TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 9 (2017) 374–394

ISSN:1756-8757 print / 1756-8765 online
DOI: 10.1111/tops.12225

This article is part of the topic "Game-XP: Action Games as Experimental Paradigms for Cognitive Science," Wayne D. Gray (Topic Editor). For a full listing of topic papers, see http://onlinelibrary.wiley.com/doi/10.1111/tops.2017.9.issue-2/issuetoc.

Interrogating Feature Learning Models to Discover Insights Into the Development of Human Expertise in a Real-Time, Dynamic Decision-Making Task

Catherine Sibert, Wayne D. Gray, John K. Lindstedt

Cognitive Science Department, Rensselaer Polytechnic Institute

Received 8 October 2015; received in revised form 17 April 2016; accepted 24 June 2016

Abstract

Tetris provides a difficult, dynamic task environment within which some people are novices and others, after years of work and practice, become extreme experts. Here we study two core skills; namely, (a) choosing the goal or *objective function* that will maximize performance and (b) a feature-based analysis of the current game board to determine where to place the currently falling zoid (i.e., Tetris piece) so as to maximize the goal. In Study 1, we build cross-entropy reinforcement learning (CERL) models (Szita & Lorincz, 2006) to determine whether different goals result in different feature weights. Two of these optimization strategies quickly rise to performance plateaus, whereas two others continue toward higher but more jagged (i.e., variable) heights. In Study 2, we compare the zoid placement decisions made by our best CERL models with those made by 67 human players. Across 370,131 human game episodes, two CERL models picked the same zoid placements as our lowest scoring human for 43% of the placements and as our three best scoring experts for 65% of the placements. Our findings suggest that people focus on maximizing points, not number of lines cleared or number of levels reached. They also show that goal choice influences the choice of zoid placements for CERLs and suggest that the same is true of humans. Tetris has a repetitive task structure that makes Tetris more tractable and more like a traditional experimental psychology paradigm than many more complex games or tasks. Hence, although complex, Tetris is not overwhelmingly complex and presents a right-sized challenge to cognitive theories, especially those of integrated cognitive systems.

Correspondence should be sent to Catherine Sibert, Cognitive Science Department, Rensselaer Polytechnic Institute, Carnegie 108, 110 8th Street, Troy, NY 12180. E-mail: siberc@rpi.edu

Keywords: Tetris; Cognitive skill; Experts; Expertise; Strategies; Methods; Cross-entropy reinforcement learning; Perceptual learning; Machine learning

1. Introduction

Tetris^{TM} is one of the most played games in the world (Stuart, 2010), one of the games most used for behavioral science studies (Lindstedt & Gray, 2015; Mayer, 2014), and a favorite task for the machine learning community (Fahey, 2015; Gabillon, Ghavamzadeh, & Scherrer, 2013; Szita & Lorincz, 2006). The latter became interested in Tetris as a challenging machine learning problem. Behavioral scientists have viewed *Tetris as treatment condition*; that is, as an *Independent Variable* in their experimental designs. The world's many game players enjoy Tetris simply because it provides an entertaining challenge.

In our research, we use tools from the machine learning community to further behavioral science and, more specifically, cognitive science goals. However, we see major differences between our work and almost all prior behavioral science work on Tetris. The vast majority of behavioral science studies have viewed Tetris as a potential "treatment" for things as diverse as ameliorating sex differences in spatial skills (Linn & Petersen, 1985; Okagaki & Frensch, 1994; Sims, 1996; Terlecki, Newcombe, & Little, 2008), relief from "flashbacks for trauma" (Holmes, James, Coode-Bate, & Deeprose, 2009), or improving the abilities of engineering students (Martin-Gutierrez, Luis Saorin, Martin-Dorta, & Contero, 2009).

In contrast, our focus is on the acquisition of extreme expertise in real-time, dynamic decision-making tasks—in situations in which "even hesitating requires a decision."[1] For us, Tetris provides a complex, but not too complex, environment within which some people are novices and others, after years of work and practice, become extreme experts. Understanding how the former become the latter presents a challenge that cannot be solved by relying on traditional experimental psychology paradigms in which more than three variables are never varied at once, 1 hour of practice is the norm, and 6 hours of practice is considered expert performance. The only prior work with Tetris we have found, which also analyses in-game behaviors to draw inferences for cognitive theory, is the study of complementary action (aka epistemic action) (Destefano, Lindstedt, & Gray, 2011; Kirsh & Maglio, 1994; Maglio, Wenger, & Copeland, 2008).

The work presented here is part of a larger project to study the acquisition of extreme expertise in dynamic task environments. We refer the reader to our other study of Tetris in this issue of topiCS (Lindstedt & Gray, 2016), our framework for considering *dips and leaps* in skill acquisition (Gray & Lindstedt, 2016), and empirical work using statistical techniques from changepoint analysis (Killick & Eckley, 2014) to identify dips and leaps (Destefano & Gray, 2016) in the video game Space Fortress.

1.1. The Tetris challenge to cognitive theory

In addition to gifting us with the variety of reinforcement learning models used in our study, our machine learning friends imply that Tetris is hard, at least in part, because it contains a huge number of possible board configurations: $2^{200} \simeq 10^{59}$ (Thiery & Scherrer, 2009a)! Although we do not doubt their math nor the problem that Tetris poses for machine learning, we do doubt whether the human brain *sees* things the same way.

For example, the game tree complexity of chess is Shannon's number, 10^{120}, which includes all possible positions (legal or not). Much greater than for Tetris! However, estimates of the number of chunks that human chess masters have at their command are far lower than that. The cannonical estimate, based on computer programs that simulate aspects of chess masters' behavior, was less than 100,000 chunks but more than 10,000 (Simon & Gilmartin, 1973) with a more recent update that puts that number closer to 300,000 chunks (Gobet & Simon, 2000). Whatever the correct estimate, both numbers suggest that the positions considered by human chess masters are extremely far below the number of possible arrangements of chess pieces on a chess board. By extending this rough logic to Tetris, we would be very surprised if human Tetris players had to deal with more complexity than human chess players. Whatever its merits for estimating the difficulty that Tetris poses to machine learning models, the estimate of 10^{59} possible board configurations would seem to vastly overestimate the human challenge.

But what is the human challenge and can it be explained by cognitive science? Posing the challenge of Tetris in this way raised a question we did not know how to answer; namely, what do Tetris players see when they look at a Tetris board (e.g., Fig. 1)? Which features are important and which are not? The falling light-blue "I-beam" could be placed to rest on the board at any one of 17 positions, defined by its two rotations (i.e., vertical or horizontal) and horizontal displacement. From Fig. 1, it looks as if the current player is planning to drop the I-beam zoid (i.e., Tetris pieces are called "zoids") straight down and get points for clearing the bottom two rows. In this case, where is our player planning to put the orange "L" shown in the Preview Box (upper-right in the figure)? We "see" two reasonable placements for that zoid, but we do not know whether this player has either of them "in mind." However, if she does see what we believe are her two best placements, the one she chooses might be informed by the next zoid; that is, the zoid that will appear in the Preview Box, once the light-blue I-beam is dropped and the orange "L" is in play.

As we have described her choices, the player is clearly not a novice, but is she an expert? Hard to say. If she were an expert, she would be thinking ahead as to how to best arrange the board so as to clear four lines at once (called a *Tetris*) and thereby get 7.5 times the points she would get if she cleared one line, four times. As part of her planning she would be doing "contingency planning"—saving places on the Tetris board to temporarily store a "difficult" piece so it was out of the way until she cleared her four lines with one I-beam.

Sometimes a zoid comes along, as our expert is trying to set the board up for a Tetris, that just has to be placed in an awkward location ("awkward" in terms of our expert's plans). Sometimes even an expert must leave a hole or "pit" in the board. A Tetris board that is full of pits is bad because only filled rows can be cleared and rows with pits can quickly lead to a pile that overflows the board and causes the game to end. However, our player is not just an expert; she is an extreme expert. With a pit to deal with, she goes into "disaster recovery" mode and begins the task of eliminating pits, reducing the average height of the board, and resuming her quest to score Tetrises.

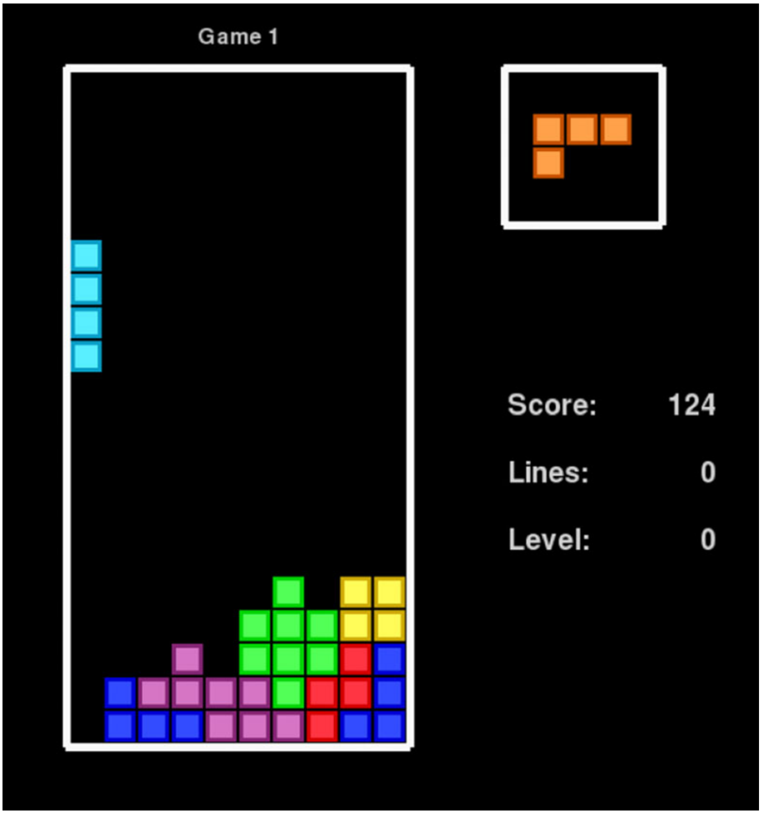

Fig. 1. Tetris board, with a falling *I-Beam-zoid*, the *pile* at the bottom, and a new *L-Beam-zoid* in the Preview Box on the right.

In summary, Tetris is a complex, dynamic task requiring problem-solving, contingency planning, and disaster recovery. In many ways it seems a perfect-sized challenge to modern cognitive theory and behavioral laboratory techniques (Lindstedt & Gray, 2015). Tetris is divided into a series of discrete episodes where the current zoid falls and the next zoid appears in a Preview Box. This repetitive task structure makes Tetris more tractable and more like a traditional experimental psychology paradigm than many more complex games or tasks. However, unlike many experimental psychology tasks, Tetris has contingencies that carry over from one episode (trial) to another that necessitate immediate, short-term, and longer-term planning. Hence, although from a distance Tetris seems like a simple immediate response type task (aka a *twitch game*); the closer we get, the more complex are the challenges to the novice player to simply "see" good placement locations and "move" the zoid to those locations. As some expertise is acquired, the challenge becomes one of longer range planning so as to score more points by clearing multiple lines at once. As expertise continues to be acquired schemes are hatched for contingency planning, but when those schemes fail, all is not always lost, as often an expert's tried and true disaster recovery plan can be invoked so that she can recover from all but the most dire situations.

1.2. Preview of the studies

For Study 1, we built several cross-entropy reinforcement learning (CERL) controllers, each of which attempted to optimize one of four goals or *objective functions*. Similar to genetic algorithms, each CERL model optimizes performance by adjusting the weight given to each of six board features over many generations. In Study 2, the two best of these controllers were used to classify each of 370,131 zoid placements (aka "episodes") collected from 67 human Tetris players.

2. Playing Tetris

Tetris is played by using a keyboard or a special game controller to rotate the pieces (called *zoids*, see Fig. 2), 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 Fig. 1.) Despite Tetris's widespread appeal, it is unwinnable. If you play it long enough, you will lose (Baccherini & Merlini, 2008; Fahey, 2015; Kendall, Parkes, & Spoerer, 2008)!

Players earn points by clearing lines, but they earn variable amounts of points based on the number of lines they clear in a single move. At the first level of the game, clearing a single line is worth 40 points, clearing two lines is worth 100 points, clearing three lines is worth 300 points, and clearing four lines, a maneuver known by players as a Tetris, is worth 1,200 points. These base scores are multiplied by a constant modifier defined by the player's current level. The level of the game increases each time a player clears 10 lines.

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 second and 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 second, taking only 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.

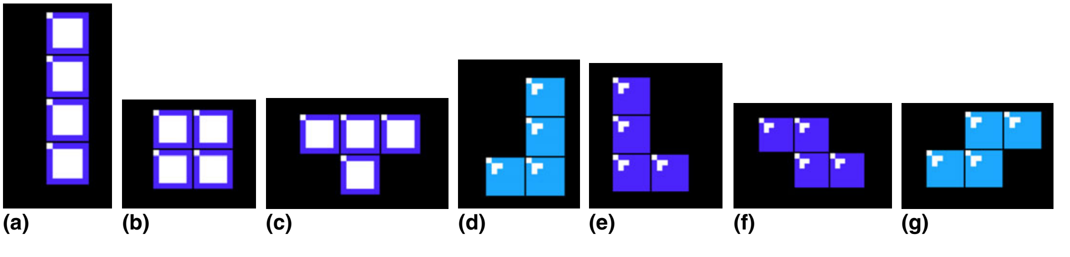

Fig. 2. The seven (7) Tetris zoids, commonly called the: I-Beam, Square, T, J, L, Z, and S.

Table 1
Useful Tetris features proposed by Dellacherie

Feature	Description
Landing height	Height where the last zoid is added
Eroded zoid cells	No. of cells the current zoid eliminated due to line clears
Row transitions	No. of full to empty, or empty to full, row transitions between cells on the board
Column transitions	Same as above for vertical transitions
Pits	No. 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

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. Interviews with 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 Fig. 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 landing height of the last zoid added, number of cells the current zoid eliminated (i.e., *eroded*) due to line clears, and the number of empty cells covered by at least one full cell (pits). When deciding where to place a piece, the model evaluates all possible placements. For each of these possible placements, the feature values will differ slightly, and these differing values are multiplied by the corresponding weight. The move score for each placement is calculated by taking the sum of all the feature values multiplied by their weight, and the model makes the move with the highest total score. Occasionally, two or more placements will tie for the highest scoring move. In these relatively rare cases, the model selects one of the highest scoring placements at random.

3. Study 1

The first study explored the performance of our four different objective functions. We view each objective function as a different goal that a human player could choose to optimize.[2] In terms of feature sets, we adopted the *Dellacherie* set (Fahey, 2015) of six features that has been widely used in the machine learning literature (Szita & Lorincz, 2006; Thiery & Scherrer, 2009a b) (see Table 1).

3.1. 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, 100 models are generated. When generating each model, its feature weights are initially randomly selected from a normal distribution with a mean of 0 and a standard deviation of 100.[3] Hence, the feature weights in these first 100 models form a normally distributed “cloud” around the origin of 0. After all of these 100 models have completed one game of Tetris (often losing the game quite quickly!), the distribution of the weights of each of the features of the top 10 performing models are then used to generate the 100 models for the next generation. In this way, with each passing generation, the cloud of models gradually narrows in on an optimal configuration of feature weights for a particular objective function. The objective function defines which models are considered the best performing, for example, the models that clear the most lines, or the models that achieve the most points. This procedure is followed until 80 generations of controllers have played Tetris, resulting in a highly optimized controller.

To avoid early convergence on a suboptimal local maximum, a constant noise value of 4 is added to the standard deviation for each feature, meaning that for the first generation, the standard deviation was 104. This noise value of 4 remains constant throughout the generations; however, the standard deviation to which it is added narrows with each passing generation, allowing some amount of exploration even when a strong solution has been found.

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.

3.2. Adapting CERLs for our purposes

The models as designed by the machine learners produce very good “players,” but play under different conditions and for different goals (i.e., objective functions) than human players.

3.2.1. Conditions of play

The most dramatic difference between CERL players and human players is the elimination of time pressure for CERL players. Although Tetris has been shown to be unwinnable—if you play it long enough, you will encounter a sequence of zoids that cannot be handled (Baccherini & Merlini, 2008; Fahey, 2015; Kendall et al., 2008)—it is unclear how important this feature is in human wins and losses. Rather, more prominent for human players (especially for those who are not expert) is “loss due to time pressure.” As the game level increases, it reaches speeds where humans are physically unable to plan and move fast enough. CERLs completely avoid this aspect of the game. Implementing a comprehensive and realistic sense of time pressure for the models proved difficult, however, and for our explorations, we went for a simple approximation of time pressure by limiting the game length; that is, no CERLs were allowed to play games of more than 506 zoids—that being the length of the longest game played by any player in our laboratory.

3.2.2. Goals/objective functions

There is also a big difference between why CERL players and human players play. In the machine learning literature, as far as we can determine, all CERL players have been rewarded for the number of lines they clear, whereas most human players work to gain a high point score. These two objective functions are correlated but, as any human expert will tell you, simply clearing single lines is not the way to get a high score. Humans make a rational tradeoff. We cannot handle the highest levels due to the speed and every time we clear 10 lines the level and speed increases. Therefore, humans tend to maximize the point value of clearing those 10 lines by clearing multiple lines at once.

3.2.3 Our models

For our exploration, we trained the models using different objective functions: reaching the highest level, clearing the most lines, and scoring the most points. In addition, we added a fourth objective function that rewarded models for completing the most simultaneous four line clears, the highest scoring move in the game.

Not all differences between CERL and human players are addressed in this initial exploration. As with previous CERL researchers, we determined that allowing the CERLs to consider an upcoming piece (i.e., the "preview piece" typically shown in the Preview window—see the right side of Fig. 1) would increase computation time too much, and so the models we produced are single-zoid optimizers, rather than the two-zoid optimizers that we suspect human players to be. Similarly, the ability for the models to detect and execute overhang maneuvers was too computationally expensive to include in this study. We hope to include both of these capabilities in future models.

Another issue that we do not address is the feature set. In the machine learning literature, many different feature sets have been used, with the most common being the six feature, Dellacherie set (Fahey, 2015). For our explorations of objective functions, we chose to go with the machine learning norm; namely, the Dellacherie set. In future work, we hope to use a variety of methods, including some of the work proposed by Lindstedt and Gray (2016), in the hopes of finding features which can be shown to be similar with those humans use.

3.3. Discovering the best *controller*

Each objective function model is run for 80 generations whereupon for each generation 100 models are spawned. Our models used the Dellacherie features shown in Table 1.

For our studies, the 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 (*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 Fig. 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 30 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 report that setting up and executing these “Tetris” moves composes a large part of their game strategy.

For each generation, we ran each model 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. 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.

3.4. Results

For Study 1 we discuss three types of results: (a) the learning of feature weights, (b) comparing performance of each model across objective functions, and (c) comparing each model to human performance.

3.4.1. Learning feature weights

Table 2 shows the final weights (normalized) of the six Dellacherie features for each of the four models tested. Key differences between the strategies employed by each model can be observed within these numbers. For example, maximizing the number of cells cleared with each zoid, the “eroded zoid cells” feature, is favored by the Level and Lines model (with z-scores of +1.450 and +1.507) and less favored by the Score and 4Lines models (z-scores −0.823 and −0.971) as line count and level reached does not benefit from the score boost that comes from clearing multiple lines. In contrast, building Wells (empty columns surrounded by filled columns) is much more favored by Score and 4Lines (z-scores +1.315 and +1.289) than by Level and Lines (z-scores +0.466 and +0.512).

Table 2
Feature weights, in z-scores, of the Level, Lines, Score, and 4Lines objective functions for the Dellacherie features set

Feature	Level	Lines	Score	4Lines
Landing height	−0.096	−0.197	+0.637	+0.351
Eroded zoid cells	+1.450	+1.507	−0.823	−0.971
Row transitions	+0.187	+0.216	+0.505	+0.881
Col transitions	−0.409	−0.629	−1.364	−0.391
Pits	−1.577	−1.409	−0.271	−1.159
Wells	+0.446	+0.512	+1.315	+1.289

Note. For ease of comparison, the feature weights for each model have been normalized for each objective function by first centering and then normalizing. To center, we subtract the column mean from each value in its column. To normalize, we then divided each centered number by its column’s standard deviation.

3.4.2. *Comparing model performance*

Each of the four models was trained to maximize each of the four objective functions where these objective functions can be considered analogous to human goals (as discussed in Janssen & Gray, 2012). Table 3 shows the value that each model achieved on each of our four objective functions.

Level model—was rewarded for increases in game levels (row 1 in Table 3). Across its 30 post-training games, it reached an average of 19.7 levels. This score ties the highest level achieved by the Lines model but is better, by this metric, than the Score or 4Lines models.

Lines model—cleared more lines than the Score or 4Lines models but tied with the Level Model.

3.4.2.1. Levels and lines and longevity: Even though most Tetris games, including ours (Lindstedt & Gray, 2015), provide the players with their current Level, Lines, and Score (see Fig. 1, the right side of the gameboard), Level and Lines actually tally the same behavior; that is, the level increases as a function of the number of lines cleared. However, although we were initially surprised by the similarity of the Lines and Levels CERLs on these two measures, we should not have been as the math here is quite simple.

As discussed earlier, 506 episodes is the maximum that any model can play. Each zoid contains 4 squares. Clearing one line requires filling it with 10 squares. Hence, it requires 2.5 zoids to clear one line. As we limited our CERL controllers to 506 zoids total, then the maximum number of lines that any CERL could clear would have been 506 *zoids*/2.5 zoids-per-line; that is, 202.4 lines. As these calculations show, our Levels and Lines CERLs are at asymptote in lines cleared. Indeed, this result is probably what makes "lines cleared" such an attractive metric for the Machine Learning community; namely, it provides a simple and unambiguous metric of model longevity. The model that plays the longest is the one that clears the most lines.

Score model—achieved the highest score, 175,455. This is almost twice the score of the Level model and 1.7 times the score of the Lines model and it achieves this score while clearing only 168 lines (see Fig. 1).

Table 3
Optimizing the objective functions for the Dellacherie feature set: One model was trained on each of four different Objective Functions; levels reached, lines cleared, score achieved, or number of times four lines were cleared at once

Model Type	Level	Lines	Score	4Lines
Level model	19.70	200	91,943	0.10
Lines model	19.70	200	103,342	0.57
Score model	16.33	168	175,455	8.56
4Lines model	11.86	124	136,413	9.66

Note. For Model Type, the red number indicates its value on the Objective Function on which it was trained.

4Lines model—obtained more Tetrises (9.66) than other models. In contrast, it reached the lowest level and cleared the fewest lines. Although it achieved the second highest score, this score was closer to the score achieved by the Lines model than to the Score model.

3.4.2.2. Summary: Janssen and Gray (2012) observed that the choice of when, what, and how much to reward are important considerations in comparing reinforcement learning models to human performance. In our studies we have held "when" and "how much" constant and focused on "what." Despite our expectations, we find ourselves surprised by how much the four scores varied depending on which objective function the CERL pursued.

3.4.3. Optimizing performance

As discussed in Section 3.1, at the end of each of 80 generations of CERL controllers, the mean feature weights of the best 10 models were averaged to create a new starting model for the next generation. However, before the 100 models in the next generation were spawned, the starting model played 30 games of Tetris. The mean *scores* (i.e., not levels, lines, or number of 4lines cleared) for each set of these 30 games is plotted, for each objective function, in Fig. 3.

As emphasized in Section 3.2, our interest in CERLs is in what they can tell us about human performance. Hence, as human Tetris tournaments most typically award prizes to those with the highest score and as this is what we tell players in our lab to optimize, we are most interested in comparing score as a human goal versus as a CERL objective function.

Congruent with Table 3, Fig. 3 shows the biggest effect on skilled performance came from the choice of objective function and the effects occurred early in training. Models

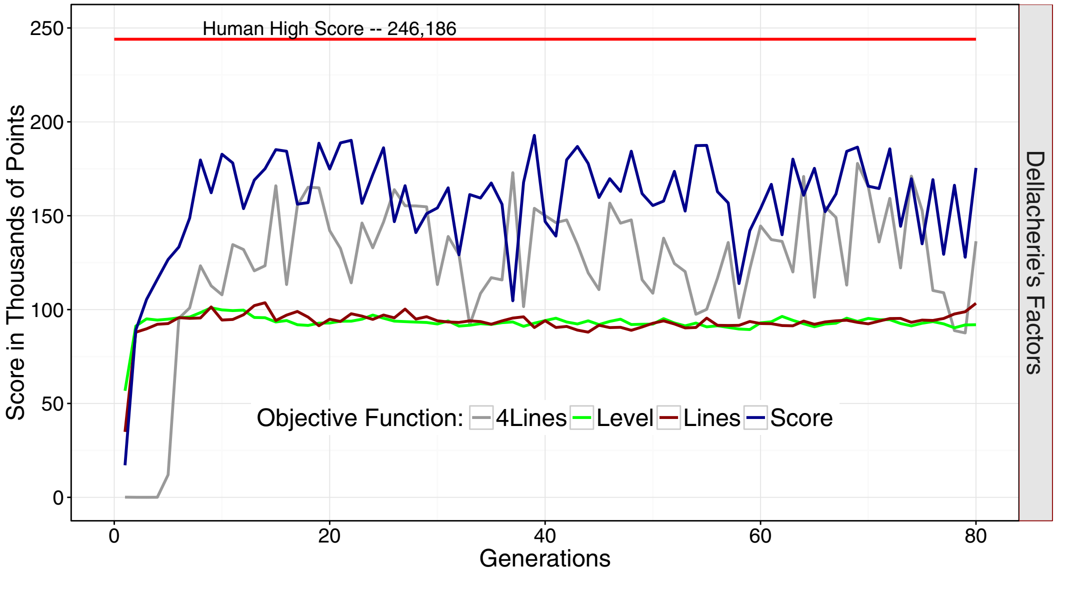

Fig. 3. Learning curve of models. (See text for discussion of the Human High Score.)

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. The *Human High Score*, shown in Fig. 3, 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 toward the top of each panel in Fig. 3).

3.5. Discussion

Study 1 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 at once. Likewise, the better human players also work deliberately over many successive zoids to prepare the playing board for high scoring opportunities, while preventing disastrous buildups (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 instantaneously 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 1 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 optimization and whether or when such optimizations need to be superseded by other human strategies.

4. Study 2

In Study 2, we used each of the trained controllers from Study 1 (one for each of four objective functions), to classify 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 controller's highest rank placement for the same zoid in the same board configuration.

Study 2 explored three alternative hypotheses; namely, that the model's classification of human moves would be (a) random with respect to human expertise, (b) inversely proportional to human expertise, being better able to predict novice than expert zoid placements, or (c) vary proportionally to human expertise.

All three hypotheses assumed that both model and human performance would show a rational adaptation to the task environment (Anderson, 1990 1991; Gray, Sims, Fu, & Schoelles, 2006) to the limits of their respective cognitive, perceptual, and motor capabilities.

Hypothesis 1 seemed plausible as we knew what the model was doing but we had no idea how reliant human performance was on long-term memory, planning, and higher level strategies, or whether the reinforcement learning processes for the model zoomed in on the same longer term contingencies as the humans. In this case, the correlation between human and model move selection would be random with respect to human expertise.

Hypothesis 2 seemed plausible if the placement decisions of human novices emphasized the same features and the same sets of longer term plans and goals as the models but if human experts developed completely different sets of longer term plans and goals. Hence, in this case, the correlations between human and model move selection would decrease with expertise.

Hypothesis 3 seemed reasonable if as human expertise increased, decisions as to where to place the current zoid were increasingly influenced by an awareness (of sorts) of how current placements contribute to the success of longer termed plans.

As humans were told to optimize score, both the second and third hypotheses suggest that at least some of human performance should mirror the choices made by the models that emphasized score and clearing four lines, more than it would the lines cleared or levels achieved models.

As Fig. 5 shows, Hypothesis 3 was supported with the model matches to human performance increasing linearly from 43% to 65% from the poorest to best players.

4.1. Methods

4.1.1. Human gameplay

All human Tetris games were collected in Session 1 of a four-session, 6-h Tetris study. Session 1 was "free play" as the scores obtained in Session 1 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 to also collect 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.

4.1.2. Matching humans to models

For each of the 370,131 episodes in the human dataset, the board configuration and current zoid were given to each of the four final models from Study 1. Each model evaluated a move score for all available moves and returned the set of highest scoring moves. The human's move was considered to have matched the model if it was in the model's set of highest scoring moves. Note that in 1.39% of the matches the model returned more than one move in its set. We consider that the human's choice matched the model's if it was in this set of tied choices.

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 Fig. 4). 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.)

4.2. Results

4.2.1. Normalizing expertise

For purposes of statistical analysis, our measure of human expertise needs to follow a normal distribution. Unfortunately, Tetris has a peculiar reward structure that escalates as players reach higher levels of performance. For this and for possibility other, unknown reasons, our measure of Tetris expertise—the mean of each human player's highest four Tetris games—is not normally distributed.

For purposes of this analysis, we transformed the criterion score by calculating the fourth root of each player's criterion score (the mean of their highest four games). By the Shapiro–Wilk test, the transformed data were not significantly different from normal with a $W = 0.975$ and $p = .1837$.

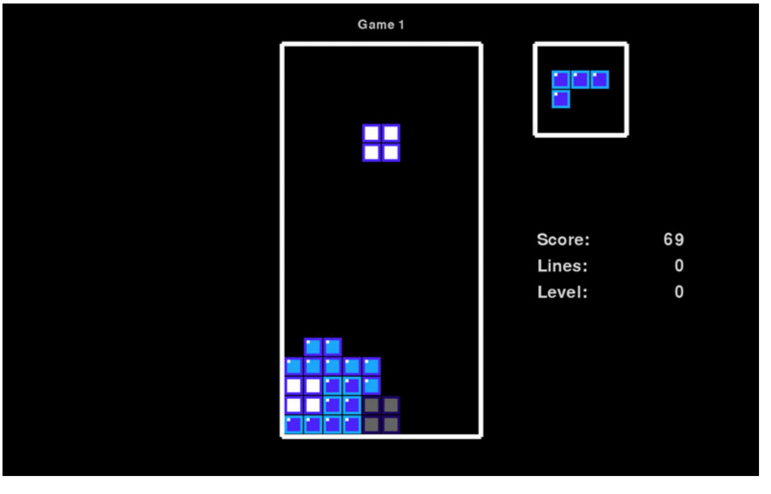

Fig. 4. An *overhang* maneuver entails placing all or part of a zoid underneath parts of the pile. In the example, the player wishes to place the falling zoid where the shadow zoid is.

4.2.2. Comparing objective functions

Our four models were trained on one of each of the four objective functions (Lines, Levels, Score, and 4Lines). Applying these models to classifying human performance produces four statistical models of human performance. In this section, we seek to identify the best statistical model where *best* is defined in terms of the model's success at classifying human moves by our regression analyses' adjusted R^2, the p-value of its F test, and likelihood (Glover & Dixon, 2004). The probability, F test, and likelihood can be considered complementary measures. As Gallistel (2015) tells us, "probabilities attach to results; likelihoods attach to hypotheses."

In calculating the likelihood ratio for each of our four models, we rely on the *Akaike Information Criterion* (AIC). Crawley (2013) describes AIC as "penalized log-likelihood" as it rewards goodness of fit. It is often used to weigh the fit of a model against the number of parameters used; the more parameters, the more the model is penalized. However, as each of our four models has the same number of parameters, the "penalties" for each are the same and the AIC gives us each model's likelihood estimate.[4]

4.2.3. Eliminating two objective functions: Lines and levels

The results of our four regression models are summarized in Table 4. As can be seen by the table, the Score and 4Lines models are similiar to each other on Adjusted R^2, F test, and AIC.[5] Indeed, these two are much better on all three measures than Lines which, in its turn, is better than Levels. Hence, we conclude that there are sufficient differences among our objective functions to justify eliminating models with Lines and Levels from further consideration.

4.2.4. 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 about the human data. Both of the remaining two models, Score and 4Lines, provide significant fits to the data (both p-values are well under .00001) with the Adjusted R^2 for Score accounting for 0.63 of the variance and that for 4Lines accounting for 0.60. However, a clearer picture can be gained by looking at Fig. 5, which plots the proportion matched for each of the 67 Tetris players based on their level of expertise.

Table 4
Summary of regression analyses by objective function

Objective Function	Adjusted R^2	$F(1, 65)$	p	AIC
Level	0.099	8.284	0.005	−198
Lines	0.372	40.06	2.60e−08	−237
Score	0.631	113.60	6.63e−16	−291
4Lines	0.601	100.20	8.55e−15	−285

Note. Regressions on the fourth root of each Human Player's Criterion Score (the mean score of the highest four games).

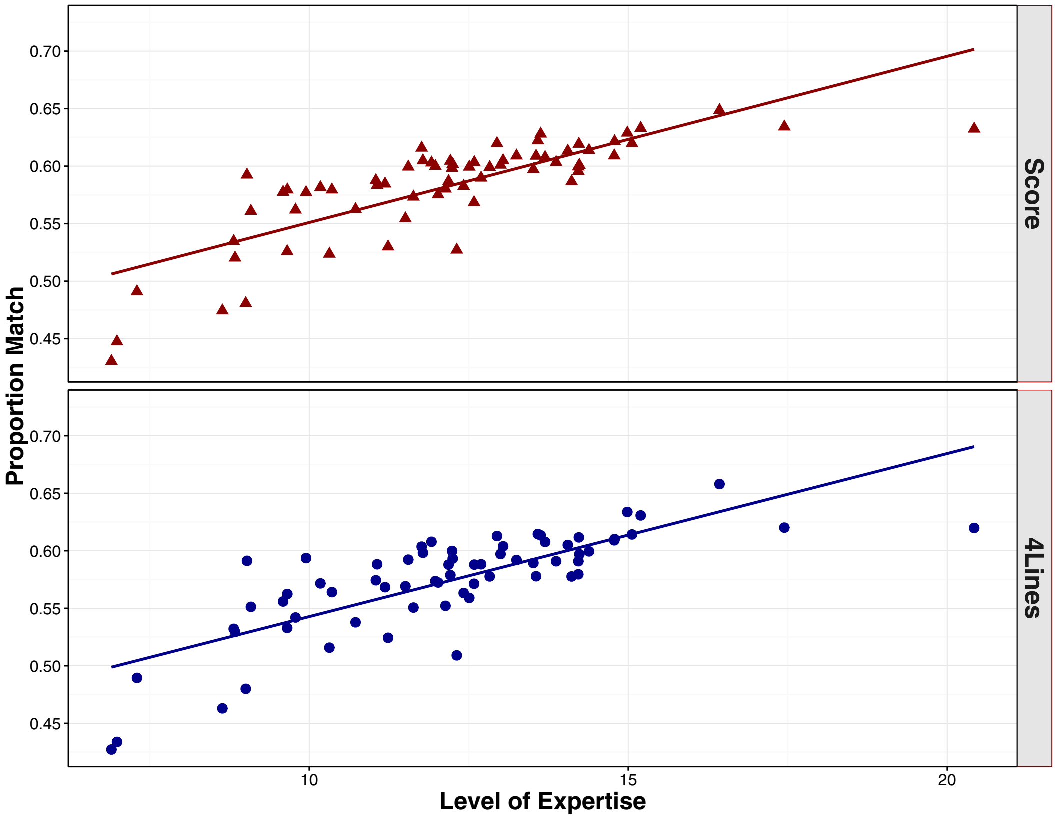

Fig. 5. Proportion of moves classified for each of our 67 human players for the Dellacherie feature set and the two best matching Objective Functions—Highest Score and Number of 4Lines. Each data point in the figure represents the model's ability to predict the moves that one human would make. The *x*-axis ranks human performance in terms of the fourth root of the mean of each human's four highest game scores.

4.3. Discussion

As foreshadowed in our introduction to Study 2, the data support Hypothesis 3; namely, that a significant component of human expertise requires recognizing the implication of configural features of the current board for long-term planning such that the decisions made for where to place the *current* zoid on the *current* board *now* are influenced by an awareness (of sorts) of how current placements influence achievement of those longer termed plans.

To a large degree, support for Hypothesis 3 highlights the power of reinforcement learning to achieve sophisticated long-term goals. Although each move is evaluated based on the board features as weighed by the current objective function, the reward is not given until the end of the game when the goodness of the feature weights assigned to each of the 100 models of a given generation is judged by how well that model did, compared to the other 99 models, on that objective function. The feature weights for the 10 best models are then averaged and used to spawn a new generation of 100 models, and so on for 80 generations. (Note that a more detailed discussion of this process was provided in Section 3.1.) Some of these interpretations seem easy to see as, for example, comparing Table 3 with Table 2 shows that the models reinforced for high scores and tetrises (the Score and 4 Line Models) weighed Wells more highly and Pits as much more detrimental than did the models reinforced for Lines or Levels.

The model begins its success by classifying 45% of the moves of novice players and ends by classifying 65% of the moves of expert players. What are we to make of the moves that the model does not capture? There are several categories of logical possibilities.

First, it seems likely the case, that no matter how slow the early levels of the game are, they are too fast for our novices. That is, there is a significant perceptual-motor component to Tetris that novices clearly lack. Simply recognizing a side-ways L-zoid as an “L” and not as a “J” (see Fig. 2, the third and the fourth zoids from the left) in all of its rotations, requires domain-specific learning (Sims, 1996). Likewise, maneuvering zoids as they are falling, even as they are slowly falling, requires the acquisition of a certain amount of game-specific, perceptual-motor skill. Hence, it seems likely that our novices are consumed with mastering the move mechanics and visual features of the game rather than game strategy. (This point is also supported by Ackerman’s [1988 1990] extensive work on the progress of skill acquisition.)

Second, what are we to make of our CERL match with our experts of 65%? Is that the tops? If we found better experts (and ours are very good but none would be competitive nationally) would we expect the match to be higher? Or have the CERL and our humans reached some sort of parting of the ways?

Our experts certainly seem to be engaged in higher level cognitive skills such as planning for whatever zoid comes next and for taking the least bad move when good moves are not available. In some cases, the least bad move results in a pit—and then the good players tell us that they work to eliminate the pit over time. A good example, if true, of disaster planning and recovery.

5. Conclusion

Comparing human data to CERL models suggests two differing directions for future work. The first focuses on perceptual and motor components of game performance that must be mastered before novice players can begin to consider more strategic aspects of performance. The second focuses on the acquisition of expert skills required for contingency planning and disaster management (Wikipedia, 2016a, b).

5.1. Schemas, perceptual learning, and motor skills

The three, classic, Tetris studies from the 1990s and 2000s (Okagaki & Frensch, 1994; Sims & Mayer, 2002; Terlecki et al., 2008) were motivated by the difficulty of recognizing Tetris zoids in all of their various rotations with the goal of deciding where in the Tetris board they should be fit. The work on Epistemic/Complementary Action (Kirsh & Maglio, 1994; Maglio et al., 2008) was motivated by the various transpositions and rotations required to identify good zoid placements and physically move the zoids to those locations. The former set of studies sought to evaluate whether playing Tetris might transfer to a general increase in spatial skills, especially mental rotation. The latter set of studies argued that rotations and transpositions served as physical replacements for mental simulations of movement. In our view, however, the one clear conclusion, which is relevant to both sets of studies, is that as expertise increases, the number of rotations and transpositions decline:

> As even by our most lenient criteria we find that the use of complementary actions all but disappears as expertise increases, we believe that our claim that expert Tetris players engage in very few complementary actions is rock solid. (Destefano et al., 2011, p. 2713)

The Destefano et al. (2011) finding suggests that at least part of the 20% increase in matching of CERL to human placements as human expertise increases might be cast in terms of some sort of domain-specific learning. One alternative might be the type of schema learning or pattern acquisition discussed by Simon and colleagues for chess players (Chase & Simon, 1973; Gobet & Simon, 2000; Simon & Gilmartin, 1973). A second alternative might be some type of domain-specific perceptual learning (Fine & Jacobs, 2002; Goldstone, Landy, & Son, 2010; Goldstone, de Leeuw, & Landy, 2015), where as the task is acquired, an individual becomes, "tuned to frequently recurring patterns in its specific local environment that are pertinent to its goals without requiring costly executive control resources to be deployed" (Goldstone et al., 2015, p. 24). However tantalizing these suggestions might be, the current study was not designed to test them. Rather, although it is clear from our work and past work that some sort of increase in pattern matching, perceptual skill, or motor performance occurs in Tetris, exactly how to categorize that skill is unclear.

5.2. *Studying the development of higher level skills with Tetris*

In adopting Tetris as a research paradigm, a motivating focus was the acquisition of extreme expertise in real-time, dynamic decision-making tasks—in situations in which, "even hesitating requires a decision" (see note 1).

We are surprised that our simple CERL models seem to capture as much of human expertise as they do. While we have emphasized the six feature Dellacherie models, a parallel effort (Lindstedt & Gray, 2016) has focused more on accounting for human performance via multiple regression and principle component analyses (PCA) of expertise. In future work, we plan to incorporate these PCAs as features for CERL controllers as well as use these to explore possible differences in features between novice and expert players.

Another planned use of CERL models will be to characterize changes in the dynamic task environment as indexed by the range between the summed feature weights of the best possible move and the worst possible move. Preliminary work shows that for some of our best players that range remains fairly steady over episodes but shows a few dramatic dips and rises. As a preliminary hypothesis we are targeting these dips as periods of disaster and disaster recovery. It will be interesting to understand how skilled players handle such episodes and interesting to determine what sorts of models we need to capture the development and execution of these highly skilled moves.

5.3. Final words: Tetris as experimental paradigm

A strength of Tetris as an experimental paradigm for cognitive science is that only one feature changes at a time. This aspect makes Tetris similar to many human tasks and games but very different than other tasks such as kayaking in rapid water or playing a first-person shooter. Hence, although complex, Tetris is not overwhelmingly complex and presents a *right-sized* challenge to cognitive theories especially those of integrated cognitive systems.

All in all, we conclude here that "Tetris as an Experimental Paradigm" is more complex than we initially realized and that a full explanation will be more difficult to achieve and far more interesting for cognitive theory than we had anticipated.

Acknowledgments

This work was supported, in part, by grant N000141310252 to Wayne Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer. Thanks to the other members of the CogWorks Laboratory (Ryan Hope and Matthew-Donald Sangster) with whom we read and discussed many of the papers cited and ideas incorporated in this work. Thanks to Özgur Simsek (Max Planck Institute for Human Development—Berlin) for introducing the authors to CERLs.

Notes

1. Lec (1962): Auch zum Zögern muß man sich entschließen—"Even the hesitation you have to decide."
2. See Janssen and Gray (2012) for a discussion of how parameter choices affect the "when," "what," and "how much" to reward in reinforcement learning models of cognition.
3. These values are somewhat arbitrary as their absolute values cannot be compared across models. Within a model they represent how that model will weigh one feature relative to another when evaluating a particular board state.
4. Note that the AIC is very similar to the Bayesian Information Criterion (Gallistel, 2015) in that both reward goodness of fit based on penalized log-likelihood.
5. For AIC, lower is better. Hence, −291 for the Score model is much better than −198 for the Levels model.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*(3), 288–318.

Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 883–901.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–517.

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

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Crawley, M. J. (2013). *The R Book* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Ltd. doi:10.1002/9781118448908

Destefano, M., & Gray, W. D. (2016). Where should researchers look for strategy discoveries during the acquisition of complex task performance? The case of space fortress. In A. Papafragou, D. Grodner, D. Mirman, and J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society (pp. 668–673)*. Austin, TX. Cognitive Science Society.

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. (2015). Tetris AI. Available at http://www.colinfahey.com/tetris/. Accessed September 9, 2015.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of Vision*, *2*(2), 190–203. doi:10.1167/2.2.5

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).

Gallistel, C. R. (2015). Bayes for beginners 1: Probability and likelihood. Available at: http://www.psychologicalscience.org/index.php/publications/observer/2015/september-15/bayes-for-beginners-probability-and-likelihood.html. Accessed October 5, 2016.

Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review*, *11*(5), 791–806. doi:10.3758/BF03196706

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*, *24*(4), 651–682. doi:10.1016/S0364-0213(00)00031-8

Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science*, *2*(2), 265–284. doi:10.1111/j.1756-8765.2009.01055.x

Goldstone, R. L., de Leeuw, J. R., & Landy, D. H. (2015). Fitting perception in and to cognition. *Cognition*, *135*(SI), 24–29. doi:10.1016/j.cognition.2014.11.027

Gray, W. D., & Lindstedt, J. K. (in press). Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance. Cognitive Science. doi: 10.1111/cogs.12412. [Epub ahead of print].

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*(3), 461–482.

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

Janssen, C. P., & Gray, W. D. (2012). When, what, and how much to reward in reinforcement learning based models of cognition. *Cognitive Science*, *36*(2), 333–358. doi:10.1111/j.1551-6709.2011.01222.x

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

Killick, R., & Eckley, I. A. (2014). Changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, *58*(3), 1–19.

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

Lec, S. J. (1962). *Unfrisierte gedanken (unkempt thoughts)* (J. G. (translator), Ed.). New York: St. Martin's Press.

Lindstedt, J. K., & Gray, W. D. (2015). MetaT: Tetris as an experimental paradigm for cognitive skills research. *Behavior Research Methods*, *47*(4), 1–23. doi:10.3758/s13428-014-0547-y

Lindstedt, J. K., & Gray, W. D. (2016). Can a snapshot predict performance? Classifying Tetris players as expert, novice, or in between after 2, 10, 100, or all zoids have been placed. *submitted for publication on 15.10.19*.

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.

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.

Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, *5*(1), 29–46. doi:10.1016/0010-0285(73)90024-8

Sims, V. K. (1996). Domain specificity of spatial skills: The case of the expert video game player. Doctoral dissertation, University of California, Santa Barbara, Santa Barbara, CA.

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology*, *16*(1), 97–115.

Stuart, K. (2010). Tetris and snake—The biggest games in the world. Available at: http://gu.com/p/2e2kk/em. Accessed April 10, 2015.

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. (2009a). Improvements on learning Tetris with cross-entropy. *ICGA Journal*, *32*(1), 23–33.

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

Wikipedia. (2016a). Contingency theory—Wikipedia, the free encyclopedia. Available at https://en.wikipedia.org/w/index.php?title=Contingency_theory&oldid=713028874. Accessed April 4, 2016.

Wikipedia. (2016b). Disaster recovery plan—Wikipedia, the free encyclopedia. Available at https://en.wikipedia.org/w/index.php?title=Disaster_recovery_plan&oldid=698880602. Accessed April 4, 2016.