Tetris™: Exploring Human Performance via Cross Entropy Reinforcement Learning Models

Catherine Sibert, Wayne D. Gray, and John K. Lindstedt
Cognitive Science Department
Rensselaer Polytechnic Institute

Abstract

What can a machine learning simulation tell us about human performance in a complex, real-time task such as Tetris™? Although Tetris is often used as a research tool (Mayer, 2014), the strategies and methods used by Tetris players have seldom been the explicit focus of study. In Study 1, we use cross-entropy reinforcement learning (CERL) (Szita & Lorincz, 2006; Thiery & Scherrer, 2009) to explore (a) the utility of high-level strategies (goals or *objective functions*) for maximizing performance and (b) a variety of features and feature-weights (methods) for optimizing a low-level, one-zoid optimization strategy. Two of these optimization strategies quickly rise to performance plateaus, whereas two others continued towards higher but more jagged (i.e., variable) plateaus. In Study 2, we compare the zoid (i.e., Tetris piece) placement decisions made by our best CERL models with those made by the full spectrum of novice-to-expert human Tetris players. Across 370,131 episodes collected from 67 human players, the ability of two CERL strategies to classify human zoid placements varied with player expertise from 43% for our lowest scoring novice to around 65% for our three highest scoring experts.

Keywords: Tetris, human expertise, strategies, methods, cross-entropy reinforcement learning

Introduction

Tetris™ is one of the most played games in the world (Stuart, 2010), one of the games most used for psychological studies (Lindstedt & Gray, 2015; Mayer, 2014), and a favorite challenge for the machine learning community (Fahey, 2013; Gabillon, Ghavamzadeh, & Scherrer, 2013; Szita & Lorincz, 2006). The latter became interested in Tetris as a challenging machine learning problem. The former has seen Tetris as potentially important for its presumed side effects for things as diverse as ameliorating sex differences in spatial skills (Linn & Petersen, 1985; Okagaki & Frensch, 1994; Sims, 2011; Terlecki, Newcombe, & Little, 2008), relief from "flashbacks for trauma" (Holmes, James, Coode-Bate, & Deeprose, 2009), and improving the abilities of engineering students (Martin-Gutierrez, Luis Saorin, Martin-Dorta, & Contero, 2009). The world's many game players, presumably, enjoy Tetris simply because it provides an enjoyable and entertaining challenge.

Within cognitive science, Tetris has been used to develop (Kirsh & Maglio, 1994; Maglio, Wenger, & Copeland, 2008) and refine (Destefano, Lindstedt, & Gray, 2011) the construct of Epistemic (or Complementary) Action. This use of Tetris is qualitatively different from other uses as the researchers were interested in the detailed interactions among cognition, perception, and action that forms the basis of interactive behavior in Tetris. The scientific arguments relied on a deep analysis of the instance by instance interactions of humans-with-zoid (i.e., the Tetris pieces), the tradeoffs made between *cognition in-the-head* and *cognition in-the-world*, and how those tradeoffs changed as a function of expertise with Tetris.

The current work is well within this cognitive science tradition. In Study 1, we build 8 cross-entropy reinforcement learning (CERL) controllers that attempt to optimize Tetris performance using four different strategies (*objective functions*) crossed with two different sets of features such as the landing height of the last zoid added, pits (the number of empty cells that are covered by other zoids), and many more. Similar to genetic algorithms, each CERL model optimizes performance by adjusting the weight given each feature in the feature set over several generations. In Study 2, the two best of these 8 controllers are used to classify each of 370,131 episodes collected from 67 human Tetris players.

Following Newell's (1973) injunction to "accept a single complex task and do all of it" this work is part of a larger effort that seeks to understand the acquisition of extreme expertise in Tetris (Gray, Hope, Lindstedt, & Destefano, 2014).

Playing Tetris

Players use the keyboard or special game controllers to rotate *zoids*, as they are falling, into an accumulating *pile* of zoids at the bottom of the screen. When a player fills an entire row, the row vanishes, and the score increases. Since it is not always possible to clear rows, the pile gradually rises. The game ends when the pile rises above the top row in the board. (A game in progress is shown in Figure 1.) Despite Tetris' widespread appeal, it is unwinnable. If you play it long enough, you will lose (Baccherini & Merlini, 2008; Kendall, Parkes, & Spoerer, 2008; Fahey, 2013)!

The standard Tetris board is 10 squares wide by 20 squares high. At the beginning of the game the zoids fall at the rate of 1.25 rows per s and would take 25 s to fall from the top to the bottom row. This drop speed increases with the game level, and at level 9 the pieces fall at 10 rows per s or 2 s to fall from the top to the bottom row. Mastering decision-making and physical placement at these rates is a significant challenge for human players.

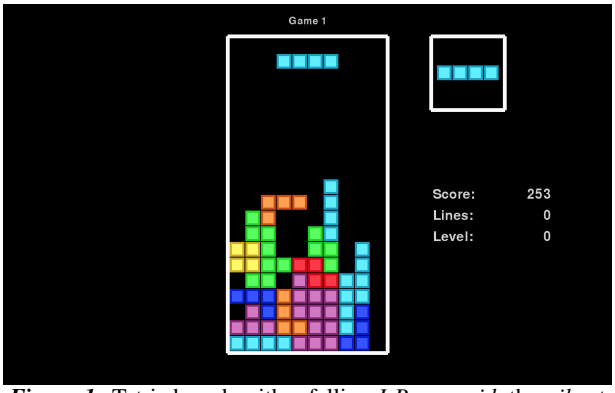

Figure 1. Tetris board, with a falling *I-Beam-zoid*, the *pile* at the bottom, and a new *I-Beam-zoid* in the Preview Box on the right.

When people play Tetris, we somehow consider both the current move and some number of future moves to determine where to place a zoid to maximize points and minimize height. Data from our best human players suggest that they have a web of contingency plans that span the current zoid, the next zoid (which is shown in the Preview box in Figure 1), and several unknown future zoids.

In contrast, our CERL models are one-zoid optimizers which make move decisions by evaluating all potential zoid placements using sets of weighted features and selecting the highest scoring move. As Table 1 shows, these features are metrics such as the total height of the pile, the number of unfilled squares, or pits, and the number of lines that will be cleared by the given placement. For any given game board configuration, the feature values will differ slightly for each possible placement. Ties among the highest rated zoid placements are decided randomly.

Study One

The first study explored the performance of our four different objective functions on two different feature sets. We consider each objective function as one goal or strategy that a human player could choose to optimize. In terms of feature sets, we adopted the *Dellacherie* set (Fahey, 2013) of 6 features that has been widely used in the machine learning literature (Szita & Lorincz, 2006; Thiery & Scherrer, 2009, 2009) (See Table 1). We also created our own set of 48 features. This set is composed of features developed in our prior work (Lindstedt & Gray, 2013; Lindstedt, 2013) combined with the 6 Dellacherie features as well as other features described in the machine learning literature. Unlike the machine learners, who were interested in claiming bragging rights as to which approach cleared the most lines, we are interested in human level results. Hence, for this purpose we ran each model on each generation until it died or until it completed 506 Tetris episodes (i.e., where each episode is the placement of one zoid), as 506 episodes is the longest game played by any player in our laboratory.

Table 1
Useful Tetris Features Proposed by Dellacherie

Feature	Description
Landing height	Height where the last zoid is added
Eroded zoid cells	# of cells of the current zoid eliminated due to line clears
Row transitions	# of full to empty or empty to full, row transitions between cells on the board
Col transitions	Same as above for vertical transitions
Pits	# of empty cells covered by at least one full cell
Wells	A series of empty cells in a column such that their left cells and right cells are both occupied

Cross Entropy Reinforcement Learning

Four things are required to train the CERLs: an objective function, a set of features, an assignment of weights to those features, and patience. Patience is required as each controller is trained for 80 generations where each generation consists of 100 controllers completing one game of Tetris each.

For the first generation, the starting controller sets all factor weights to zero and the standard deviation for each factor to 100. Hence, the first 100 models for this first generation form a cloud around the zero starting point, with a standard deviation of 100. Each successive generation begins with a new *starting controller* defined by the mean values of the best performing 10 models from the prior generation. To avoid early convergence, a constant noise factor of 4 was introduced in each generation, meaning that for the first generation, the standard deviation was 104. This noise factor remains constant throughout the generations, while the standard deviation to which it is added is adjusted and potentially converges on an optimal feature value. As for the first generation, the new starting controller is used to spawn one hundred new controllers that form a cloud around this new starting point. This procedure is followed until 80 generations of controllers have played Tetris, resulting in a highly optimized controller.

Of course, within each run of 80 games, the definition of "best controller" depends on the objective function being optimized. For our studies these objective functions were (a) *Score*, (b) total number of *Lines* cleared, (c) highest *Level* reached, and (d) the number of episodes in which four lines

Table 2
Feature Weights of the Lines and Score Objective Functions for the Dellacherie Features Set

Feature	Lines	Score
Landing height	−0.45	−0.29
Eroded zoid cells	+0.32	−0.81
Row transitions	−0.26	−0.34
Col transitions	−0.64	−1
Pits	−1	−0.61
Wells	−0.13	−0.05

(*4Lines*) were cleared at once. The first three objective functions are typically displayed to human players during the game (as shown by the right-middle portion of Figure 1). However, in all of our human games, humans were told to optimize the first; namely, Score. The fourth objective function, clearing four lines with one zoid, rewards 13.3 times as many points as does using four zoids to each clear one line. Note that clearing four lines at once is called a "Tetris" and gives the game its name. Human experts often report that the setting up and execution of these "Tetris" moves composes a large part of their game strategy.

At the end of each generation, before the next set of 100 controllers was generated, the new starting controller played 30 test games, consisting of 3 games each of 10 preselected game seeds (the game seeds produce different randomizations of the sequence of zoids). The average score of these 30 games was used to track the learning of the model over each generation. The mean scores for these 30 games is plotted, for each of the 8 objective functions, in Figure 2.

Results

Learning Feature Weights. Table 2 shows the final weights (normalized) of the six Dellacherie features for the Lines and Score models. Key differences between the strategies employed by each model can be observed within these numbers. For example, "eroded zoid cells" is clearly favored by the Lines model's moderately positive weight of +0.32, but less emphasized by the Score model's strongly negative weight of −0.81. These different weightings are characterized by behavioral differences between the models in that the Lines model seeks to clear as many lines as possible (thus eroding the zoid cells), whereas clearing lines is not as emphasized in the Score model, allowing it instead to build up to higher score payoffs.

Controller Performance. As Figure 2 shows, the biggest effect on skilled performance came from the choice of objective function. Models optimized for Lines and Level quickly reached a score threshold of a little under 100,000, and then performed consistently at that threshold. Models optimized for Score and 4Lines took longer to reach a score threshold, and that threshold, while higher than those optimized for Lines and Level, was much more variable.

To put the CERL results into a human context, we can compare their performance with the mean of each human's four highest scoring games. The second to fourth best humans averaged scores of 93,000, 73,000, and 53,000, respectively. Our very best human's average score was 174,000.

Our *Human High Score*, shown in Figure 2, was contributed by one very determined human. As, for each move, the model was allowed as much decision time as it needed and as the time for the model to move a piece into position was essentially instantaneous, our representative of humanity was allowed to play with *gravity turned off*; that is, unless the human held down the drop key, the zoids did not drop. This enabled our champion to score 246,186 points (see the red line towards the top of each panel in Figure 2).

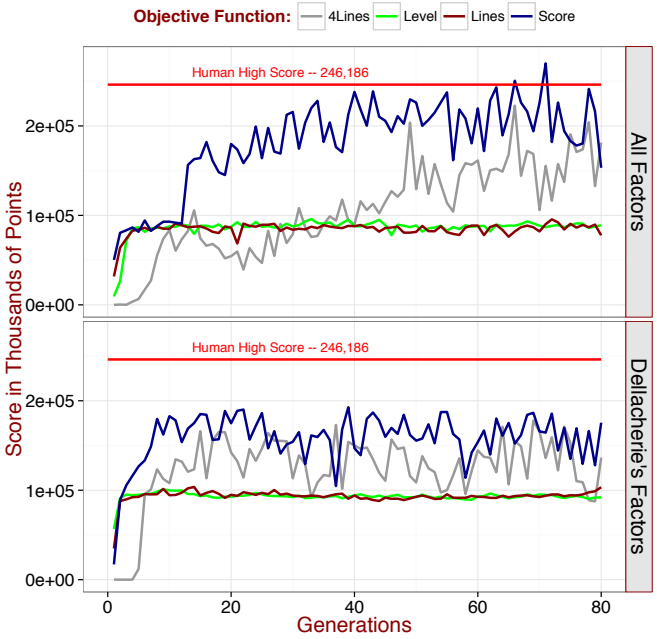

Figure 2. Learning curve of models. (See text for discussion of the Human High Score.)

Discussion

Study One tells us that machine learning models that focus on 1-zoid or 1-step optimization do pretty well compared to humans who presumably are attempting to optimize placements of two successive zoids (i.e., the current zoid and the zoid shown in the Preview Box) while planning for longer sequences, such as deliberately arranging the pile so as to clear off four rows with one I-Beam zoid. Likewise, the better human players also work deliberately over many successive zoids to prepare the playing board for high scoring opportunities while preventing disasterous buildups (of course, the CERLs do some of these things as well, see Table 1, Feature 1, Landing Height).

Of course, our humans are working under more constraints

than our CERLs. Unlike humans, once a decision is made, the CERLs instantianeously rotate, move, and place the current zoid into the desired location. Hence, it may be easy for humans to match CERL performance when the game is at level one and it takes a zoid 25 s to drop from top to bottom, but not nearly as easy when the game is at level 9 and a zoid takes only 2 s to fall the same distance!

These observations raise the question as to how much of human performance could be accounted for by 1-step optimizations and whether or when such optimizations need to be superseded by other human strategies.

Study Two

In Study Two, we used each of the 8 trained controllers from Study One (two sets of factors by four objective functions), to classify nearly 370,131 episodes of Tetris (all episodes from each of our 67 human players) as to whether the location where the human placed the zoid, matched the location that the controller would have placed the same zoid on the same board configuration.

Methods

Human Gameplay. All human Tetris games were collected in session one of a four session, 6-hr Tetris study. Session one was "free play" as the scores obtained in session one were used to assign players to Tetris conditions for the remaining three sessions of the study. All humans used the *Meta-T* (Lindstedt & Gray, 2015) experimental task environment to play Tetris and which also collected all keystrokes, eye data, and system events with millisecond accuracy. Hence, these games can be considered as normal play, uninfluenced by experimental manipulations, albeit under laboratory conditions.

Matching Humans to Models. For each episode in the human dataset, the board configuration and current zoid were given to each of the 8 final models from Study One. Each model evaluated a move score for all available moves, and returned the highest scoring move. The model's chosen move was compared to the move made by the human, and was considered to have matched the human if the model chose the same move as the human.

A move was also counted as a match under two additional conditions. First, it is often the case that the model will equally rank two or more moves. In this case, the model breaks its tie by random choice. Therefore, in cases where the human's move was equally ranked with the model's, we considered that the model's move, matched human choice. (These sitations occurred 1.39% of the time.) Second, humans were capable of making one move that the models could not; namely, humans could slide a zoid under an overhang left by another zoid (see Figure 3). Our current search algorithm considers these overhangs as inaccessible pits, but experienced human players recognize these moves quickly and tend to use them whenever available. Given the desirability of closing such pits whenever possible, we ranked the human use of a slide as equivalent to the best move considered by the model. (Overhang moves were made by humans 0.74% of the time.)

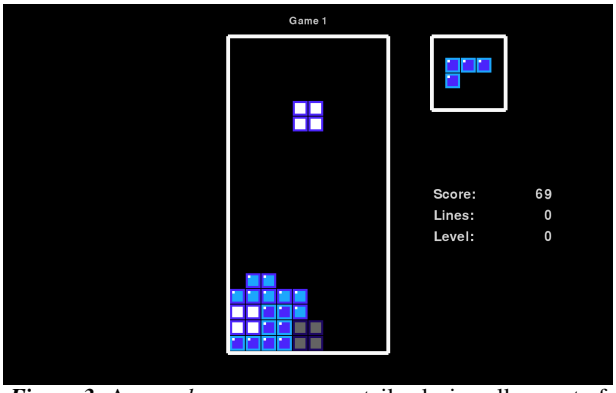

Figure 3. An *overhang* maneuver entails placing all or part of a zoid underneath parts of the pile. In the example, the player wishes to place the falling zoid where the shadow zoid is.

Results

Our 8 models were derived by crossing the four objective functions (Lines, Levels, Score, and 4Lines) with our two sets of features (ALL 48 and DELL 6). Applying these models to classifying human performance produces 8 statistical models of human performance. In this section, we seek to identify the best statistical model where *best* is defined both in terms of the model's success at classifying human moves and by parsimony; that is, the ALL 48 and DELL 8 feature sets vary in size and a trivial prediction would be that the larger feature set fits the data better than the smaller set. Hence, as we move from the realm of purely machine learning considerations, to models that help explain human performance, we want to be sure that the models we settle on have the fewest assumptions that are reasonable.

In winnowing out our 8 models we rely on the results of Multiple Regression modeling and the *Akaike Information Criterion* (AIC). Crawley (2013), describes AIC as "penalized log-likelihood" as it weighs the fit of a model against the number of parameters used; the more parameters, the more the model is penalized, and a better fit is required if the model is to be seriously considered.

Eliminating Two Objective Functions: Lines and Levels. We begin by collapsing over feature sets to compute four regression models of the form, $lm(percent\_match \sim feature\_set)$, and computing one AIC for each model. This yields AICs for Lines and Levels of -401 and -398 and AICs for Score and 4Lines of -446 and -452. As for AICs, *smaller is better*, we conclude that there are sufficient differences among our Objective Functions to justify eliminating models with Lines and Levels from further consideration.

Eliminating the All Features Models. We now compare the remaining 48 feature, ALL models with the 6 feature, DELL ones. Following Crawley (p. 416), we find that R's AIC function, e.g., $AIC(Feature.ALL.OF.Score)$ produces the same result as the loglinear model $-2 * logLik(data.ALL.score) + 2 * (3)$. Hence, we modify the loglinear model by adding the number of features into the final df term with the following results:

$-2 * logLik(model.DELL.Score) + 2 * (3 + 6)$ for an AIC of -237
$-2 * logLik(model.DELL.4Lines) + 2 * (3 + 6)$ for an AIC of -236
$-2 * logLik(model.ALL.Score) + 2 * (3 + 48)$ for an AIC of -152
$-2 * logLik(model.ALL.4Lines) + 2 * (3 + 48)$ for an AIC of -165

As lower is better for an AIC score, these comparisons led us to conclude that the simpler Dellacherie models provide the better fit and to drop the remaining All Features models from further consideration.

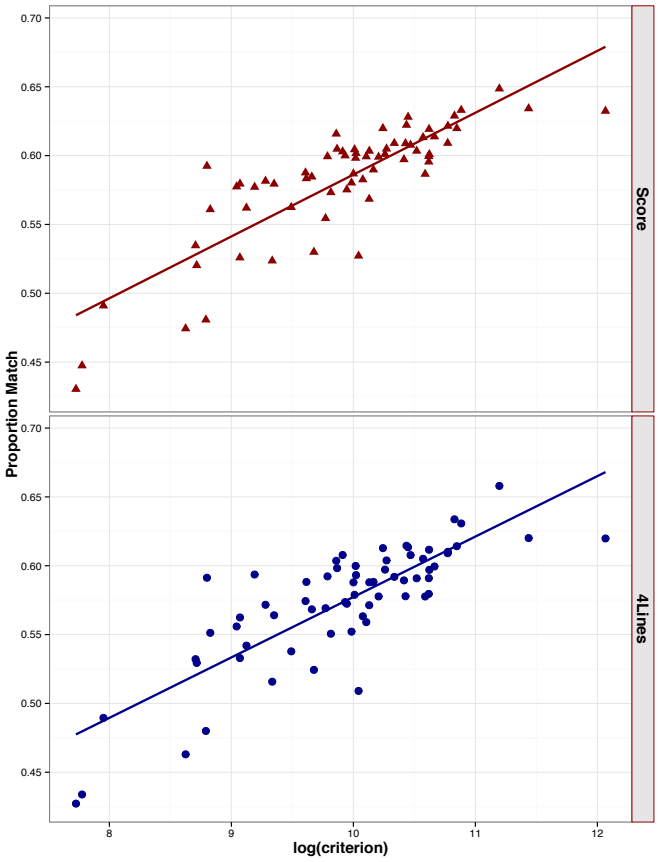

Figure 4. Proportion of moves classified for each of our 67 human player for the best feature set (Dellacherie) and the two best matching Objective Functions – Highest Score and Number of 4Lines. Each data point in the figure represents the model's ability to predict the move that the human would make. The x-axis shows human performance in terms of the log of each human's highest game score.

Fit of CERL Models to Human Data. So far we have been more concerned with reducing our set of models than we have been with what they suggest to us about the human data. Both of the remaining two models, the Dellacherie versions of the Score and 4Lines models, provide significant fits to the data (both $p's < .00001$) with the Multiple R-Sq for Score accounting for 0.33 of the variance and that for 4Lines accounting for 0.32. However, a clearer picture can be gained by looking at Figure 4 which plots the proportion matched for each of the 67 Tetris Players based on the log of their criterion score (i.e., the mean of their highest 4 game scores).

Discussion

When we began the modeling work, we entertained the hypothesis that the models would do better at predicting novice than expert performance, as we believed that novices tended to think only one zoid ahead whereas experts constantly planned for Tetrises and other manuevers. To our surprise, as Figure 4 shows, models match human performance from about 45 to 65 percent of the time with the by-player match increasing linearly from the poorest to best players.

Of course, a reasonable question to ask is always, "what is chance?" Although we are not quite sure how to calculate chance for a zoid placement, with 10 columns in which a zoid can be dropped and 1 orientation for the Square, 2 each for the I-Beam, S, and Z, and 4 each for the T, L and J, we believe that a conservative estimate of chance is around 5%. Hence, we are somewhat surprised at how good of a job these machine learning models are doing at classifying human behavior.

Finally, we are intrigued by the two right-most points in both halves of Figure 4. These are our two very best human players and they seem to be showing a prediction plateau. Do these points represent the limits of prediction for models based on one-zoid optimization? Or would our data show a continued upward slope if our dataset included more and stronger human Tetris players?

Conclusion

We see four directions forward for this line of Tetris research. First is the question as to how the objective functions of Score and 4Lines vary. It is clear for humans that clearing four lines increases score tremendously (13.3 times more points for clearing four lines with one zoid than than clearing one line with each of four zoids). Are these two really separable or does their similiarity in matching human choice imply they are measuring the same thing? Second, we want to revisit our All Factors set to determine if some of the potential power from the additional factors was wasted as subsets of different factors were calculating the same outcome in some situations but different outcomes in other situations. This might result in two subsets of predictors that were competing with each other rather than cooperating. Third, we are very intrigued with the increasing success of CERL classifications with increasing human expertise. Tutoring complex perceptual-motor-cognitive skills in real-time is very difficult and seldom done well. We wonder whether our CERL models could be effectively used to provide immediate feedback to novice or intermediate human Tetris players. Fourth, it is intriguing to ponder whether the two left-most points in Figure 4 represent a flattening of CERL powers of classification at the higher levels of human expertise or whether the slope would continue upward if more and stronger human players were found. Perhaps the apparent flattening reflects the limits of one-zoid optimization for Tetris and the beginning of a role for human strategies requiring extreme expertise?

Acknowledgements

Address all correspondence to Wayne Gray. <grayw@rpi.edu>. The work was supported, in part, by grant N000141310252 to Wayne D. Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer. Thanks to Özgur Simsek for introducing the authors to CERLs.

References

Baccherini, D. & Merlini, D. (2008). Combinatorial analysis of Tetris-like games. *Discrete Mathematics*, *308*(18), 4165–4176. doi:{10.1016/j.disc.2007.08.009}

Crawley, M. J. (2013). *The R Book* (Second). John Wiley & Sons, LTD. doi:10.1002/9781118448908

Destefano, M., Lindstedt, J. K., & Gray, W. D. (2011). Use of complementary actions decreases with expertise. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2709–2014). Austin, TX: Cognitive Science Society.

Fahey, C. P. (2013). *Tetris AI*. Retrieved from http://www.colinfahey.com/tetris/

Gabillon, V., Ghavamzadeh, M., & Scherrer, B. (2013). Approximate dynamic programming finally performs well in the game of Tetris. In *Advances in neural information processing systems* (Vol. 26, pp. 1754–1762). Retrieved from http://media.nips.cc/nipsbooks/nipspapers/paper_files/nips26/881.pdf

Gray, W. D., Hope, R. M., Lindstedt, J. K., & Destefano, M. (2014). Elements of extreme expertise: searching for differences in microstrategies deployed by experts and novices. In *Plenary Presentation at the 12th Biannual Meeting of the German Cognitive Science Society*. Universität Tübingen. Tübingen, Germany. doi:10.13140/2.1.3809.1529

Holmes, E. A., James, E. L., Coode-Bate, T., & Deeprose, C. (2009). Can playing the computer game "Tetris" reduce the build-up of flashbacks for trauma? A proposal from cognitive science. *PLoS ONE*, *4*(1), e4153. doi:10.1371/journal.pone.0004153

Kendall, G., Parkes, A., & Spoerer, K. (2008). A survey of NP-complete puzzles. *ICGA JOURNAL*, *31*(1), 13–34.

Kirsh, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*(4), 513–549. doi:10.1016/0364-0213(94)90007-8

Lindstedt, J. K. (2013). *Identifying Expertise: Data Exploration in Tetris* (Master's thesis, Rensselaer Polytechnic Institute).

Lindstedt, J. K. & Gray, W. D. (2013). Extreme expertise: Exploring expert behavior in Tetris. (pp. 912–917).

Lindstedt, J. K. & Gray, W. D. (2015). MetaT: Tetris as an Experimental Paradigm for Cognitive Skills Research. *Behavior Research Methods*. doi:10.3758/s13428-014-0547-y

Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Development*, *56*(6), 1479–1498. doi:10.1111/j.1467-8624.1985.tb00213.x

Maglio, P., Wenger, M. J., & Copeland, A. M. (2008). Evidence for the role of self-priming in epistemic action: expertise and the effective use of memory. *Acta Psychologica*, *127*(1), 72–88.

Martin-Gutierrez, J., Luis Saorin, J., Martin-Dorta, N., & Contero, M. (2009). Do Video Games Improve Spatial Abilities of Engineering Students? *International Journal of Engineering Education*, *25*(6, SI), 1194–1204.

Mayer, R. E. (2014). *Computer games for learning: an evidence-based approach*. Cambridge, MA: MIT Press.

Newell, A. (1973). You can't play 20 questions with nature and win: projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York: Academic Press.

Okagaki, L. & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: gender effects in late adolescence. *Journal of Applied Developmental Psychology*, *15*(1), 33–58. Retrieved from http://search.ebscohost.com.libproxy.rpi.edu/login.aspx?direct=true&db=psyh&AN=1994-44678-001&site=ehost-live&scope=site

Sims, C. R. (2011). *Internal models of embodied dynamics: a computational theory of learning in routine interactive behavior* (Doctoral dissertation, Rensselaer Polytechnic Institute, Troy, NY).

Stuart, K. (2010). Tetris and Snake - the biggest games in the world. *The Guardian*. Retrieved October 20, 2012, from http://gu.com/p/2e2kk/em

Szita, I. & Lorincz, A. (2006). Learning Tetris using the noisy cross-entropy method. *Neural Computation*, *18*(12), 2936–2941.

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: gender differences in growth patterns. *Applied Cognitive Psychology*, *22*(7), 996–1013.

Thiery, C. & Scherrer, B. (2009). Improvements on learning Tetris with cross-entropy. *ICGA Journal*, *32*(1), 23–33.

Thiery, C. & Scherrer, B. (2009). Building controllers for tetris. *ICGA Journal*, *32*(1), 3–11.