# Strategy Detection in Wuzzit: A Decision Theoretic Approach

Truong-Huy D. Nguyen, Shree Subramanian, Magy Seif El-Nasr, Alessandro Canossa
Northeastern University, 360 Huntington Avenue, Boston, MA 02115
tru.nguyen, sh.subramanian, m.seifel-nasr, a.canossa@neu.edu

**Abstract:** During the past few years there has been an explosion on the number of games that are used for learning. Recently, there has been an emphasis on determining what players learn when they play a game, how they problem solve or make decisions and how such decisions progress through time showing some aspect of learning. This paper focuses on uncovering player learning behavior over time – specifically targeting progression analysis. We use a commercial game called *Wuzzit Trouble*, developed to teach players mathematical thinking through simple game mechanics of collecting keys to free a cute creature. We developed a new algorithm based on decision theory to investigate how players played the game over time and how they shifted their own strategies. The contribution of this paper is a new analysis approach towards understanding how players progress through the game, and how such progression can reveal insights on their learning process and goals they hoped to achieve.

## Introduction

In the past decade or so, research in games and learning has examined different ways in which individuals learn through games and effective ways in which games can be used within and outside traditional classroom settings to teach critical and academically valued skills, like writing and literacy (Magnifico, 2012), computational thinking and programming (Kafai, Fields & Burke, 2010; Durga, 2012), scientific thinking (e.g. Steinkuehler & Chmiel, 2006; Gaydos & Squire, 2012) and, more recently, engineering design skills (Mayo, 2007; Chesler et al., 2013). The main challenge of understanding learning within game contexts is that game environments are playful environments, vastly different from classroom settings where knowledge is disseminated and tested through examinations. Specifically, in a classroom environment assessment relies on tests, where test takers usually have a single intent, and that is to maximize their scores. In such situations, it is straightforward to interpret participants' behaviors. In contrast, learning games present participants with novel experiences filled with choices of unknown consequences; thus, it is much harder to determine the hidden motivation that drives participants' behaviors and how that may or may not lead to learning (Young et al., 2012; Shute, 2011; Shaffer et al., 2009). Players may take actions that help them understand the game's rules and mechanics instead of optimizing any specific utility function, especially during the early stage of playing. We thus distinguish between two types of behaviors, which have been discussed by previous literature (Jensen 2013; Cohen et al., 2007), *exploration* and *exploitation*. Exploration is exhibited when players select actions to gain a better understanding of their possibility spaces within the game environment (Jensen 2013). Exploitation behavior is observed when the players already have a good idea of how the game operates and aim to seek pleasure in beating the game while maximizing some utility function (e.g. score or fun). In fact, when facing a dynamic or unknown environment, people usually have to deal with the so-called *exploration-exploitation trade-off* in order to gain the highest long-term benefit (Cohen et al., 2007; Leonard et al., 2007). Moreover, as there is no time constraint in playing a game, players are enticed to explore because that allows them to enjoy their discoveries better in later stages (Carstensen et al., 1999). Due to these iconic patterns of behavior in playful environments, learning and assessment techniques cannot be applied as-is to learning games.

In order to model and understand learning in such a game environment, one needs to model and understand behaviors and choices made over time specifically capturing the exploitation vs. exploration theory (Cohen et al., 2007). It should be noted that while games are designed with finite mechanics, the behavioral outcomes and their combinations over time are not a closed space this is due to the stochastic and Markovian nature of learning in game environments. In other words, we believe that the meaning (specifically learning) of a behavior X in a situation Y cannot be interpreted based solely on the current situation, but rather we need to take into account the behavior sequence preceding that behavior and the situations of each of these behaviors (Baker et al., 2009).

In this paper, we adopt the reinforcement learning perspective (Sutton, 1999) to make inferences about players' decision-making process and motivations in their moment-to-moment actions. Using reinforcement learning, we model player strategies in a commercial Math game called *WuzzitTrouble*, with the objective of revealing possible player motives for the actions that they take in the game. The resulting model is a domain-independent mechanism to associate the raw action data with exploratory-exploitative patterns. We will show how players' action data viewed from this perspective helped us understand players' progression and quitting behavior, and reveal players' problem-solving and decision making strategies that have been extensively theorized to be critical in developing fluency and competence in mathematical problem-solving (Schoenfeld, 1992; 1985).

## Related work

Works in learning assessment in games and complex interactive environments fall in to two categories, (a) evidence-based assessment techniques that focus on design of assessment frameworks and task designs to indicate content learned and (b) user modeling and user-adapted AI techniques that seek to model patterns in learning and participant strategies. Evidence-based automated learning assessment techniques use prescribed task and competence or skill models, apply them to an open-ended environment, thereby scoping it, and determining whether or not participants succeed (or how well they succeed) in completing these tasks, so as to assess the strength of the evidence in what and how well the participants learned something (Messik, 1994; Mislevy & Hartel, 2006). Such assessment techniques can work extremely well in cases where evidence from each action can cumulatively amount to total learning, such as in many computer-based tests, like GRE or GMAT. In games, to some extent, clusters of skills and abilities may be inferred from player actions, thereby amounting to the evidence for learning, such as in *stealth* assessment design (Schute, 2011). The strength in such evidence-centric approaches lie on their ability to demonstrate alignment of skills or competencies to possible player actions (Schute et al., 2008; Rupp, et. al, 2010). However, these approaches work best when competence and action mapping are unambiguous. In other words, these evidence-centric approaches need to make (or enforce) definitive assumptions about player motives, e.g., excel or succeed, and by doing so they assume that players pursue a certain set of strategies aligned with the designers' intent. As we have argued earlier in this paper, this may not always be the case, particularly in games where choices and motives are linked rather too closely and tend to shift during play.

In parallel, there have also been several studies in learner modeling or profiling, which is the idea of grouping learners based on a pre-established screening algorithms that are derived either from static questionnaires (Tzouveli et al., 2008) or a limited set of pre-determined choices that participants are allowed to make in problem-solving, for instance see early work on cognitive tutors (Anderson et al., 1995; Corbett & Anderson, 1985). More recently, learner models are derived through data-mining and using decision trees to determine learner characteristics (Lin et al., 2013). The focus is on learners' strategies when immersed in open-ended environments with fairly unlimited or a large set of choices and models are used to detect strategy differences across individual participants or patterns of problem solving in large open-ended space (Berland et al., 2013). Such approaches that seek to *detect* player strategies, or salient patterns of play allow for more in-depth analyses on how player strategies shift over time (Desmarias & Baker, 2011; Baker, et. al. 2013). Some studies use unsupervised learning techniques that take in large raw or unlabeled data and detect learning patterns through clustering on sets of aggregated behavior data identifying "clusters of behaviors" (Berland & Martin, 2013; Drachen et al., 2009). However, a limitation with such techniques is that they provide an overview of the population's behavior and operate on aggregated data, which inherently imposes limitations for modeling individuality, i.e., nuances in individual actions are washed out (Recker & Piroli, 1995). In order to account for individual differences in strategies, some player-modeling approaches connect hypothetical implicit playing or learning factors (such as motivations, beliefs, intentions, desires or skillfulness and competency) to actual action observations. These models provide a means to logically infer the players' implicit states from their specific observed actions. Such models, while extremely flexible because they can operate both on aggregated and raw data, require domain knowledge (to build hypotheses) and extensive testing (to validate hypotheses) before they can be used reliably. This imposes a non-trivial overhead on human labor. Our technique marries these two approaches, providing a domain-independent method to contextualize raw action data using a well-tested theory of human behavior: decision theory. That is, we seek to model a learning phenomenon such that it can predict or be used to '*mine'* for relationships that are more useful for examining complex relationships between actions and motives for strategies. Similar examples for possible analyses of complex relationships include looking at how impact of moderating design decisions impact student behavior over time, or looking for related sub-categories of learners that can benefit from similar learning material, but differently (Baker & Yacef, 2009).

## Modeling *WuzzitTrouble* using a reinforcement learning perspective

Sutton explains reinforcement learning as

> "learning what to do, how to map situations to actions so as to maximize a scalar reward signal. The learner is not told which action to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them" (Sutton, 1999).

From this viewpoint, a strategy in a game is a way to select actions so as to maximize a certain utility or reward function. Several goals may describe how a said level in the game was completed and thus, being able to predict consistent strategies from moment-to-moment player actions. This would allow us to calibrate player expertise or ability to play the level. For example, a player who seeks to acquire the highest possible score will try to complete as many sub-quests as possible to obtain most points possible. Through this strategy, the player consciously values obtaining points in a level over quickly passing it. In contrast, a player who seeks to minimize the time spent in each level will focus only on the bare minimum set of quests needed to move on to the next level. Therefore, while the game mechanics and the reward structure of the game stays the same, players actions ought to be interpreted differently because different actions are prioritized due to adoption of different strategies. Normalizing player actions would likely suppress individual differences in play styles. In the following sections, we describe briefly, the game mechanic, how we modeled the game system based on two hypothesized player strategies, and subsequently we show how identifying these strategies allowed us to interpret player behavior.

### Brief description of the game *WuzzitTrouble*

*WuzzitTrouble* is a commercial game, released as a free download by the startup company BrainQuake in fall 2013. The game is designed to provide arithmetic-based puzzles with increasing difficulty in a fashion that circumvents the usual symbolic notation of arithmetic in an effort to break the symbol barrier, a widely known obstacle in arithmetic problem solving (Devlin, 2013). The goal of the game is to free creatures called Wuzzits from their traps by collecting all the keys in a level. The keys hang on to a large wheel that is rotated such that the position of the key aligns with the marker at the top. Note, in Figure 1 below, the marker is at number '0' and needs to be moved to number '20' and also to number '50' to obtain both the keys needed to free the trapped Wuzzit. To align the marker with the said number, players rotate the large wheel clock-wise or anti-clockwise. The distance or number of units moved by the large wheel depends on the gears (smaller cogs beneath the large wheel, with numbers 3, 8 &13 marked on them). For example, if a player rotates the gear numbered 3 clockwise, the marker moves 3 pegs to the left; or if the player rotates the gear marked 13 anti-clockwise, the marker moves 13 pegs to the right.


Figure 1. Wuzzit Trouble, Stage 2, Level 3 (Image used courtesy of BrainQuake Inc.)

Each small cog can be turned up to five times to generate a five-step turn of the wheel, offering up to five opportunities to collect a key (or other item) with a single move. This is a critical gameplay mechanic to learn in order to free the Wuzzit with the smallest number of cog rotations such that the player beats the level with most stars, from a range of one, two or three stars. In addition, the pegs with unique gems attached to them, like the green gem between 25 and 30, in the figure above, rewards the player with bonus points. At higher levels, there are items that deduct points from the players, so for players who value high points, they should avoid these items as much as possible. The overall score at each level is determined by obtaining the keys and the bonus rewarding gems with least possible moves.

### WuzzitTrouble In-Game Player Action Data

WuzzitTrouble has been instrumented such that the moves made by players in each level are recorded in the database with tags differentiating different player sessions uniquely. Players' personal information, such as names, is obfuscated to protect their privacy. Gameplay logs from the telemetry allow us to simulate how the game was been played in one unique session. The logs include information about player performance, like the number of points and stars earned at the end of the level and moment-to-moment player actions, like all the moves made in a level and the items collected in that level with each move.

## Analysis

By modeling and solving each game level as sequential action planning problems, we can obtain the ideal, and thus optimal, behavior patterns with respect to two main strategies that we think are prominent within Wuzzit: Score Maximizer (SM) and Level Passer (LP). The solutions then allow us to impose a ranking order on all actions available at each state. The values, namely Q-ranks, capture how *optimal* each action is when executed at each state as compared to the ideals of strategies SM and LP. Since the conformity of a player's selected moves to a strategy's optimal actions, if happening consistently, is an indication of the strategy's adoption, Q-ranks provide a means to track how a player adopts strategies throughout the play session. For example, a sample output after applying the Q-ranking procedure in a level goes as follows

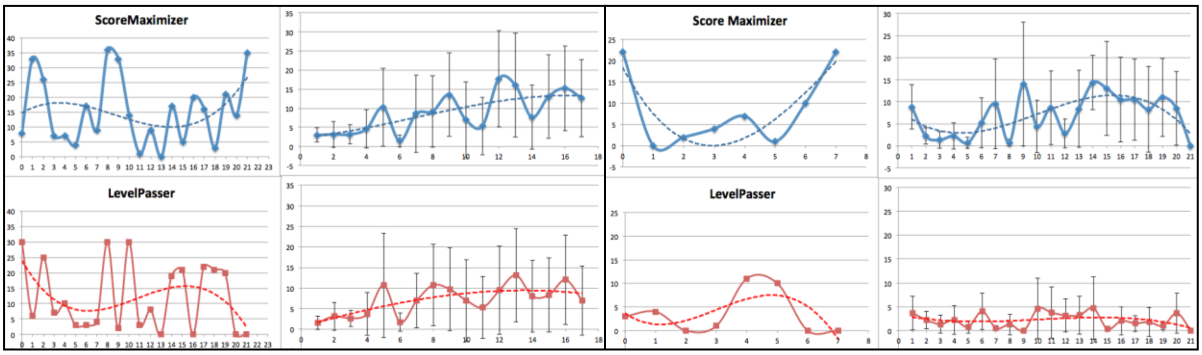
```
ScoreMaximizer:     Avg = 8.25; Q-Ranks = [3, 11, 1, 18, 0, 3, 9, 21]
LevelPasser:        Avg = 1.62; Q-Ranks = [1, 4, 3, 0, 1, 4, 0, 0]
```

The average number is simply the average of the series of Q-ranks that follow. In the above example, the player's Q-ranks show that her behavior is closer to a Level Passer than a Score Maximizer. Additional statistical test can be applied to arrive at a statistically significant conclusion in terms of strategy proximity.

## Interpretation

With the above analysis of player behavior, each player's moves can be interpreted in relation to known strategies. This allows us to form hypotheses about possible developments of a player's strategies over time, both intra-level and inter-level, some of which are described below.

**1. Explore then quit:** Figure 2a depicts the strategy proximity of Player 14091's behavior in a randomly selected level (in this case, level 16 in her session) and in all levels played before quitting. Observe that in the levels leading to the end of the session, the average proximity of her moves to both strategies (right column of Figure 2a) is large and has high variance. The lack of convergence on any strategy shows that the player was still in an exploration mode when she quitted. The quit can be explained as being the result of her frustration or boredom in the game. When zooming in onto a specific level (left column of Figure 2a), we can see a similar pattern of moves, in which the player was selecting actions that are not optimal in any strategy, before deciding to end the game quickly by selecting actions similar to those of a Level Passer. The fact that the player was almost never able to solve a level optimally with respect to any strategy hints that the player might not even discover the strategies.




(a) (b)

Figure 2. The Q-ranks of (a) player 14091 and (b) player 17380; the y-axis denotes Q-rank values. The first column in each subgraph depicts Q-ranks in One level while the second shows average Q-ranks in each levels of their session with error bars being standard deviations; dotted lines are respective trend lines of the graphs.

**2. Converge to LP then quit:** Figure 2b depicts the same examination done on Player 17380's play session. Overall, the player made moves that are closer to those of an LP than SM. There seems to be a period (level 10 to 14, Figure 3b) in which the player attempted exploration by deviating away from LP's optimal moves (high variance), but eventually gave up and fell back to LP before quitting the game.

**3. Exposed to both strategies then quit:** Figure 3 shows the analysis on Player 2280. Notably, this player had both moments in which her moves are optimal in SM (moves 2, 3 in Figure 4a and level 11, 12 in Figure 4b), and moments when it is clearly that she was just LP-ing (level 17). At the end of the session, this player had been exposed to both strategies.

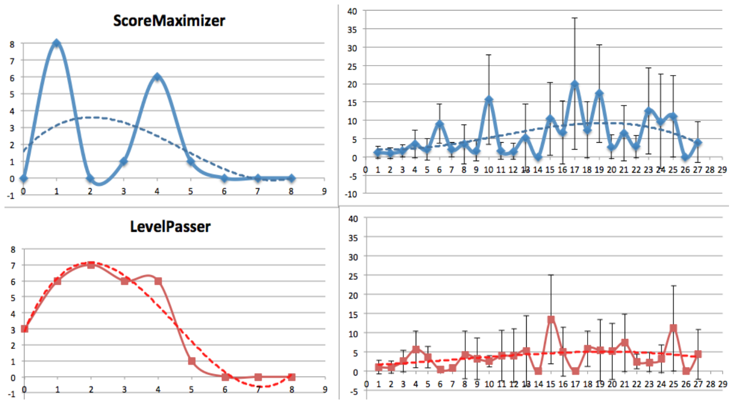



Figure 3. Q-ranks of player 2280

## Concluding remarks and implications for future directions of the Study

In this paper, we presented a decision-theoretic approach to model players' in-game behavior. The analysis opens new doors towards understanding different phases in the playing pattern, leading to insights that can in turn be leveraged to improve the playability of the existing game. In the future, we want to pursue the following directions:

1. Detection of exploration/exploitation transition point: A player in the process of exploring the game does not adhere consistently to any specific strategy yet. If we plot the proximity of their moves in relation to strategies' expected moves, the result is that players in exploration often select actions that are rarely close to any strategy over a long period of time.
2. Detection of quitting patterns: By observing the player's behavior, in relation to exploration and exploitation, towards the end of his/her play session, we can learn quitting patterns to answer the following questions: Does the player stop playing upon frustration (i.e. quitting while exploring) or due to boredom (quitting while exploiting)? If this analysis can be done in real time while the player is still playing, we can adjust the game accordingly to avoid their upcoming frustration or boredom.
3. Difficulty analysis: By analyzing quitting patterns of all players, we will be able to obtain the current difficulty level of the game to its audience, i.e. whether most players quit while exploring or exploiting. This insight can then be fed into the next iteration of game development so that the game's difficulty can be adjusted suitably to raise player retention.
4. Pedagogical intervention: In relation to learning games or games with serious purposes, detection of adopted strategies allows the instructors to timely intervene each individual students' learning process should they are detected to be embarking on a wrong path.

## References


Baker, C. L., Saxe, R., and Tenenbaum, J. B. (2009). Action understanding as inverse planning. Cognition, 113(3), 329-349.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. JEDM-Journal of Educational Data Mining, 1(1), 3-17.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. International Journal of Human-Computer Studies, 68(4), 223-241.

Bellman, R. (1957, 12). A Markovian Decision Process. *Indiana University Mathematics Journal, 6*(4), 679-684. doi: 10.1512/iumj.1957.6.56038

Carstensen L.L, Isaacowitz D, Charles S.T (1999). Taking time seriously: a theory of socioemotional selectivity. Am. Psychol. 54, 165–181.

Chesler, N. C., Arastoopour, G., D'Angelo, C. M., Bagley, E. A., & Shaffer, D. W. (2013). Design of a Professional Practice Simulator for Educating and Motivating First-Year Engineering Students. Advances in Engineering Education, 3(3).

Cohen, J. D., McClure, S. M., and Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1481):933-942.

D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. Cognition & Emotion, 25(7), 1299-1308.

Devlin, K. (2013, March). The Music of Math Games. *American Scientist*, Volume 101, Number 2, Page: 87 [http://www.americanscientist.org/issues/pub/the-music-of-math-games]

Desmarais, M. C., & d Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. User Modeling and User-Adapted Interaction, 22(1-2), 9-38.

Drachen, A., Canossa, A., Yannakakis, G.N. (2009). Player modeling using self-organization in Tomb Raider: Underworld. IEEE Symposium on Computational Intelligence and Games (CIG 2009), 1-8.

Durga, S. (2012). Learning to mod in an affinity-based modding community. Learning in Video Game Affinity Spaces, New York: Peter Lang, 84-102.

Gaydos, M. J., & Squire, K. D. (2012). Role playing games for scientific citizenship. Cultural Studies of Science Education, 7(4), 821-844.

Gee, J. P. (2008). Learning and games. The ecology of games: Connecting youth, games, and learning, 3, 21-40.

Jensen H. G. (2013). Making Sense of Play in Video Games: Ludus, Paidia, and Possibility Spaces. Eludamos. Journal for Computer Game Culture. 2013; 7 (1), pp. 69-80

Leonard, N. E., Paley, D. A., Lekien, F., Sepulchre, R., Fratantoni, D. M., & Davis, R. E. (2007). Collective motion, sensor networks, and ocean sampling. Proceedings of the IEEE, 95(1), 48-74.

Lin, C. F., Yeh, Y. C., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: an application of decision trees. Computers & Education, 68, 199-210.

Magnifico, A. M. (2012). The game of Neopian writing. Learning in video game affinity spaces, 212-234.
Mayo, M. J. (2007). Games for science and engineering education. Communications of the ACM, 50(7), 30-35.
Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher*, *23*(2), 13-23.
Mislevy, R. J., & Haertel, G. D. (2006). Implications of Evidence-Centered Design for Educational Testing. Educational Measurement: Issues and Practice, 25(4), 6-20.
Nunes, T., Carraher, D., Schliemann A. (1993). *Street Mathematics and School Mathematics*, Cambridge University Press.
Recker, M. M., & Pirolli, P. (1995). Modeling individual differences in students' learning strategies. The Journal of the Learning Sciences, 4(1), 1-38.
Schoenfeld, A. H. (1985). Mathematical problem solving. New York: Academic press.
Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. Handbook of research on mathematics teaching and learning, 334-370.
Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., ... & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning.
Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You Can't Fatten A Hog by Weighing It–Or Can You? Evaluating an Assessment for Learning System Called ACED. International Journal of Artificial Intelligence in Education, 18(4), 289-316.
Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. Computer games and instruction, 55(2), 503-524.
Steinkuehler, C., & Chmiel, M. (2006, June). Fostering scientific habits of mind in the context of online play. In Proceedings of the 7th international conference on Learning sciences (pp. 723-729). International Society of the Learning Sciences.
Sun, P. C., Tsai, R. J., Finger, G., Chen, Y. Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. Computers & Education, 50(4), 1183-1202.
Tzouveli, P., Mylonas, P., & Kollias, S. (2008). An intelligent e-learning system based on learner profiling and learning resources adaptation. Computers & Education, 51(1), 224-238.
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.