abort. A very sparse cluster, barely distinguishable from Cluster 1. Possibly indicates a deletion and restart of the solved level in progress.
- Cluster 4 (in cyan): ***FUTILITY***. Students make a few changes, but fail exactly in the same place in the map against the same obstacle as their previous attempt.
- Cluster 5 (in dark blue): ***WINNING***. Students make one or more changes or additions that result in a successful attempt, thus completing the level.
- Cluster 6 (in pink): ***PLANNING***. Students spend a long time and add actions as well as trajectory elements (i.e., added both categories of elements). These attempts are occasionally successful, but not always.

Further investigation revealed that cluster assignments have some structure both in terms of *when* they occur in the order of play (i.e. early levels vs. later levels in Figure 9), and in terms of learning outcomes of the student that produced them (i.e., in terms of pre-post learning gains in Figure 16).

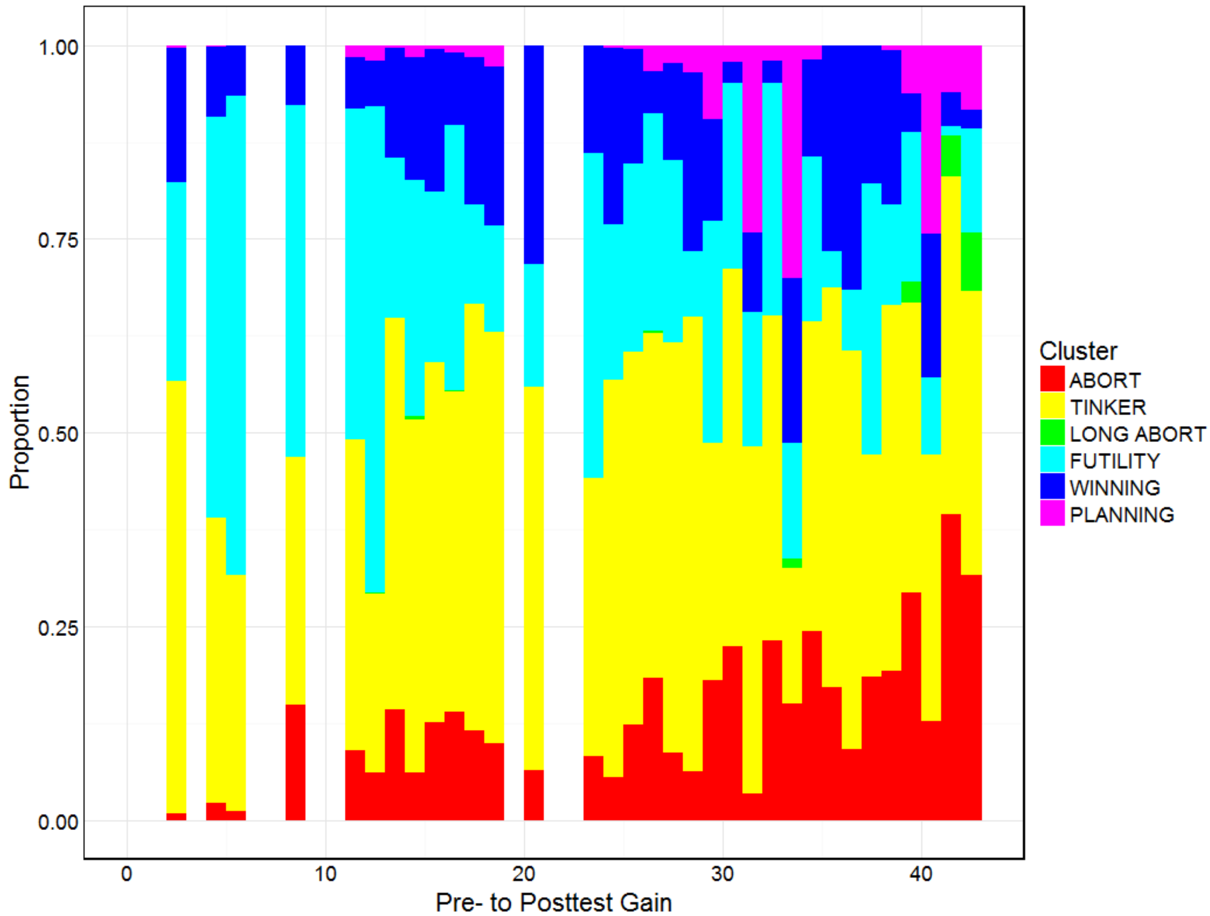

Figure 15. Frequency of part.6 cluster assignment by level. Blank columns represent steps without student interaction.

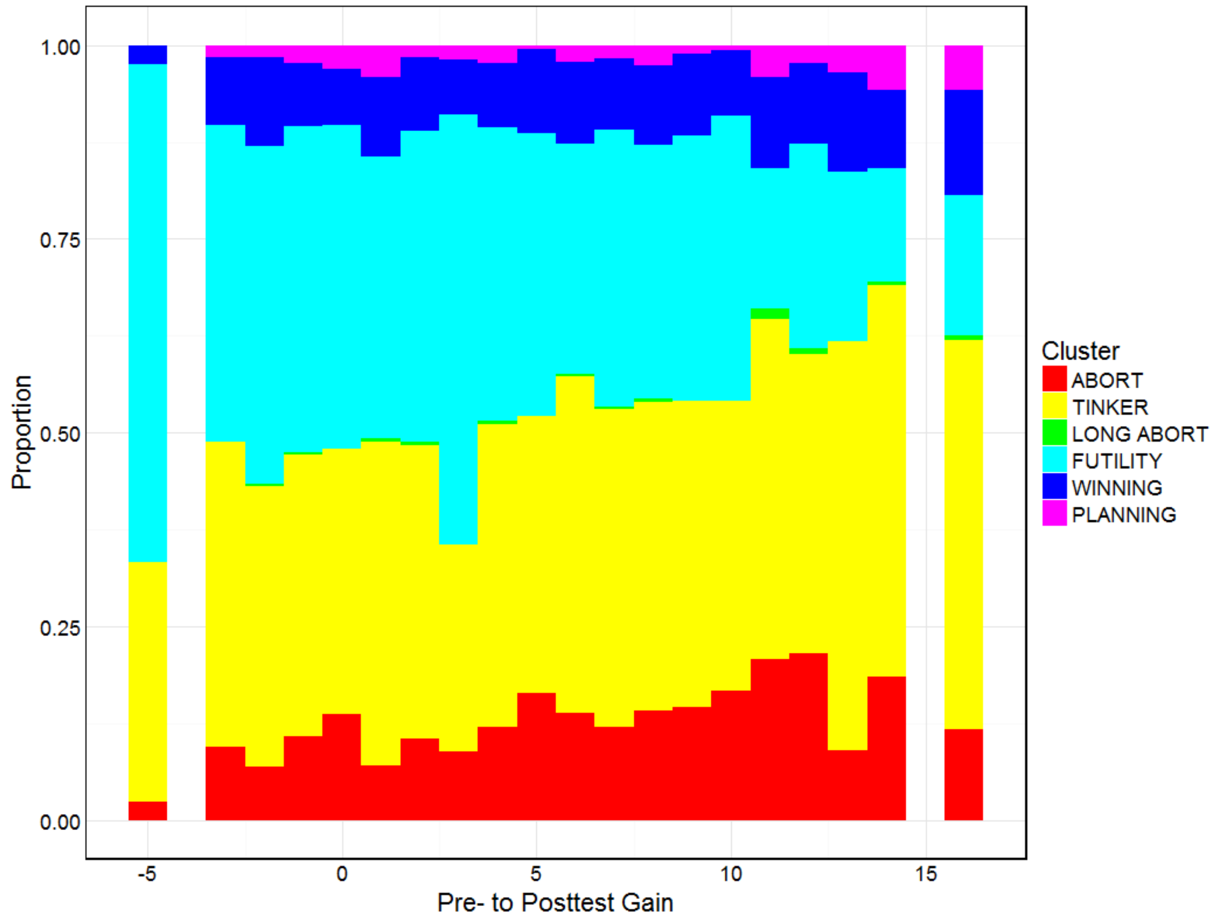

Figure 16. Frequency of part.6 cluster assignment by pre-post test score gain. Blank columns are unobserved gain scores.

As we can see, the relative distribution of the cluster assignments may be sensitive to the learning outcome of the student (Figure 16). In other words, the levels played by students at a given level of pre-post test performance may have different ratio of cluster assignments than those of students at a different level of performance. The different frequency profiles in Figures 5 and 6 suggest, furthermore, that the differences are not entirely due to how far students progress into the game. It is clear, then, that the *part.6* cluster solution provides not only a set of meaningful code assignments that describe students' play, but also that these assignments are related somehow to learning outcomes. Figure 10 further suggests that cluster patterns evolve as students progress through the game.

Sequence Mining

Sequence mining is a methodology intended to find patterns in sequence data, e.g. words in a sample of natural language or genes in a protein. The main requirement is that the order of the components be as significant, or more significant, than their frequency. The question sequence mining asks is, “given a set of items that form sequences, what are the most common smaller sequences to be found within and across those sequences?” In the case of EPIGAME data, the components to be sequenced are cluster membership codes; in other words, my goal is to investigate how students’ actions, described individually in general terms by the clustering procedure, appear in succession as a part of a chain of actions intended to solve a level.

The dataset contained 2730 such sequences, i.e. the students’ combined attempts to solve any level totaled 2730, an average of 22.2 levels per student. Each sequence was comprised of the series of each student’s attempts to solve a single level; thus, the length of these sequences ranged from 1 to 140 (i.e. the minimum and maximum number of consecutive attempts recorded in a single level). To perform the sequence mining, I used the *TraMineR* package for *R* (Gabadinho, Ritschard, Muller, & Studer, 2011). This package has the capability to calculate the relative importance of *subsequences* of elements within the element chains of sequence data. The relative importance of subsequences is measured not in terms of their frequency but in terms of "support", i.e. what proportion of sequences in the overall sample can claim a given subsequence as a subsequence of itself? The mining algorithm was configured to seek only first-order subsequences (i.e. only events that happen exactly consecutively are considered to be in sequence), and the minimum support level was set at 0.01. Thus to qualify for analysis, a subsequence would have to be supported by at least 27 sequences. An additional parameter was set so that the support of subsequences of *n* identical codes would be consolidated across all sequences found of one or more identical codes. The algorithm returned 47 candidate subsequences, which were then ordered by support. The results of the sequence mining are given in Figure 17, below:

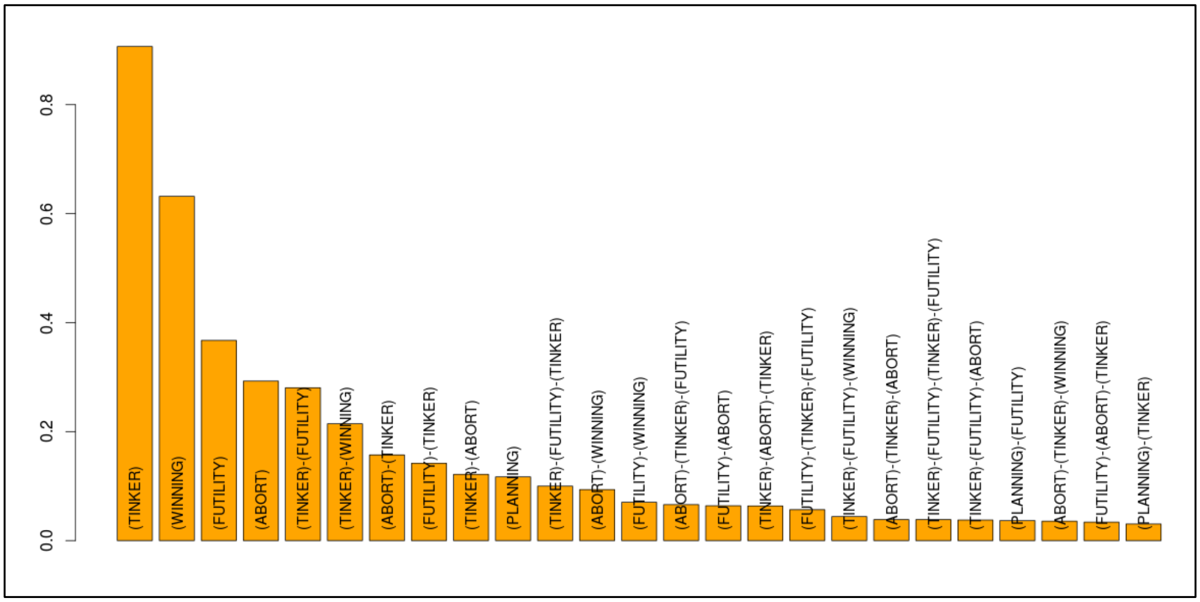

Figure 17. The 25 highest-supported subsequences. Words in parenthesis indicate cluster assignment of the sequenced items, following the part.6 solution (above).

The height of the bars in the graph indicate the support for that subsequence, and they are ordered by decreasing support. Support for the unitary subsequences, e.g. (TINKER), the most common one, are quite high since, for example, a sequence of (TINKER) - (TINKER) - (TINKER) can claim the subsequence (TINKER) a total of 6 times. Recalling that Cluster 2 stands for TINKER, thus, there are an above-random number of sequences containing long chains of TINKER, and similar above-random chains of FUTILITY. The high support value of WINNING is to be expected since 97% of all sequences end with WINNING (that is how students advance in the game, after all).

To investigate the relationship between play sequences and learning, I then classified students according to their pre-post test performance. Since the group of students as a whole gained significantly in their pre- to post- test scores, I chose a classification strategy that would qualify their gains relative to the group. The resulting classification scheme is summarized in Table 7 (below). The “High Prior” group consisted of students who scored in the upper quartile in both pre- and post-tests. The “Low Prior” group is likewise formed of students who scored in the bottom quartile of the pre- and post- test. A third group, “Learned” contains students whose pre-test scores were in the lower three quartiles but who improved their score by at least one quartile. A fourth group, “Null”, collected students whose pre-test was in the higher three quartiles but did not show a significant increase in their scores. The number of students in each classification was 14, 13, 23, and 54, respectively.

Table 7.
Classification of Students by relative pre-post gains

	Post-test score (quartile)		
Pre-test score (quartile)	1^{st}	$2^{nd} – 3^{rd}$	4^{th}
1^{st}	Low Prior	Learned	Learned
$2^{nd} – 3^{rd}$	Null	Null	Learned
4^{th}	Null	Null	High Prior

These assignments were used as discriminant groups, so that each detected subsequence’s support could be tested for correlation with learning outcomes via a Chi-square test. Table 8 contains the 18 subsequences with the highest Chi-square statistic; support for these subsequences thus varies by discriminant group in a statistically significant way. The graph of the resulting support values for each subsequence according to the student classification group is provided in Figure 18 (below).

Table 8.
Sequence analysis by discriminant group

			Support by Group			
Subsequence	*p*	Chi-Sq	High Prior	Learned	Low Prior	Null
(WINNING)	0.0000	115.7115	0.4229	0.5991	0.8068	0.6943
(FUTILITY)	0.0000	113.2765	0.1676	0.3251	0.5398	0.4332
(TINKER) – (FUTILITY)	0.0000	98.5601	0.1144	0.2391	0.4489	0.3351
(FUTILITY) – (TINKER)	0.0000	33.9205	0.0638	0.1254	0.2159	0.1667
(TINKER) – (FUTILITY) – (TINKER)	0.0000	32.1598	0.0293	0.0904	0.1477	0.1221
(PLANNING) – (WINNING)	0.0001	29.0500	0.0426	0.0087	0.0000	0.0089
(PLANNING)	0.0003	26.8821	0.1915	0.1181	0.0795	0.0971
(FUTILITY) – (TINKER) – (FUTILITY)	0.0006	25.2825	0.0133	0.0466	0.1023	0.0704
(FUTILITY) – (ABORT) – (TINKER)	0.0006	25.2800	0.0160	0.0364	0.1023	0.0357
(ABORT) – (TINKER)	0.0013	23.7677	0.0851	0.1516	0.2273	0.1738
(ABORT) – (TINKER) – (FUTILITY)	0.0021	22.7801	0.0213	0.0598	0.1193	0.0766
(FUTILITY) – (ABORT)	0.0051	20.8867	0.0372	0.0554	0.1364	0.0668
(ABORT)	0.0178	18.2497	0.2181	0.2770	0.3636	0.3164
(TINKER) – (ABORT) – (TINKER) – (FUTILITY)	0.0560	15.7918	0.0080	0.0204	0.0625	0.0321
(TINKER) – (WINNING)	0.0983	14.5475	0.1489	0.2259	0.2784	0.2193
(LONG ABORT)	0.1368	13.7978	0.0346	0.0160	0.0057	0.0089
(FUTILITY) – (WINNING)	0.1510	13.5705	0.0293	0.0671	0.0795	0.0847
(TINKER)	0.1528	13.5422	0.8617	0.9038	0.9432	0.9180

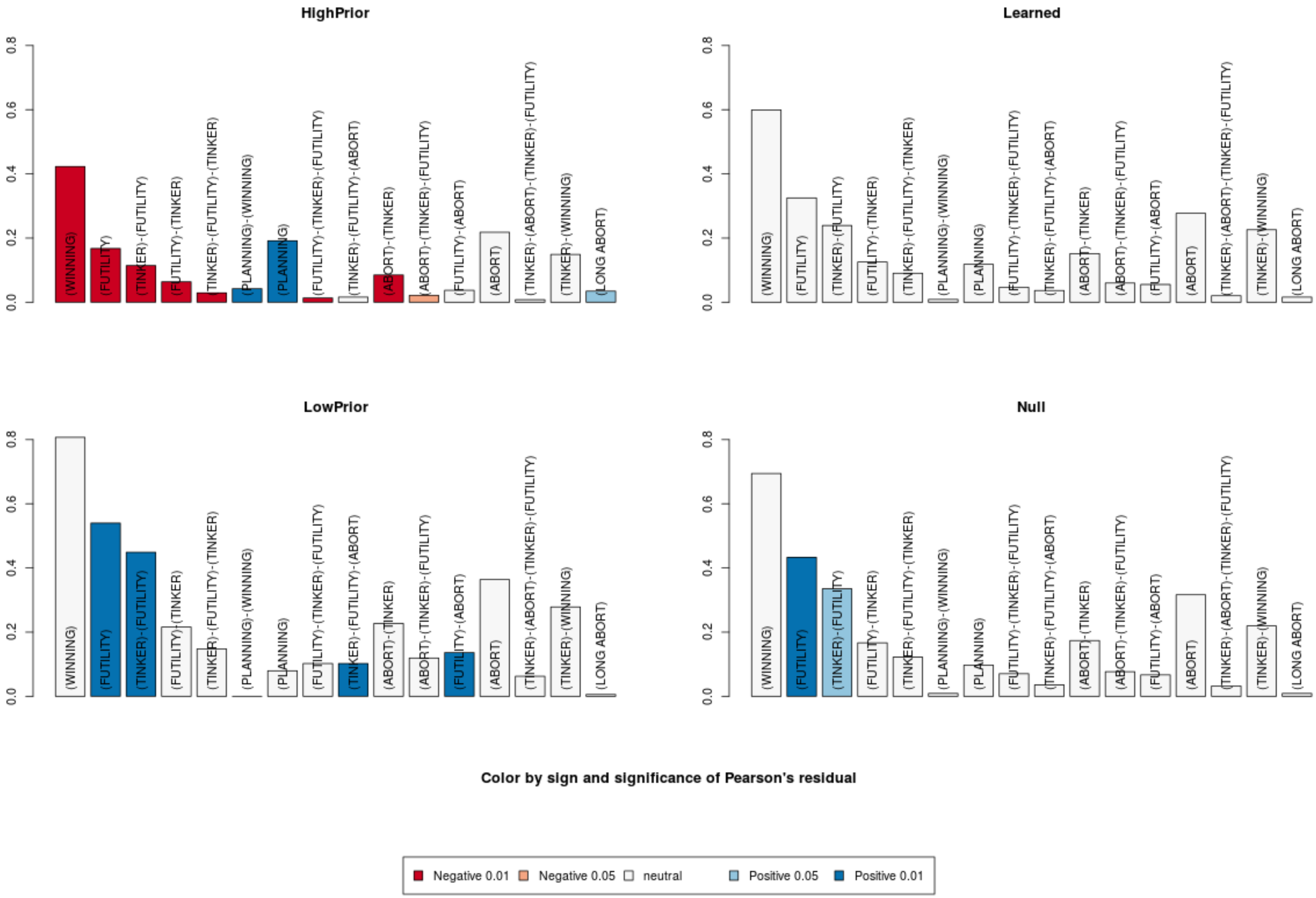

Figure 18. Sequencing analysis by group.

In Figure 18, red bars indicate subsequences with significantly less support than under the assumption of independence. Conversely, blue-colored bars indicates significantly more support. Sequences in white show no statistically-significant across all four groups. These significances are computed at the 0.01 level; light-blue and light-red bars indicate significance at the $p = 0.05$ level. For significance testing, the *p*-values were Bonferroni-corrected for the multiple comparison. This correction increases the probability of false negatives is compared to the probability of false positives, but protects against incorrectly rejecting the null hypothesis (i.e. that the support values for the subsequences do not vary across discriminant groups).

This group-discriminant sequencing analysis suggests that students in the High Prior knowledge group have sharply fewer FUTILITY subsequences, fewer TINKER-FUTILITY and FUTILITY-TINKER, ABORT-TINKER, TINKER-FUTILITY-TINKER, and FUTILITY-TINKER-FUTILITY cycles, and substantially more PLANNING chains. Conversely, students with Low Prior knowledge are more likely to present longer FUTILITY chains, and more TINKER-FUTILITY cycles. These students are also more likely to follow FUTILITY with ABORT, ostensibly because they recognize the probable outcome of that attempt would also have been FUTILITY. Students in the middle two quartiles who do demonstrate a relative increase in their conceptual understanding also show more FUTILITY chains and slightly more TINKER-FUTILITY chains.

Discussion

The Importance of Prior Knowledge. The sequence analysis reveals that students with High or Low Prior knowledge play very differently than their peers. Students who have High Prior knowledge plan more and exhibit very few sequences of attempts in which they are stuck. They are not as likely to attempt small iterative fixes, preferring more complex and thought-out solutions. On the other hand, if students consistently demonstrate repeated failure on the same obstacle over large numbers of attempts (i.e. in a FUTILITY sequence), it is less likely that they improved their learning, regardless of their level of prior knowledge. No particular way of playing (in other words, no subsequence exhibited by students in the *Learned* group) seems to correlate with relative learning gains independently of prior knowledge**.**

This finding suggests that students' gameplay choices are strongly influenced by their prior knowledge. It may be fairly argued that *High Prior* knowledge students played a very *different game* than their *Low Prior* peers. The former group approaches the game as a "planning game", preferring the creation of complete solutions that require only small adjustments, making full use of the Solve-and-Debug strategy hypothesized in the Student Model. The latter group likely sees the game as a "guess and check game" or "tweaking game", where a solution emerges gradually out of extended iterating cycles of more-or-less purposeful trial-and-error - described earlier as an Additive-Iterative strategy.

Why are students in the Low Prior group more likely to use the Additive-Iterative strategy? From the 2SM perspective, these students could be said to prefer low-effort, low-information, control-oriented processing strategies. The 2SM conceptualizes these as being closer to the Player Stance, which privileges feedback from the game environment to evaluate success. Students who play in the Additive-Iterative mode are more reliant on feedback from the game, since such feedback (rather than evaluation of internalized models) represents their main source of information about how the game operates. On the other hand, students who play the "planning game" can rely more on their own ability to visualize and predict how the game will respond to their input, and thus probably require less feedback from "tweaking" or "guessing and checking”. This distinction correlates well with the general descriptions in the Two-System Framework of the Player Stance and Learner Stance, respectively.

Implications for Design. The results of the gameplay data analysis from EPIGAME generally support the notion that patterns of play related to the Player Stance are not optimal for learning. Students who persist in fast strategies are not likely to improve their learning relative to their peers, and students who make the highest relative gains do not prefer fast strategies overall. The analysis shows that a tolerance or preference for Attempt sequences with a high reliance on FUTILITY are associated with lower learning outcomes.

In the 2SM framework, multiple FUTILITY attempts with low average time per attempt can be understood as a strategy for obtaining feedback from the game’s model as a way to avoid having to use slower, more intensive reasoning processes such as the second-order model. The goal of this strategy is to serve the player’s agency and sense of control, and preserve the momentum of play. It may be argued that use of the Player Stance helps students remain motivated and engaged even in the face of failure, and long after the novelty of the game has worn off. Yet, as we have seen, in the case of EPIGAME, the Player Stance and its associated play strategies are associated with lower learning gains. Then, the immediate question becomes, can the Player Stance be disrupted in order to promote learning? Or in the context of EPIGAME, can a student playing the “tweaking game” be nudged towards playing the game more as a “planning game”? Can a game be designed in such a way that this “nudge” occurs automatically?

In the case of EPIGAME, the tutorials might provide a clue as to how this “nudge” can occur early in play (see Question 2). Whatever the eventual form that this encouragement takes, the effectiveness of this feature depends on having a method to detect whether or not a student has settled in a Player Stance. This “detector” could be built upon the analysis here described:

the game could use a similar process of unsupervised clustering I used to arrive at the *part.6* solution as a guide to classify students' actions in real time, and then detect the sequences of play which, as we have seen, are not strongly associated with learning. This added functionality would allow specific feedback to be provided to students, e.g. early in their play before they commit to playing a "tweaking game" (c.f. Clark, Martinez-Garza, Biswas, Leucht, & Sengupta, 2012). Lastly, if the game can be made so that, once it has gathered enough student data to predict a student's play characteristics, the game can modulate its difficulty to make sure that the student faces challenges appropriate to the student's level of skill and knowledge, while compensating for the tendency of students to choose low-effort strategies if doing so preserves the momentum of play. These three additional functionalities could all be potentially very powerful ways to promote student learning with games, and they are all made possible by an expanded understanding of how students actually play.

These findings should also motivate discussion about how much and what kinds of support students should receive during game-based learning opportunities. Lower-performing students' over-reliance on fast strategies might be more of an adaptive response to being forced to play a game that is too difficult as opposed to an intentional strategy choice in response to their perceptions of what the game is about. In this case, automated feedback and adaptation as discussed above would also be useful. Students who are facing intractable difficulty could be detected and helped automatically. It is also possible that students perseverating in fast strategies are doing so transgressively (see Aarseth, 2007), i.e. as a rejection of the game's challenge and a personal disinvestment from the game's outcomes. This low-effort position is radically different from the low-ability position described above, but in terms of data logs it would look rather similar. The analytics used in this study are not well-suited to detect the difference between low effort and low ability, although some scholars have had success with specific detection algorithms for disengagement in the context of science simulations (Gobert, Baker, & Wixon, 2015).

In this study, as in much of classroom-based educational game research, I relied on pre-existing classroom norms for expectations on student behavior and effort. Also, the presence and expert eye of the teacher to help identify and gently correct students who were off-task, and to offer guidance to those few students who may have found the game too demanding, was indispensable. I observed and respected these practices while fully knowing that their effects would disturb the central assumption that the data logs record students' actions *and only students' actions*. This tension points to an inherent limitation of the data logging approach. Data logging can only account for what happens within the student-computer interaction, and classroom technology use often involves, or even privileges, person-to-person interactions. It is during these kinds of interactions that teachers (and often peers) help students make sense of the game when the game itself doesn't offer the necessary scaffolds, whether motivational or content-related. These interactions may have effects on participating students' play that would be captured by data logging but be difficult for LA techniques to correctly explain or attribute. It may be that future work that harnesses data log analytics for adaptive feedback might approach, or perhaps even duplicate, the classroom teacher's ability to identify apathy and helplessness in the classroom context, or the knowledgeable peer willingness to dispense timely hints. Until that time, however, we accept some imperfection and "mangling" of the record, and look for opportunities to more deeply integrate log-based analytics with observational and grounded methods.

Research Question 2: Effect of Changes in Students' Game Performance on Pre-post Test Performance

The main learning goal of EPIGAME is to help students build deeper understanding of Newtonian kinematics. Thus, a portion of the game's rules and systems deal with inertia and the relationship between force and velocity. Ideally, as students improve in their ability to solve inertial challenges, their conceptual understanding (per an external measure) should likewise improve. From the previous analysis (see Question 1 section, above), we know that students with different degrees of prior knowledge approach the game differently and play in sharply different ways. In terms of Question 2, I investigated whether these differences in game play correlate with differences in performance in the specific game situations intended to help students develop concepts of inertia.

The first step in in this analysis was coding the conceptual challenges. Each challenge is a situation on the game map where a student has to apply one or two maneuvers to advance past that situation. The selected challenges all deal with *inertia* and/or *Newton's second law of motion.* These concepts can be portrayed in EPIGAME in one of four ways:

1. The student must navigate Surge from rest up to a certain velocity by applying an unbalanced force (Figure 19). There are 46 such challenges, and they were coded as *fromStop.*
2. The student must bring Surge from a constant velocity to a stop by applying one or more forces opposed to the direction of motion (35 challenges, coded as *toStop).* (Figure 20)

3. The student must increase the velocity of Surge to a certain level while Surge is in motion by applying an unbalanced force in the direction of motion (4 challenges, coded as *speedUp).* (Figure 21)

4. The student must decrease the velocity of Surge to a certain level while Surge is in motion by applying an unbalanced force opposite the direction of motion (5 challenges, coded as *slowDown).* (Figure 21)

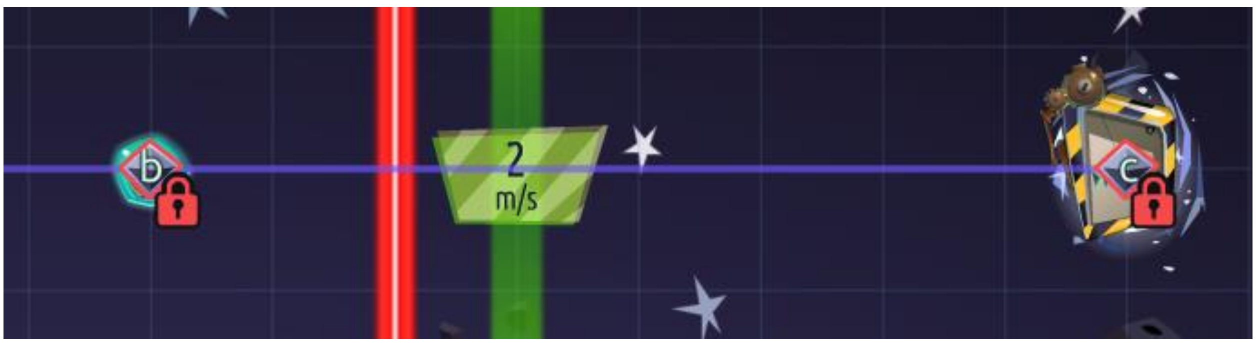

Figure 19. A fromStop challenge. Players begin motion from rest at point B and navigate toward C.

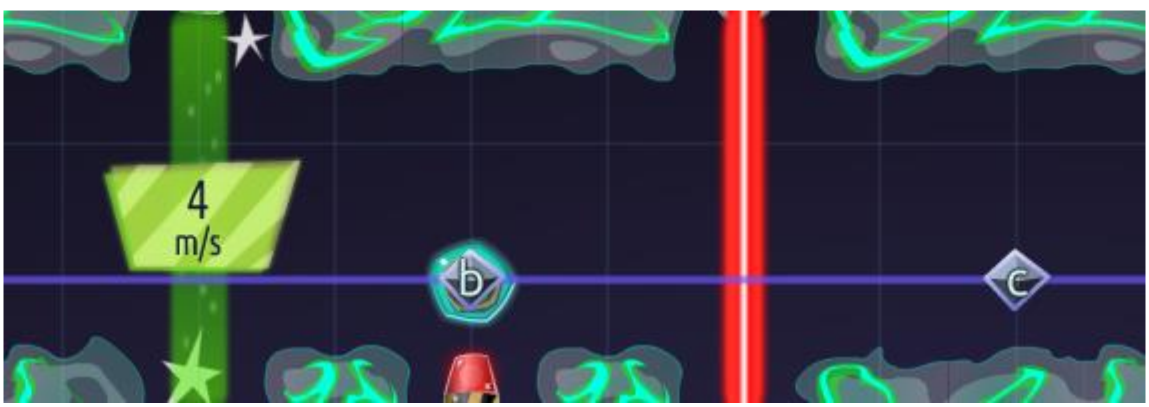

Figure 20. A toStop challenge. Players must completely stop at B before proceeding to C.

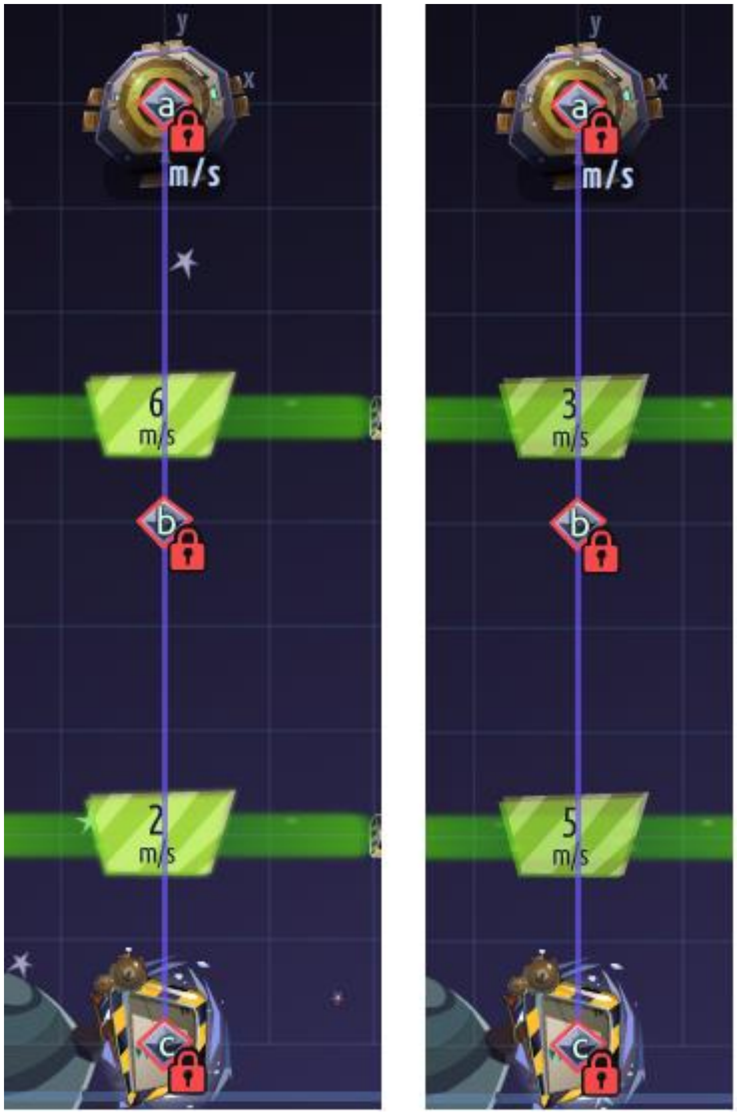

Figure 21. A slowDown (left) and a speedUp challenge (right). In both cases, student must apply an unbalanced force at B.

Each challenge was identified through visual inspection of the levels, its location and type recorded (*fromStop, toStop, speedUp*, or *slowDown*), and a consecutive serial number assigned. Only the first 90 challenges students encounter while playing EPIGAME were coded. The rationale for this limit is that the conceptual nature of these challenges changes in the latter levels, first when changes in mass are introduced, and then when students have to deal with forces applied in action-reaction pairs. Thus, the first 90 challenges students encounter before these increases in complexity are the most conceptually similar and can be safely compared.

Furthermore, these first 90 challenges are where we might be most likely to see trajectories of improvement because it tracks students from the beginning of the game where the learning curve may prove the clearest.

Then, the overall gameplay dataset was filtered through a conditional join in order to identify which attempts ended at one of the coded challenges. A total of 2175 attempts were identified. Later, I decided to reduce the sample to 1282 attempts corresponding to the first 15 challenges of each type, under the rationale that the unbalanced number of challenges per type (e.g. 46 *fromStop* vs. 4 *slowDown*) would likely lead to problems with the model fit if I used the challenge type as a covariate.

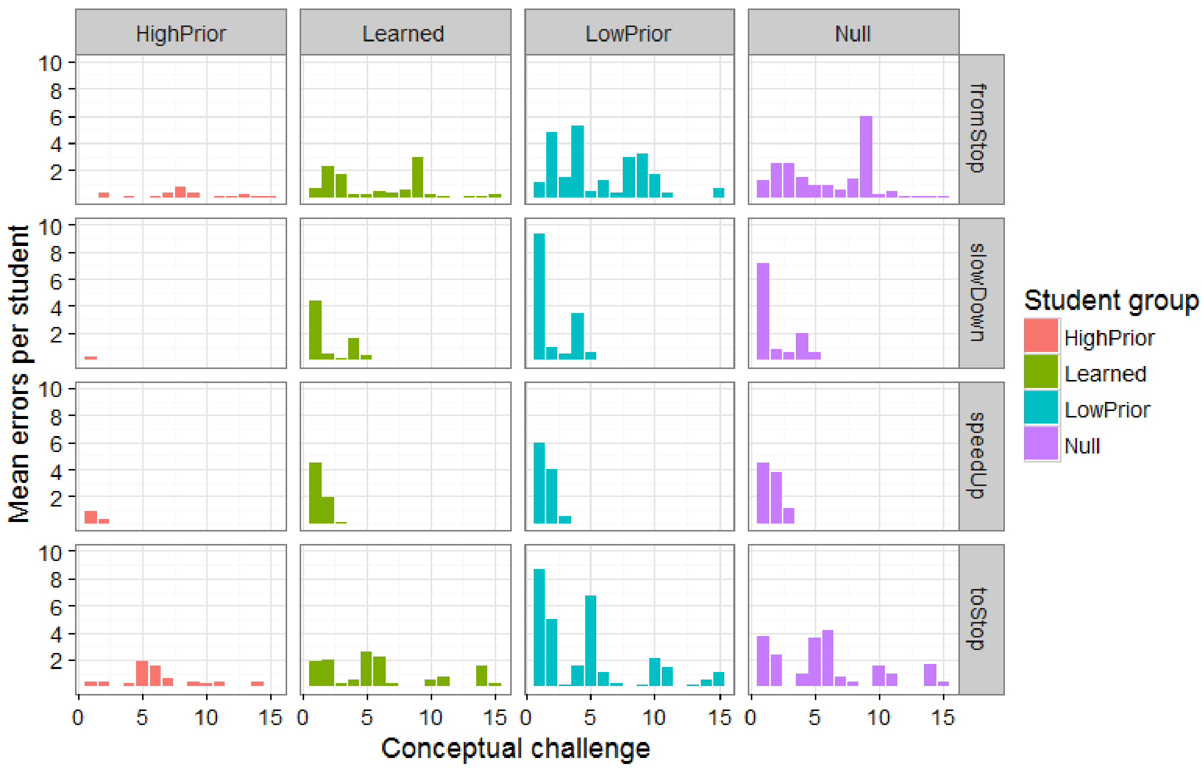

Figure 22. Mean errors per student per Conceptual challenge. Student achievement groups are in columns. Challenge types are in rows.

Next, I proceeded to fit a generalized linear model to the data. Since the dependent variable is a count (i.e. positive whole numbers only), then a Poisson regression would be most appropriate. However, the data showed considerable overdispersion, and thus a negative binomial regression was chosen.

Results

The statistics of the generalized linear model are provided in Table 9. In this model, the High Prior classification and *fromStop* challenge type are the model references. The statistically-significant predictors of student errors per Conceptual challenge are Challenge instance, and as noted, the type of challenge is not a statistically significant predictor. Furthermore, a previous iteration of the model showed that the interaction terms of the predictors were also not statistically significant. Thus, the variables best suited to predict the number of errors students commit are the number of similar challenges already faced and the students' prior knowledge.

Table 9.
Coefficients of the negative binomial regression model

	Dependent variable: number of errors per Challenge		
	Estimate	Std. Error	*p*-value
Challenge instance	-0.064^{***}	0.008	*>0.001****
Learned	0.484^{***}	0.148	*0.001****
Low Prior	0.744^{***}	0.153	*>0.001****
Null	0.654^{***}	0.140	*>0.001****
slowDown Challenge	0.112	0.090	*0,21*
speedUp Challenge	0.008	0.093	*0.93*
toStop Challenge	0.040	0.067	*0.55*
Constant	1.117^{***}	0.150	*>0.001****
Observations	1,282		
Log Likelihood	-3,221.669		
theta	1.371^{***}	0.066	
Akaike Inf. Crit.	6,459.338		
Note:	$^{***}p<0.01$		

Discussion

The generalized linear model fit to the challenge data confirms the (admittedly straightforward but encouraging) hypothesis that students tend to make fewer errors on a challenge each successive time they encounter a challenge of the same type. More surprising is that the *mean* number of errors can also be predicted on the basis of a student's prior knowledge grouping. In other words, the first 15 times students face challenges of a given type, students who score highly on the pre-test are likely to commit as few as half as many errors as students who did not score highly.

A possible explanation is suggested by the bar chart matrix on Figure 22. We can see there that students in the High Prior column make fewer errors overall, but more importantly, commit nearly no errors the first time they face a challenge of a given type. Students in other groups commit at least 3 errors on average, often more. Unless High Prior students have played EPIGAME before (which they have not), one could assume that High Prior students would make at least a few errors when they initially encounter a challenge, while they internally navigate how their understanding of physics does or does not apply to the situations and rules of the game. But the near total absence of errors on initial contact with challenge types suggests that High Prior students *already know* something directly relevant to these challenges.

There are at least two other sources of knowledge (besides any prior EPIGAME experience) students might be drawing on when they face new challenges. First, they may be drawing on inferences made from the pre-test. However, I demonstrated in Study 1 that EPIGAME and the EPIGAME assessment are free of testing effects (see Methods section), so a "priming" effect is unlikely. The other source might be the tutorial animations embedded in EPIGAME. There are two types of tutorials. At levels 1, 4, 8, 10 and 11, the tutorial animations are essentially *worked examples* – students watch as the Mentor character demonstrates, e.g. how to apply forces, how to draw Waypoints, how to start the trial, etc. These animations are intended to guide students as they learn the game's interfaces, design conventions, etc. On the other hand, the tutorials at levels 2 and 7 are *contrasting cases* (Figure 23). These animations take the form of experiments: a challenge is approached with several combinations of parameters, of which only one is correct. The student must deduce from this demonstration *why* that particular maneuver was effective. While the "worked example" tutorials show the *hows* of EPIGAME, the "contrasting case" tutorials show the *whys.*

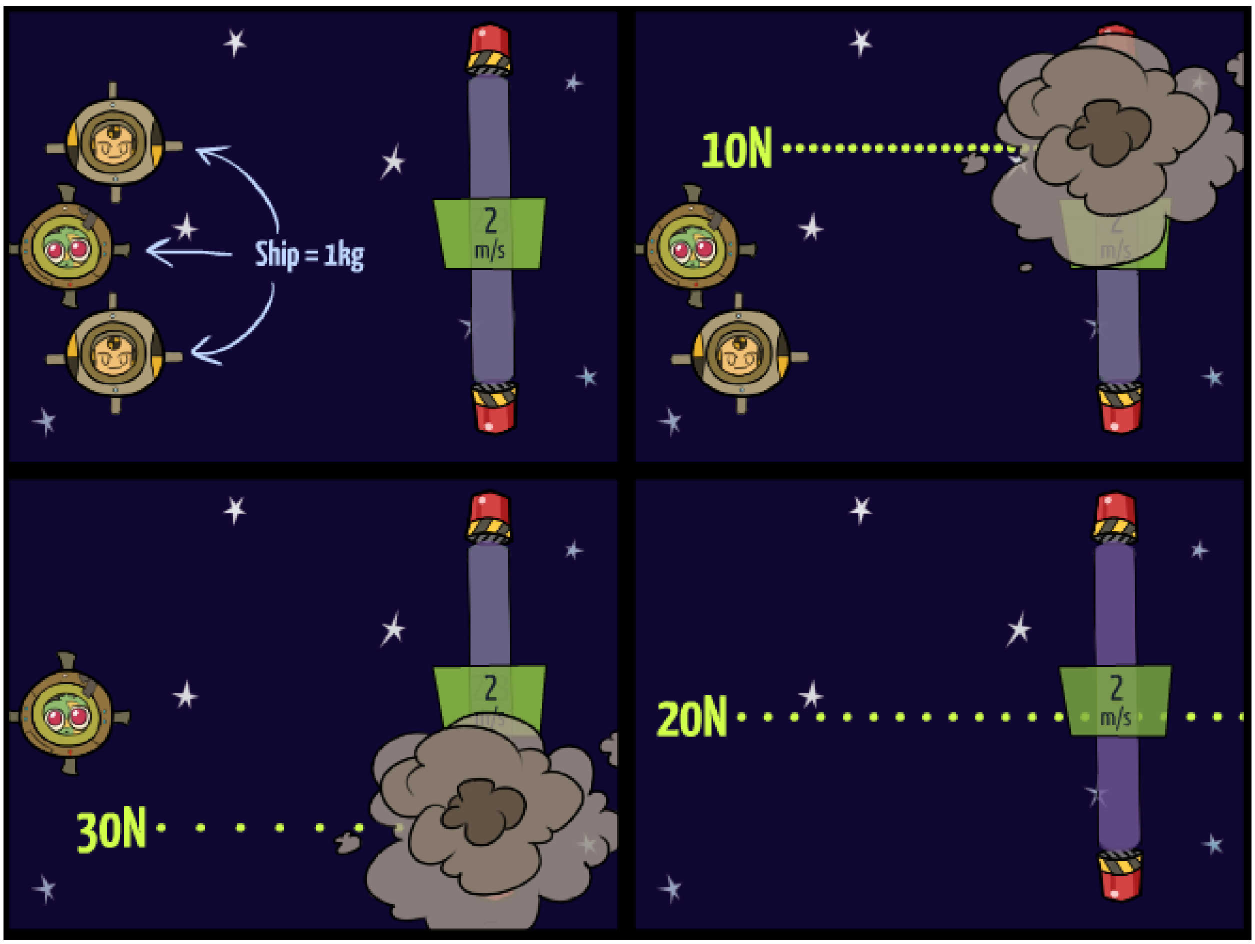

Figure 23. A "contrasting case" tutorial. The use of 10N and 30N are both incorrect for a 2m/s Velocity Gate.

My explanation for the low rate of error of High Prior knowledge students during initial trials relative to their lower-prior-knowledge peers is grounded in the 2SM. The 2SM defines two broad classes of knowledge regarding “how to play”: heuristics and internal models. The “worked example” tutorials, with their emphasis on how to execute specific maneuvers, have more “heuristicness” than “modelness”. Conversely, the “contrasting cases” tutorial focus strongly on the variables and relationships at play, suggesting more model quality. It may be that the main difference between High Prior students and their peers is *which form* of tutorial they chose to focus on. Since each form of tutorial primes a different form of knowledge about “how to play”, students with a strong preference for one form of tutorial over the other may approach the game with different kinds of knowledge and thus play in different ways. And in fact, these differing styles of play do emerge (see Question 1), with High Prior students (and only High Prior students) showing a marked preference for slow, deliberate play and small tolerance for error. In contrast, students in the Low Prior and Null learning groups prefer iterative, “tweaking” gameplay that is inherently more fast-paced, yet they tend to accrue errors at each challenge, often as many as 10, 20 or more (see Figure 22). In summary, it may be that the tutorials – necessary parts of the game experience – can “prime” the 2SM stances according to (a) the forms the tutorials take (prescriptive vs. descriptive) and (b) how salient and useful the player finds the information presented in the tutorials themselves.

Implications

In order to access the potential and intended benefits of educational games, students must first *learn to play* the game itself. This step, while commonsensical, can easily be glossed over during design; when it comes to introducing unfamiliar digital games into the classroom, we might hold the notion that young learners can simply “pick it up” and “figure it out”, since they may already be “gamers”. Thus, materials intended to help students orient themselves in the game environment and learn how to reach gameplay goals may not receive as much design attention as they should. Furthermore, when games are used in an educational setting, these materials compete for classroom time with the main game, where the target curricular material is most likely to reside. Ideally, we would prefer if students spend only a little time “learning to play” and as much time as possible simply “learning.”

Our findings problematize these design assumptions. First, by showing how prior knowledge can structure gameplay to a great extent (see Question 1). Students who enter the game experience with a good working knowledge of the concepts and relationships are less reliant on more feedback-rich and iterative, yet ultimately more laborious, “tweaking” styles of play. Also, the analysis suggests that the way the game teaches students to play, i.e. by following a procedure or operationalizing a relationship, may also be an important influence on learners, even when this learning is focused squarely on game-specific knowledge and not on curricular concepts and relationships.

If prior knowledge and differential use of tutorial materials can structure and influence play (and thus, learning), then a greater emphasis must be placed on game functionality that supports students who do not initially enjoy or leverage these advantages. For example, lack of prior knowledge can be addressed with *scaffolding*, and gameplay difficulty can be adapted to reduce repeated error. These and other measures should be considered as means to ensure that all students can access substantially similar game experiences and thus, hopefully, more equitable positive learning outcomes.

Overall Conclusions

On the Generalizability of the Current Methodology

The work described in this paper has followed a methodology that is not limited to investigating EPIGAME logs. The general methodology is versatile and feasible for use in other contexts. Starting from a robust and detailed record of students' interactions with a digital environment and a theoretical framework that supports conjectures as to why certain patterns of action create opportunities for the desired change, researchers can define the important features of those patterns and then use those features to investigate the data record using whatever LA techniques are most appropriate for that particular type of data.

A more novel focus of this analysis (one featured in Research Question 2) is that it aims to track development of students' conceptual understanding at the level of particular concepts of inertia using finer-grained observations centered on particular gameplay regions. These regions are intended to highlight specific content, and thus student performance in these regions is more closely tied to conceptual understanding than gross-level summative measures. These summative measures have been successfully used in the past and may be appropriate and sufficient for some research questions. However, the use of finer-grained contextual data offers the advantage of supporting claims of students' conceptual understanding of individual concepts (e.g., inertia or First Law), rather than broad performance constructs (e.g., knowing how to play EPIGAME).

Which is not to say that the EPIGAME data structure and focus (and thus the associated analysis) are universal. The trial-retrial structure of game play and the grain size of the data capture are not necessarily common to all educational games. The specific combination of play structure and grain size warranted the sequence mining and contextual feature mapping. Other digital environments will have different interactive structures, and thus algorithms and techniques possibly better suited to the questions being asked. Fortunately, the state of the art of learning analytics is increasing both the accessibility and variety of statistical computing software, making it suitable for a wider variety of data structures, game mechanics, and learning foci.

One thing that will likely remain invariant, however, is the expertise of the analyst and his or her familiarity with the context and the data. In this paper, my own long association with EPIGAME data, as well as observations accumulated over multiple opportunities to facilitate students' play of EPIGAME, facilitated the creation of the derived variables, the process of interpreting the *part.6* clustering, and the use of sequencing as a way to add meaning to the cluster assignments. It is unlikely that this kind of intimate understanding of the affordances and constraints of particular games and data can be substituted by generic software. Although it can, perhaps, be supplemented. Until that time, however, the skill of the analyst, as in all interpretative observational methods of research, will be crucial to success.

What the Findings Say about the 2SM

The 2SM is intended as a general-purpose framework for player cognition during game play; it is comprehensive and not intended as specific to any kind of game or any target domain. Because of this generality, it requires many constructs and mechanisms to explain phenomena of play. Furthermore, most of these constructs and mechanisms are entirely latent, existing only in the player's mind, perhaps only for brief moments of time. For these reasons, it is unlikely that a single study, however ambitious, could prove the 2SM as a theory.

The findings in this paper suggest that, while still very much unproven as a whole, the basic underpinnings of the 2SM pass muster. We see the indicia of both fast, low-information play and slow, deliberative play. More importantly, these styles of play co-vary strongly with learning outcomes, indicating that fast styles of play may not support students in developing knowledge of a form transferable beyond to the game (in this case focusing on Newtonian kinematics). Some students perseverate in guess-and-check iteration, relying entirely on the game to provide the necessary feedback, instead of using all available information to infer some generalizable rule they can use to increase their effectiveness. We can see from the Contextual Mapping analysis how some students never seem to stop making errors in parts of the game relating to a specific concept, even when they've already cleared a similar challenge 10 times or more.

That said, the finding that prior knowledge strongly influences play, even in the early stages, is problematic in terms of the 2SM. First, because it inverses the proposed way that Stances get cued. In the original framing of the 2SM, the Learner Stance is cued by a task that is too demanding, where the player has no fast effortless rule to apply. However, the results in this paper strongly support the claim that the opposite may be true, i.e. that (perceived) high task demands cue the Player stance as an effort-saving strategy that is ultimately maladaptive in terms of learning. The second challenge to the 2SM comes from the necessity of having students "learn to play" the game before they actually "play" it. This instructional phase, and its consequences, were not addressed originally in the 2SM. Yet as we have discussed previously, the Tutorial materials and other instructional affordances might bias students towards one form of reasoning or another, independently of how the student would otherwise organize his or her epistemic Stance.

These findings suggest that revisions of the 2SM are warranted in at least two lines. On one hand, the role of prior knowledge as an epistemic resource, largely ignored in the original framing. The 2SM envisions a player with well-defined goals for play but a "blank slate" in terms of pre-existing knowledge about the game. Further research that specifically targets the effect of prior knowledge, and of knowledge gleaned early in play from tutorial materials is warranted, and those findings integrated into the 2SM. Another possible revision involves the issue of task demands and their possible role in cueing other resources, such as mastery or performance orientations (Pintrich, 2000). The 2SM does not explicitly consider whether a player finds a given game situation “easy” or “difficult”; rather it only considers what epistemic resources the player has at hand (heuristics or second-order models). Yet the findings in this study highlight that Player Stance related patterns of play may also be a coping strategy to deal with game situations students find too difficult. For the 2SM to properly account for these coping strategies, a study might be designed where versions of the game of various difficulty levels are assigned to students at different levels of achievement, either by pre-test score or by an automated adaptive functionality. All that said, these revisions, while necessary and warranted, do not necessarily threaten the validity of the framework as a whole. They merely remind us that the Two-System Framework is, after all, young theory, and must co-evolve with the evidence, whether that evidence supports or resists it.

References

Aarseth, E. (2007). I fought the law: Transgressive play and the implied player. *Situated Play. Proc. DiGRA*, 24-28.

Annetta, L. A. (2010). The “I's” have it: A framework for serious educational game design. *Review of General Psychology*, *14*(2), 105.

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M.-T. (2009). Investigating the impact of video games on high school students’ engagement and learning about genetics. *Computers & Education*, *53*(1), 74–85. http://doi.org/10.1016/j.compedu.2008.12.020

Baker, R. S. J. D., & Clarke-Midura, J. (2013). Predicting successful inquiry learning in a virtual performance assessment for science. In S. Carberry, S. Weibelzahl, A. Micarelli, & G. Semeraro (Eds.), *User Modeling, Adaptation, and Personalization* (pp. 203–214). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-38844-6_17

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: applications to constructionist research. *Technology, Knowledge and Learning*, *19*(1-2), 205–220. http://doi.org/10.1007/s10758-014-9223-7

Braver, M. W., & Braver, S. L. (1988). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Psychological Bulletin*, *104*(1), 150–154. http://doi.org/10.1037/0033-2909.104.1.150

Buckley, B. C., Gobert, J. D., Kindfield, A. C. H., Horwitz, P., Tinker, R. F., Gerlits, B., … Willett, J. (2004). Model-based teaching and learning with Biologica^{TM}: What do they learn? How do they learn? How do we know? *Journal of Science Education and Technology*, *13*(1), 23–41. http://doi.org/10.1023/B:JOST.0000019636.06814.e3

Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Shaffer, D. W. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of Biomechanical Engineering*, *137*(2), 024701.

Clark, D. B., & Martinez-Garza, M. (2012). Prediction and explanation as design mechanics in conceptually-integrated digital games to help players articulate the tacit understandings they build through gameplay. In C. Steinkuehler, K. Squire, & S. Barab (Eds.), *Games, learning, and society: Learning and meaning in the digital age*. Cambridge, Mass: Cambridge University Press.

Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M. M., Slack, K., & D’Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education*, *57*(3), 2178–2195. http://doi.org/16/j.compedu.2011.05.007

Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education*, *42*(3), 309–328. http://doi.org/10.1080/15391523.2010.10782553

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*(4), 253–278. http://doi.org/10.1007/BF01099821

Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience* (First Edition). Harper Perennial.

Dede, C. (2011). Developing a research agenda for educational games and simulations. In S. Tobias & J. D. Fletcher (Eds.), *Computer Games and Instruction* (pp. 233–247). Charlotte, NC: Information Age Publishing.

Dondlinger, M. J. (2007). Educational video game design: A review of the literature. *Journal of applied educational technology*, *4*(1), 21-31.

Emerson, J. W., Green, W. A., Schloerke, B., Crowley, J., Cook, D., Hofmann, H., Wickham, H. (2012) The Generalized Pairs Plot. *Journal of Computational and Graphical Statistics, 22* (1), pp. 79-91.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. http://doi.org/10.1146/annurev.psych.59.103006.093629

Foster, A., & Mishra, P. (2008). Games, Claims, Genres and Learning. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education.* (Vols. 1–3, pp. 33–50). Hershey, PA: Information Science Reference.

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, *315*(5814), 972-976.

Gabadinho, A., Ritschard, G., Müller, N. S., Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software, 40*(4), 1-37. Retrieved from http://www.jstatsoft.org/v40/i04/.

Gee, J. P. (2007). What Video Games Have to Teach Us About Learning and Literacy. Second Edition: Revised and Updated Edition (2nd ed.). New York, NY: Palgrave Macmillan.

Geertz, C. (1973). Thick Description: Towards an Interpretive Theory of Culture. In *The Interpretation of Cultures* (pp. 3–30). Basic Books.

Gijlers, H., & de Jong, T. (2013). Using Concept Maps to Facilitate Collaborative Simulation-Based Inquiry Learning. *Journal of the Learning Sciences*, *22*(3), 340–374. http://doi.org/10.1080/10508406.2012.748664

Gobert, J. D., Sao Pedro, M. A., Baker, R. Sj., Toto, E., & Montalvo, O. (2012). Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. *Journal of Educational Data Mining*, *4*(1), 111–143.

Halverson, R., & Owen, V. E. (2014). Game-based assessment: an integrated model for capturing evidence of learning in play. *International Journal of Learning Technology*, (Special Issue: Game-Based Learning). Retrieved from http://www.inderscienceonline.com/doi/abs/10.1504/IJLT.2014.064489

Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *The Journal of the Learning Sciences*, *12*(1), 53–90.

Harpstead, E., Myers, B. A., & Aleven, V. (2013). In search of learning: facilitating data analysis in educational games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 79–88). New York, NY, USA: ACM. http://doi.org/10.1145/2470654.2470667

Hou, H. T. (2012). Exploring the behavioral patterns of learners in an educational massively multiple online role-playing game (MMORPG). *Computers & Education*, *58*(4), 1225-1233.

Husson, F., Josse, J., Le, S., & Mazet, J. (2015). *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining*. Retrieved from http://CRAN.R-project.org/package=FactoMineR

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, *93*(5), 1449–1475.

Kerr, D., & Chung, G. K. W. K. (2012). Identifying key features of student performance in educational video games and simulations through cluster analysis. *Journal of Educational Data Mining*, *4*(1), 144–182.

Lamb, R. L., Annetta, L., Vallett, D. B., & Sadler, T. D. (2014). Cognitive diagnostic like approaches using neural-network analysis of serious educational videogames. *Computers & Education*, *70*, 92–104. http://doi.org/10.1016/j.compedu.2013.08.008

Leighton, J. P., & Gierl, M. J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, *26*(2), 3–16. http://doi.org/10.1111/j.1745-3992.2007.00090.x

Linehan, C., Kirman, B., Lawson, S., & Chan, G. (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1979–1988). New York, NY, USA: ACM. http://doi.org/10.1145/1978942.1979229

Martinez-Garza, M. M., Clark, D., & Nelson, B. (2013). Advances in assessment of students' intuitive understanding of physics through gameplay data. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, *5*(4), 1–16.

Martinez-Garza, M., Clark, D. B., & Nelson, B. C. (2013). Digital games and the US National Research Council's science proficiency goals. *Studies in Science Education*, *49*(2), 170–208.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, *2003*(1), i–29. http://doi.org/10.1002/j.2333-8504.2003.tb01908.x

Mislevy, R. J., Oranje, A., Bauer, M. I., von Davier, A., Hao, J., Corrigan, S., … John, M. (2014). *Psychometric considerations in game-based assessment.* (white paper). Retrieved from http://www.instituteofplay.org/work/projects/glasslab-research/

Parnafes, O., & Disessa, A. (2004). Relations between Types of Reasoning and Computational Representations. *International Journal of Computers for Mathematical Learning*, *9*(3), 251–280. http://doi.org/10.1007/s10758-004-3794-7

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of educational psychology*,*92*(3), 544.

Plass, J., Homer, B., & Kinzer, C. (2014) *Playful Learning: An Integrated Design Framework.* (White paper #02/2014). Games for Learning Institute. doi: 10.13140/2.1.4175.6969

Prensky, M. (2006). *Don't bother me Mom - I'm learning!: how computer and video games are preparing your kids for twenty-first century success - and how you can help!*. Paragon House. Retrieved from http://www.worldcat.org/isbn/1557788588

Reyna, V. F., & Ellis, S. C. (1994). Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science*, *5*(5), 275–279. http://doi.org/10.1111/j.1467-9280.1994.tb00625.x

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development*, *44*(2), 43–58.

Rowe, E., Asbell-Clarke, J., & Baker, R. S. (2015). Serious games analytics to measure implicit science learning. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious Games Analytics* (pp. 343–360). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-05834-4_15

Sengupta, P., Krinks, K. D., & Clark, D. B. (2015). Learning to deflect: Conceptual change in physics during digital game play. *Journal of the Learning Sciences*, *24*(4), 638–674. http://doi.org/10.1080/10508406.2015.1082912

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. doi:10.1037/0033-2909.119.1.3

Shaffer, D. W., Squire, K. D., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan*, *87*(2), 104.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer Games and Instruction* (Vol. 55, pp. 503–524). Charlotte, NC: Information Age Publishing. Retrieved from http://pdf.thepdfportal.net/PDFFiles/6536.pdf

Shute, V., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. Cambridge, MA: MIT Press.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, *46*(2), 137–150. http://doi.org/10.1037/h0062958

Squire, K. D., DeVane, B., & Durga, S. (2008). Designing centers of expertise for academic learning through video games. *Theory Into Practice*, *47*(3), 240–251.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Tobias, S., & Fletcher, J. D. (2007). What research has to say about designing computer games for learning. *Educational Technology*, *47*(5), 20-29.

Vellido, A., Martin-Guerroro, J. D., & Lisboa, P. (2012). Making machine learning models interpretable. *Proceedings of the 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN). Bruges, Belgium*, 163–172.

Ventura, M., Shute, V., & Small, M. (2014). Assessing persistence in educational games. *Design Recommendations for Intelligent Tutoring Systems*, 93.

Ventura, M., Shute, V., & Zhao, W. (2013). The relationship between video game use and a performance-based measure of persistence. *Computers & Education*, *60*(1), 52–58.

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, *59*(10). Retrieved from http://courses.had.co.nz.s3-website-us-east-1.amazonaws.com/12-rice-bdsi/slides/07-tidy-data.pdf

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., … Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research*, *82*(1), 61–89. http://doi.org/10.3102/0034654312436980

Appendix A

Observed and derived variables in the EPIGAME log data

Table 10.
Catalog of variables in the Study 2 dataset

Variable name	Meaning	Type	Notes
Student ID		Identification	Anonymized to a serial number
Experiment ID		Identification	
Date and Time		Identification	
Step Visit	Number of times student has visited that Level (step)	Observed	
Attempt		Observed	Only Attempt = 1 was used
attemptTrial	Order of this Attempt within a series of Attempts (i.e. a Trial)	Observed	
totalTrials	Combined number of Attempts in all Trials of this Level by this student	Observed	
endState	Did the player succeed (=1), fail (=0), or abort (=2)?	Observed	
endScore	Score obtained by that student at the end of that Trial	Observed	
scoreImproved	Did the student increase their Score this Attempt?	Derived	
trialTime	Length of time a between this Attempt and the end of the previous Attempt	Observed	Incorrectly named in software, should be “attemptTime”
actionsUsed	How many Actions were placed on Waypoints during the Planning Phase	Observed	
isExit	Did the student leave the Level after this Attempt?	Observed	
timeLine	Position of the time cursor on the Timeline at level end	Observed	More relevant in the Timeline version of EPIGAME. Students in the present studies did not have access to the time cursor.
attemptTrial.max	Maximum value of the variable attemptTrial for that student for that level	Derived	

p.attemptTrial	Measure of **p**rogress of Attempts within the Trial	Derived	Calculated as (attemptTrial / attemptTrial.max)
p.totalTrials	Measure of progress of Attempts within the combined chain of Attempts over all Trials	Derived	Calculated as [attemptTrial + (sum of all attemptTrial.max of all previous trials) / totalTrials]
ending.event	The state of the game that caused the Level to end.	Observed	Allowed states: Success Gate, Navigation Error, Mass Gate collision, Velocity Gate collision, Laser collision, Abort.
ended.at.action	Number of actions that fired successfully	Observed	
Par	Model-derived metric of effectiveness.	Derived	See "Treatment of EPIGAME logs" for complete description.
planningTime	Duration of the Planning Phase	Observed	
tl.Adds	Addition of Actions to the Timeline	Observed	
tl.Deletes	Deletion of Actions from the Timeline	Observed	Very rare (mean = 0.05 deletions per Attempt)
tl.Modifys	Modification of parameters of Actions already in the Timeline	Observed	
tl.Moves	Actions moved within the Timeline	Observed	Very rare (mean = 0.17 moves per Attempt)
added.Tl.Total	Sum of Timeline Adds, Deletes, Modifys and Moves	Derived	
tj.Adds	Waypoints added to the Trajectory	Observed	
tj.Modifys, tj.Moves, tj.Deletes	Analogous to the Timeline (prefix: tl.) count variables	Observed	These variables exist in the record but no instance of these types of events were recorded.
locX, locY	Coordinates of Surge's spaceship when an event or Action occurred	Observed	
fail.same	Did this Attempt fail at the same location, for the same reason?	Derived	

eff.actions.added	How many Actions fired in this Attempt compared to the number that fired in the preceding Attempt?	Derived	Could be negative.
par.delta	Difference in the par metric between this Attempt and the previous one	Derived	
par.sqrt, par.delta.sqrt	Square-root transformations of the par and par.delta variables	Derived	
is.abort	Did the student press the Abort button before the level otherwise ended?	Derived	Software also registers the Abort button press in the endState variable.
ActionLog	Combined variable that registers the Actions applied to Surge during the Attempt	Observed	Includes position, type, and location of each Action applied
EventLog	Combined variable that registers important moments of gameplay not caused by Actions	Observed	Not fully functional in this version of EPIGAME
Serial	Serial number of the Conceptual challenge	Derived (from ActionLog)	
failed.to	In case of failure of a Conceptual challenge, the specific action the student did not do	Derived (from ActionLog)	Possible values: fromStop, toStop, speedUp, slowdown
is.colinear	Does this Conceptual challenge also require students to execute a turn?	Derived (from ActionLog)	
constant.mass	Do students have to account for changes to Surge's mass during the Conceptual challenge?	Derived (from ActionLog)	Only Conceptual challenges that pass this test were analyzed here
Pre, Post	The student's pre- and post-test scores, respectively	Observed	
bin.1	Classification of students according to beginning and ending quartile in assessment score	Derived	See "Question 1: Sequence Mining" for detailed description of this classification