Distinguishing experts from novices by the Mind’s Hand and Mind’s Eye

John K. Lindstedt\textsuperscript{a,⁎}, Wayne D. Gray\textsuperscript{b}

\textsuperscript{a}Department of Psychological Sciences, Rice University, United States
\textsuperscript{b}Cognitive Science Department, Rensselaer Polytechnic Institute, United States

ARTICLE INFO

Keywords:
Tetris
Choice reaction time
Principal Component Analysis
Expert
Extreme expertise
Perceptual learning
Perceptual expertise
Sequential decision-making
Time pressure
Video games
Skill

ABSTRACT

Tetris is a complex task notable for the increasingly substantial demands it makes on perception, decision-making, and action as the game is played. To investigate these issues, we collected data on 39 features of Tetris play for each Tetris zoid (piece), for up to 16 levels of difficulty, as each of 240 players played an hour of Tetris under laboratory conditions. Using only early (level 1) data, we conducted a Principle Component Analysis which found intriguing differences among its three, statistically significant, principle components. Each of these components captures different combinations of perception, decision-making, and action which suggests differing higher level skills, tactics, and strategies. Each component is presented and discussed, and then used in a series of principle component regression analyses on subsets of these data (a) from different Tetris levels, as well as (b) from players of different levels of expertise. We validate these models with data collected at a locally held Tetris tournament. These components represent evidence for an integrated complex of processes – the Mind’s Hand and the Mind’s Eye – that are the essence of expertise in the real-time, sequential-decision-making task of Tetris.

1. Introduction

Tetris is a complex task, notable for the perceptual-motor demands it makes on human players. These demands start low and slow but increase the longer the game is played. However, the game’s surface-level perceptual-motor demands are not as notable or important as the demands that Tetris places on the complex interactions between human cognition, perception, and action.

(A) In play – Tetris is a complex, dynamic decision-making task, in which even hesitating requires a decision to hesitate.\textsuperscript{1}

(B) Skilled performance in Tetris – requires perceptual learning, planning, motor skills, sequential decision-making, and more. Such topics are often studied one at a time, each isolated from the others. Explaining the role of all these within the context of Tetris would push us and push the field towards more integrated theories of cognition.

(C) Skill variations in the human subject pool – as Tetris is one of the most played games in the world (Stuart, 2010), we knew that simply by inviting a few hundred Rensselaer students into our lab for an hour each, we would be able to sample a wide-range of

https://doi.org/10.1016/j.cogpsych.2018.11.003
Accepted 21 November 2018
0010-0285/ © 2018 Elsevier Inc. All rights reserved.
skilled performance.

(D) Event structure – Tetris imposes event structures (Zacks, Speer, Swallow, Braver, & Reynolds, 2007) to which all players must adapt. To some degree, the degree of adaptation is a surrogate measure of expertise.

Points A and B (above) suggested to us that Tetris would be a fruitful task for advancing cognitive science by following Newell’s injunction to, “accept a single complex task and do all of it” (Newell, 1973), point C suggested that Tetris players, representing a wide-range of skill levels, would be readily available on a college campus, and point D adds to the richness of Tetris as a manageable complex domain for studying the acquisition of expertise.

Newell issued his injunction to the field in his infamous, “You can’t play 20 questions with nature and win” (Newell, 1973), sometimes referred to as “Newell’s complaint.” Thirty years after Newell, Wulf and Shea (2002) published their own complaint titled, “Principles derived from the study of simple skills do not generalize to complex skill learning.” Wulf and Shea worried, as Newell did, that the phenomena studied by them and their colleagues would not generalize beyond the laboratory to complex skilled performance.

Our Background section explains our subtitle, the “Mind’s Eye in Perception and the Mind’s Hand in Action”, some of our scope, as well as discussing prior human research and machine modeling work on Tetris. In the Event Structure of Tetris we introduce our domain of study to readers who play or who have never played Tetris. In Research Questions, we present three questions driving our work, address these questions in Section 5 – Laboratory Study, and validate them in Section 6 – Field Study: Validating the PCA Model. In Section 7 we synthesize our findings and examine what they have to say about human cognition and expertise. We end in Section 7.5, by summarizing our major findings and the conclusions we draw.

2. Background

In this section we (a) make clear the relevance of our referent to the Mind’s Hand and Mind’s Eye in our subtitle, (b) introduce past research on haste versus waste – and the view that skilled performers may be both faster and more deliberate than less skilled performers, (c) provide an overview of our work on machine players of Tetris, and (d) summarize this section.

2.1. Expertise: the Mind’s Eye in perception and the Mind’s Hand in action

Arguably, the modern study of expertise began with Chase and Simon’s “Perception in Chess” (1973a) and “The Mind’s Eye in Chess” (1973b). In both papers, we can paraphrase the authors as asking “what do experienced chess players ‘see’ when they look at a chess position?” In our work, we ask, “what do experienced Tetris players see” that leads them to place zoids in one position on the Tetris board rather than another. In common with Chase & Simon we are not focused on perception per se but more on the mind’s eye in playing Tetris.

Referring to the experiences of expert chess or expert Tetris players as the “mind’s eye” implies perceptual learning (Abernethy, Farrow, & Mann, 2018; Landy, 2018) and/or a high level of pattern recognition that is heavily knowledge dependent in that the same physical stimuli – a given configuration of chess pieces – can be interpreted by novice chess players as a simple pattern of white and black pieces, while expert players see a Sicilian Defense not a French Defense during the opening moves – or a Bishop and Knight fianchetto checkmate rather than Anastasia’s checkmate.

Of course, Chess is not an action game and a focus on Chess and on other “slow movement” paradigms such as the Tower of Hanoi (Anzai & Simon, 1979; Simon, 1975) may have contributed to Simon’s general inattention to an “action” component in his painstakingly detailed player-by-player analyses of “what is learned” in these paradigms. We refer to this neglected component as the “mind’s hand” in action which we see as compatible with Witt’s (2018, chap. 11) discussions of the interdependence of perception on action and of action on perception.

2.2. Quick but not rushed, fast but not hasty: are skilled performers faster but more deliberate than novices?

In several domains, it has been noted that motor elements of skilled performers are immune from task elements that impair the behavior of novice performers. In an interesting real-world study, Shinar, Meir, and Ben-Shoham (1998) found that manual gear shifting affected novice drivers’ detection of road signs whereas operating a car with an automatic transmission did not. Experienced drivers, in contrast, were equally able to detect road signs regardless of the cognitive or motor demands made by their car’s transmission.

It seems to be that the most damaging distractions are those that focus the performer’s attention on their own actions (Wulf & Prinz, 2001). However, this conclusion merits a strong and interesting caveat; namely, that the damaging effect of a performer’s self-focus varies with the performer’s level of expertise.

In a study of novice vs experienced golfers, Beilock and colleagues (Beilock, Bertenthal, McCoy, & Carr, 2004, p. 379) reported that experienced golfers did better hitting the golf ball while monitoring for an auditory tone than when attending to their swing. Novice golfers, in contrast, did better while attending to their swing than while monitoring for the tone. This finding was extended (Beilock, Bertenthal, Hoeger, & Carr, 2008) by varying the “tool” used from a regular “putter” to an oddly constructed one (referred to by the authors as “a funny putter”) with the funny putter bothering the experienced golfers more than the novices. A nearly identical finding was obtained (Gray, 2004) in a baseball batting task when the two conditions entailed attending to the frequency of a tone versus the direction of bat movement.
2.3. Machine players of Tetris

In the last 10 years, Tetris has been used as an experimental paradigm for testing various machine learning algorithms. The majority of the machine learning studies use Cross-Entropy Reinforcement Learning (CERL) modeling (e.g., Fahey, 2015; Gabillon, Ghavamzadeh, & Scherrer, 2013; Šimšek, Algorta, & Kothiyal, 2016; Szita & Lorincz, 2006; Thiery & Scherrer, 2009b, 2009a). In this subsection we briefly review two lines of research using machine models to illuminate the human challenge of Tetris.

2.3.1. Viewing the objective function of reinforcement learning models as making cognitive claims

The objective function favored by the Machine Learning Community is simply the number of lines cleared and, once trained with this objective function, their best models can clear hundreds of thousands of lines. These models make their decisions in milliseconds and, unlike human players, do not have to deal with time pressures due to increasing drop rates. Similarly, the models do not have to worry about the time needed to rotate and move the current zoid to the preferred location. Just as clearly, the models can play for hundreds of thousands of lines as they never become tired, never become bored, never become hungry, and never have to deal with anything that might distract a human.

In contrast to the machine learning perspective, Janssen and Gray (2012) argued, that for human modeling, the choice of “when”, “what”, and “how much” to reward in a reinforcement learning model can be viewed as making claims on human cognition. “What” is rewarded is referred to as the objective function and when we (Sibert, Gray, & Lindstedt, 2017) trained our models using different objective functions, we found that the model trained to optimize score rather than number of lines cleared, provided a better match to the human data with differences in score accounting for approximately 40% of novice human placements and nearly 65% of expert human placements. Unlike the machine learning studies where, during training, the models played until they died, in our study the models trained for a maximum of 506 zoids per game. ² Although the Score model cleared fewer lines than did the Lines model (168 vs 200), its score was notably higher (175,455 versus 103,342).

Studying the feature weights learned by the two models, we concluded that the Score model had developed the higher risk, higher payoff policy of using multiple line clears. For example, in Tetris, clearing 4 lines at once yields 7.5 times as many points as clearing one line, four times. However, this strategy requires building a higher board. In contrast, the risk-adverse, Lines model relied almost exclusively on single-line clears.

2.3.2. Tortoises and Hares

Our most recent machine modeling work on Tetris (Sibert & Gray, 2018), used a grid search to evaluate 11⁶ models of Tetris defined by 6 features with 11 possible weights per feature for a total of 1,771,561 models per run. The results presented here are from Sibert and Gray’s Study 1a.

For study 1a, the Hare models were run either until they died or played 200 lines (whichever came first). In a second run, all models played until they died (Tortoise models) with the longest run model playing 125,829 lines.

Table 1 shows that the behavior of the best Hare model differed from that of the best Tortoise model. The vast majority (73.6%) of line clears for the Tortoise model were 1-Line clears with very few (0.1%) 4-Line clears. In contrast, the Best Hare model showed a U-shaped distribution of line clears with more 4-Line clears than 3-Line clears and 26% fewer 1-Line clears than for the Best Tortoise model.

Table 2 shows data collected in our lab from human players. Importantly, the better human players (Ranks 1 and 2) show a stronger U-shaped function than our machine players, with a much higher proportion of 4-line clears than those shown by the Hare machine models (compare the Best Hare model in Table 1 with the Rank 1 and Rank 2 human players in Table 2).

Interestingly, the distribution of line clears for our poorest human players, Ranks 4 and 5, resembles our Best Tortoise models more than they do our Best Hare models. Perhaps more interesting is that Rank 1 shows more 4-Line clears than 1-Line clears and Rank 2 shows almost as many 4-Line clears as 1-Line clears. Whatever it is that our best human players are doing, it is something(s) very different than our Best Hare model.

2.4. Summary of section

We began with brief overviews of the cognitive science issues we investigate in this work; the Mind’s Eye in Perception and Mind’s Hand in Action (Section 2.1) and differences between skilled performers and novices (Section 2.2) in tasks involving the integration of perception, decision-making, and action.

We then reviewed results from our work that explored machine modeling approaches for understanding human trade-offs in Tetris (Section 2.3). Both our past (Sibert et al., 2017) and current (Sibert & Gray, 2018) machine modeling work enabled us to focus on subtle differences in decision-making to determine whether, when all other aspects of complex cognition are removed, a contribution of the focal aspect(s) remains. These sorts of surgeries cannot be performed on human players but can be performed on model players.

² The span of “506” zoids was picked as, at the time of the Sibert et al. (2017) study, that was the highest span of Tetris played by any player in our laboratory.
3. The event structure of Tetris

Tetris is a puzzle game in which players arrange falling blocks, known as zoids, into configurations known as the pile, while attempting to leave as few gaps as possible (see Fig. 1). Players can move the zoid left and right (translations), rotate the zoid clockwise or counterclockwise, and manually drop the zoid to expedite the fall to its destination. When a player fills an entire row of cells with zoid segments, the line is cleared and points are awarded. (Fig. 1 illustrates various elements of the game board and scoring rules in the game of Tetris.)

In common with many human events (e.g., pitching a tent, going for a hike, cooking a meal, or playing a board game), the event structure of Tetris was designed (Zacks et al., 2007; Zacks & Swallow, 2007). The most basic Tetris event structure is the episode (see Table 3). As Tetris is played, the zoid falls from top-to-bottom while being acted on by the player. The drops are step-like and if a given zoid were to fall all 20 rows from top-to-bottom it would pause 20 times. However, except in the case of an empty or nearly empty board, most zoids come to rest before falling 20 rows.

Within an episode, within each step of the zoid’s fall, there can be multiple instances of three types of player initiated movements. Players can move the zoid left or right (translations), rotate it clockwise or counterclockwise (rotations), or force an early drop to the next row. As these movements usually occur in combination with each other and as sequences of these various movements are often continuous across two or more steps, we refer to any and all combinations of these elements as player initiated movements (see Table 3). As there is only one movable zoid per each Tetris episode, it is convenient to consider all movements of that zoid as one event.

Also within an episode, a Player Initiated Movement event can result in filling 1, 2, 3, or 4 rows, thereby triggering a third type of event, clear rows(s), which clears 1, 2, 3, or 4 rows at once.

The fourth type of event is triggered at every 10th row clear event; namely, a level change (see Table 3). The number of levels for Tetris begins at level 0 and increments by one for each 10 rows cleared. The level change is made salient by changes in the level indicator (beneath the score and lines indicators in Fig. 1a) and by changes in the color schemes for the 7 zoids. Interestingly, many advanced players have memorized the correspondence between color scheme and level number. However, for the student players we study, the color scheme per se is not especially important.

The fifth type of event are speedups in the drop rate (see Table 3); however, for other than the 17 best players we study in this paper, speedups are synonymous with level changes. That is, for the first 10 levels of play, levels 0–9 (see Table 5) each change in level results in a color change and an increase in drop rate. But, as Table 5 shows, although all level changes result in color changes, above level 9, not all level changes result in speed increments. As very few of our 240 college players make it beyond level 9, the intricacies of this fifth level of event structure are largely ignored in this paper.

Tetris supports events other than those shown in Table 3. However, those events represent Extreme Expert maneuvers that our laboratory players either do not know about or cannot execute.

4. Research approach and questions

In this work, we attempt to relate the common and distinguishing behaviors of novice and expert Tetris players to the theoretical constructs of cognitive psychology. As ours is an exploratory effort, we ask the following questions:

---

3 Readers who are unfamiliar with the game Tetris or who have never seen an expert play Tetris may wish to watch the last round of play at the 2016 Classic Tetris World Championships – https://www.youtube.com/watch?v=DdfRQjb5o9k.

4 In large part, the decision where to segment movement is a level of analysis issue.
Can expert players be distinguished from novices when both are playing under low time pressure (level 1, fall rate 1.25 rows/s) and, if so, on what performance components are they distinguished?

Can we distinguish among our very best players when they are playing Tetris under high time pressure (level 9, fall rate 10.0 rows/s)?

Fig. 1. Elements of the game of Tetris. On the left of figure (a) is an example screen from the game with labels added: the “zoid” is the active game piece that falls from the top of the screen and can be manipulated by the player; the “pile” refers to the accumulated zoid segments at the bottom of the game space; and the “next” area is a preview of the next active zoid. The player’s current score, lines cleared this game, and current difficulty level are displayed, as well as the current game number in the session. On the right of figure (a) are examples of filling and clearing 1, 2, 3, and 4 lines. Point values for line clears of each size scale with difficulty level according to the formulas shown, where “level” is the current game difficulty level. Figure (b) shows the time pressure present at different levels of the game in the form of the amount of vertical space automatically traversed by the zoid in the span of 1 s.

(a) Tetris task elements.

(b) Time pressure in Tetris.

Fig. 1. Elements of the game of Tetris. On the left of figure (a) is an example screen from the game with labels added: the “zoid” is the active game piece that falls from the top of the screen and can be manipulated by the player; the “pile” refers to the accumulated zoid segments at the bottom of the game space; and the “next” area is a preview of the next active zoid. The player’s current score, lines cleared this game, and current difficulty level are displayed, as well as the current game number in the session. On the right of figure (a) are examples of filling and clearing 1, 2, 3, and 4 lines. Point values for line clears of each size scale with difficulty level according to the formulas shown, where “level” is the current game difficulty level. Figure (b) shows the time pressure present at different levels of the game in the form of the amount of vertical space automatically traversed by the zoid in the span of 1 s.

(1) Can expert players be distinguished from novices when both are playing under low time pressure (level 1, fall rate 1.25 rows/s) and, if so, on what performance components are they distinguished?

(2) Can we distinguish among our very best players when they are playing Tetris under high time pressure (level 9, fall rate 10.0 rows/s)?
Table 3
The event structure of Tetris.

<table>
<thead>
<tr>
<th>Event</th>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event 1</td>
<td>Episode</td>
<td>One zoid entering screen and coming to rest</td>
</tr>
<tr>
<td>Event 2</td>
<td>Player Initiated Movement</td>
<td>Rotation, Translation, or Forced drop</td>
</tr>
<tr>
<td>Event 3</td>
<td>Rows cleared</td>
<td>Filling, thereby clearing 1–4 rows</td>
</tr>
<tr>
<td>Event 4</td>
<td>Level change</td>
<td>Triggering next level by clearing 10 rows</td>
</tr>
<tr>
<td>Event 5</td>
<td>Speed ups</td>
<td>Increase in drop rate</td>
</tr>
</tbody>
</table>

(3) Can we then again distinguish among our very best players when they are playing Tetris under low pressure?

(4) Can our findings be generalized to a different population of players (contestants trying to qualify for a Tetris Tournament rather than college students playing Tetris for course credit) and to very different conditions of play; that is, the open and noisy environment of a Game and Anime Festival as opposed to the relative peace and quiet of the sound-controlled chambers used in the lab?

5. Laboratory study

In the laboratory study, we seek to construct models of player expertise based on the results of a Principal Component Analysis of empirically collected performance data, then compare the findings of these models of player skill under conditions of both high and low time pressure that naturally occur in the task.

This section contains five main parts and a discussion. Parts 1 and 2 are discussed in the Methodology subsection. Part 1 details the collection of an hour of Tetris gameplay data from each of 240 students in the laboratory. Part 2 details the set of 39 behavioral features extracted from the gameplay data collected in part 1 (i.e., the moment-to-moment structure of the pile and player actions while maneuvering the zoid).

Parts 3–5 are discussed in the Results subsection. Part 3 codes player expertise, or the criterion score, as a normalized transformation of the mean of the highest-scoring four games played in the one hour lab sessions from part 1. Part 4 is a Principal Component Analysis (PCA) that reduces the 39 features from part 2 into a set of 3 orthogonal components by weighting them in such a way as to account for as much of the variance in player behavior as possible (note that specific indicators of the player’s final game score or game level were not included in this analysis). Part 5 constructs Principal Component Regression (PCR) models of player skill using the criterion scores from part 3 and the three components from part 4. Models are constructed and compared for both low pressure (level 1, 239 players) and high pressure (level 9, 27 players) game scenarios, with a third model focused on skill differences among the 27 high-scoring players under low pressure conditions. Fig. 2 shows the relationship and flow of data between the parts of this laboratory study.

5.1. Methodology

5.1.1. Part 1: Data collection

Over the course of 5 semesters between 2013 and 2015, we collected one hour of gameplay performance data from 240 participants in our laboratory with the intent of comparing their relative levels of expertise as distinguished by the fine details of their moment-to-moment task performance.

5.1.1.1. Participants. Participants were recruited from the undergraduate populations of General Psychology and Sports Psychology courses. Participation was rewarded with course credit (or hourly compensation of $10 for a session during Summer semesters). Participation in the study was reviewed and approved by the Rensselaer IRB.

5.1.1.2. The Tetris task. The task of each subject was to play the Meta-T (Lindstedt & Gray, 2015) implementation of “Classic Tetris” (Wikipedia, 2016). Although there are minor visual differences stemming from our use of Python as an implementation language, the Meta-T version of Tetris is a faithful representation of the original NES Tetris used at the annual, “Classic Tetris World Championship” (CTWC). Meta-T has been examined by the software expert of CTWC and found to be a faithful version of Classic Tetris up through level 19. Above level 19 there are subtle hardware and software differences between Meta-T running on a modern computer and the original Tetris cartridge running on a 1980s-era NES machine (as used by the CTWC) that have proven difficult to reconcile. However, as none of the 240 participants in our lab or in the tournaments run by our lab has ever reached level 19, these feature differences were not important to this study.

5.1.1.3. Procedure. Each participant was seated in front of a desktop computer, given an NES (Nintendo Entertainment System) controller (retrofitted to connect to the computer via USB), and asked to play 50 min of our custom experimental Tetris software Meta-T (Lindstedt & Gray, 2015). After the game, the participant completed a brief exit survey and was debriefed.

The computer was equipped with an SMI eye tracking system. Eye tracking data were collected, but are not part of the current study.
During play, the Meta-T software tracks game states and keypresses to the millisecond level with such fidelity that a perfect replay of each player’s performance could be played back at actual (or arbitrary) speed. This high fidelity logging allows analysis of the details of moment-to-moment gameplay that goes far beyond simple analyses of overall game scores used in other Tetris studies (e.g., Holmes, James, Coode-Bate, & Deeeprose, 2009; Linn & Petersen, 1985; Martin-Gutierrez, Luis Saorin, Martin-Dorta, & Contero, 2009; Okagaki & Frensch, 1994; Sims & Mayer, 2002; Terlecki, Newcombe, & Little, 2008). Hence, Meta-T transforms Tetris from the “game as treatment condition” mode used in these prior studies, to a Game-XP (game as experimental paradigm) (Gray, 2017).

5.1.1.4. Data reduction and filtering. As a player progresses through a single game of Tetris, the time pressure rises with the level change (Event 4, see Table 3) until even the best players can no longer control the board, resulting in the end of the game. This final level of play necessarily contains the time at which things “went wrong” and the player lost the game, possibly exhibiting panicked behavior in the process. While this “game-losing behavior” is likely to be quite interesting to examine on its own, it is qualitatively different from whatever constitutes “successful” performance for a given player’s level of expertise. Many players choose to abort games early if all is not going exactly to plan. This behavior is more common during the early levels, when the player has little invested, than it is when the game’s difficulty level approaches the limit of the player’s ability. Although we asked players not to abort when they had the impulse to get a fresh start, it seems at least some Tetris players could not avoid doing so. As a data analysis issue, this game-aborting behavior is difficult to efficiently and reliably distinguish from both normal play behavior and struggling behavior.

Thus, to side-step the issues of end-of-game panic and self-termination, we only use data from “completed” game levels, excluding behavior from the final level reached. Before filtering the data, in total the 240 players played 2153 games (mean 9.0 per player), producing a total of 271,843 episodes (mean 1132.7 per player). We omit the one player who never managed to complete level 0. After filtering, the data set contains data from 239 players playing 1814 games (mean 7.6 per player) and producing 213,322 episodes (mean 892.6 per player). (Some of these filtering processes will become clearer when we walk through a sample player’s data in Fig. 5.)

Fig. 2. Graph of the flow of data through the experiment.
Part 2: Decomposing Tetris play into measurable features

For each episode of gameplay, we decomposed the Tetris task into an array of measurable features each of which reflects some aspect(s) of player behavior (a complete list of these features is provided in Appendix A). Many of these features are measures of the game state that a player can visually observe and reason over. These features include information about the shapes of the piles a player builds and the locations in which a player ultimately places the falling zoid. Some of these game state features were derived from the machine learning literature (Fahey, 2015; Gabillon et al., 2013; Szita & Lorincz, 2006; Thiery & Scherrer, 2009a), where the focus has been on developing metrics of the “goodness” of each zoid-placement.

We interpret many of our features as measuring (a) errors of commission and recovery (e.g., the number of unworkable ‘pits’ present in the game state), (b) risk and reward (e.g., how high the player allows a board to grow in favor of a high payoff maneuvers), and (c) more abstract elements a player may be sensitive to, such as relative “randomness” or “ugliness” (e.g., the general disarray of the board).

Unlike for human players, interactive elements of placing a zoid (e.g., motor times, action sequence efficiency) are trivial or non-existent for Machine Players (MPs). Hence, our set of features goes beyond those of the MP work and also includes measures of how a player maneuvers a zoid to its ultimate destination. These features correspond with classic psychometric elements of behavior present in most any interactive task, including counts and kinds of keypresses, initial and average response latencies, and measures of efficiency of move execution. Fig. 3 illustrates some of the features measured. (See Appendix A for extended definitions of each of our 39 features.)

Fig. 3. Some features of gameplay in Tetris: (a) features of the board state, including the maximum height of the pile (max_ht), the number of unworkable holes buried in the pile (pits – indicated by the red circles), the mean height of all columns in the pile (mean_ht), the “jaggedness” of the pile’s perimeter (jaggedness); (b) a feature describing how well a particular zoid-placement fits with the surrounding pile (matches); and (c) the three types of keypress actions in the game (translations, rotations, and drops) and three temporal features of the episode including the initial latency of the first keypress of the episode (init_lat), the mean latency of keypresses throughout an episode (avg_lat), and the time until the player finally (if at all) drops the zoid into the pile (drop_lat).

5.1.2. Part 2: Decomposing Tetris play into measurable features

For each episode of gameplay, we decomposed the Tetris task into an array of measurable features each of which reflects some aspect(s) of player behavior (a complete list of these features is provided in Appendix A). Many of these features are measures of the game state that a player can visually observe and reason over. These features include information about the shapes of the piles a player builds and the locations in which a player ultimately places the falling zoid. Some of these game state features were derived from the machine learning literature (Fahey, 2015; Gabillon et al., 2013; Szita & Lorincz, 2006; Thiery & Scherrer, 2009a), where the focus has been on developing metrics of the “goodness” of each zoid-placement.

We interpret many of our features as measuring (a) errors of commission and recovery (e.g., the number of unworkable ‘pits’ present in the game state), (b) risk and reward (e.g., how high the player allows a board to grow in favor of a high payoff maneuvers), and (c) more abstract elements a player may be sensitive to, such as relative “randomness” or “ugliness” (e.g., the general disarray of the board).

Unlike for human players, interactive elements of placing a zoid (e.g., motor times, action sequence efficiency) are trivial or non-existent for Machine Players (MPs). Hence, our set of features goes beyond those of the MP work and also includes measures of how a player maneuvers a zoid to its ultimate destination. These features correspond with classic psychometric elements of behavior present in most any interactive task, including counts and kinds of keypresses, initial and average response latencies, and measures of efficiency of move execution. Fig. 3 illustrates some of the features measured. (See Appendix A for extended definitions of each of our 39 features.)

This robust list of features suffers from one critical drawback: there is substantial overlap among the definitions of many of these features such that they are highly correlated. Fig. 4 shows a visual representation of the correlation matrix for this full set of features based on all of the available episode-level data, post filtering. For the exact correlation coefficients, see Appendix A. It is clear from the extent to which the correlations appear to cluster in the figure that some of these features are measuring similar underlying elements of the task and player behavior. In the next section we turn to principal component analysis to construct a reduced set of components that better explains variation in performance.
5.2. Results

The data collected from each participant captured performance as granular as the individual keystrokes in Player Initiate Movements (see Event 2, Table 3), and as coarse as summary scores for the entire game. The basic unit of analysis is Event 1, the episode (see Table 3); that is, the time between a zoid appearing, dropping, and stopping. The period of “dropping” includes all Event 2s (i.e., keystrokes to rotate, transpose, or drop that zoid), as well as less observable player decisions on where to place the zoid and how to move the zoid to the targeted location. Of course, not all plans are equally well-thought out or equally well-executed and our task was to identify those elements of player behavior that signify differences in player expertise.

Our approach is more exploratory than confirmatory, with our goal being to “observe and report” details related to phenomenon of expert player skill. We enter into this analysis trivially expecting to find significant correlations between player behavior and player expertise, but what we are most interested in are those patterns of behavior that differentially signal novice versus expert performance.

Before addressing the relationship between player behavior and player expertise, we must first: (1) appropriately codify player expertise, and (2) identify what distinct elements of player behavior are most meaningful in the game of Tetris. To address the former, we construct a metric of expertise based on players’ game outcomes, and perform a principal component analysis to address the latter.
5.2.1. Part 3: Codifying expertise in Tetris

In the present investigation, “experts” are considered to be players who score higher than their peers and do so routinely and consistently, not just in the occasional outstanding instance. As such, we want to use an average score to smooth out high outliers. But because Tetris is a sequential task where a series of perfectly executed “good” decisions can be undone with a single critical error, it is much easier for a skilled player to “crash and burn” than it is for a novice player to secure an abnormally high score. Thus we examine each player's “best consistent performance” by taking the mean of each player's best 4 game scores achieved in their 1 h of laboratory gameplay. This measure forgives the player's worst games, while considering their four best games as exemplary of their expertise. Fig. 5 shows that Player 3117's highest four games scored 142,443, 106,569, 85,028, and 178,400 points. The mean of these four is 128,110 points.

As players progress through the game the speed at which the zoids drop increases and the points received for 1-line clears, 2-line clears, 3-line clears, and 4-line clears (aka “a tetris”) are multiplied. Hence, the scoring system is highly positively skewed. In pursuit of a normal distribution of criterion scores measuring player skill, we found the transformation that maximized normality for these criterion scores was a sixth-root transform, with a Shapiro-Wilks test revealing it had the least deviation from normal compared to other roots and log transformations (W = 0.992, p = 0.24). This transformation is done not with any particular cognitive mechanism in mind, but for statistical convenience.

For comparison, Fig. 6, charts the trajectory of each player's mean score (best four games, untransformed) and mean score (best four games, 6th root transform) across each game difficulty level, as well as the distributions of players' mean high scores under each transformation. Most notable in the figure is that, for the untransformed data, there is a pile of overlapping lines and low (under 50,000 points) scoring games at the easier levels of game difficulty (x-axis) with a sparse collection of stronger players represented at the higher levels of play, all of which produces a highly skewed distribution of summary scores. By contrast, the 6th root transformation highlights the differences between players across levels, and produces a much more normalized distribution of summary scores.

Thus, a player's “criterion score” is equal to the 6th root of the mean of the best 4 games played in 1 h, a normally distributed representation of a player's level of expertise in comparison to their peers. With our operationally defined measure of player expertise, we now examine what combinations of measures of moment-to-moment player behavior best account for the variety of performance in Tetris using principal component analysis.

5.2.2. Part 4: Principal Component Analysis

Principal component analysis finds a reduced set of orthogonal components that maximally capture variation in the task behavior. Each component is a weighted combination of a subset of the 39 features of player performance. Using this process, we reduce our 39 features to 3 components and examine them for more intuitive descriptions of how behavior varies in this task.\(^6\)

\(^6\) We also performed an exploratory factor analysis (EFA) which we do not report in the main body of this article because its resultant factor loadings did not differ substantially from the PCA's. Additionally, whereas the PCA yielded three components that seem coherent (as represented by the names we assigned and our discussions of them in the text), the EFA created 12 factors that lacked coherence. We revisit this issue in our Conclusions & Summary section.
We ran the principal component analysis on the pre-filtered data set of 213,322 episode-level observations, including all 39 features of the game state described in Appendix A. (As discussed earlier and illustrated in Fig. 5, we did this for each player for each game played, the pre-filtering removed all data from any level not completed during a game.) Note that this data represents a strictly behavioral record of task performance, with no sense of the player's skill level, nor the game's current difficulty level. Fig. 7 is a scree

Fig. 6. Plot of the trajectory of the mean game score across game levels for the best four games played by all players, giving an impression of the shape of the data available for analysis under no transformation (top) and a 6th root transformation (bottom). Lines represent individual players, X-marks indicate final game scores, and the distributions on the right represent the mean high scores of all games played under each transformation.

We ran the principal component analysis on the pre-filtered data set of 213,322 episode-level observations, including all 39 features of the game state described in Appendix A. (As discussed earlier and illustrated in Fig. 5, we did this for each player for each game played, the pre-filtering removed all data from any level not completed during a game.) Note that this data represents a strictly behavioral record of task performance, with no sense of the player's skill level, nor the game's current difficulty level. Fig. 7 is a scree
plot showing the variance in the data explained by each component. Typically, one searches for the inflection point, or “elbow”, in a scree plot to determine how many components to include in further analysis, as the explanatory power gained tapers off with each additional component considered. Here we see the elbow at component 4, which suggests we use only the first 3 components in the analysis. Because the remaining components explain little variance (and are difficult to describe in terms meaningful to task behavior), they are excluded.

Table 4 shows that the three principal components included in the analysis explained 25.6%, 14.0%, and 9.3% of the variance in the observed data, for a cumulative total of 48.9%. Table B1 in Appendix A shows the exact feature loadings for each component.

5.2.2.1. Component 1, “disarray”. Disarray is associated with larger piles, more holes, general messiness, and pile structures that are generally unfavorable. A positive value here means the pile is larger and more difficult to work with, while a negative value relates more to lower piles.

5.2.2.2. Component 2, “4-line planning”. Planning for “Tetrises” requires an empty vertical space that is at least 4 rows deep. This type of structure is associated with large but orderly piles that leave a deep empty well on the far left or right of the game board – an ideal setup for performing the valuable but risky maneuver of clearing 4-lines at once (i.e., a “Tetris”). Higher values of component 2 imply more progress toward preparing for a 4-line payoff, while lower values are associated with shorter piles, and perhaps more short-sighted strategies.

5.2.2.3. Component 3, “decide-move-placed”. This component neatly corresponds to the activities in Event 2 (see Table 3), Player Initiated Movement. The “decide” subcategory measures choice-response-time on four features and, for a fifth feature, includes the proportion of time in the episode during which the player intentionally “dropped” a zoid. The “move” subcategory compares the minimum number of movements required to get the zoid to its resting location with the actual number used (i.e., sum of rotations, lateral transitions, or intentional drops). The “placed” subcategory measures the goodness of the zoid’s placement in terms of minimizing the height of the pile and maximizing the number of edges matched (as per Fig. 3b).

---

7 A scree plot displays the eigenvalues associated with a component or factor in descending order versus the number of the component or factor, see (Minitab, 2017).
8 Zoids will drop by themselves. However, players can hasten drop speed by pressing and holding the drop button. Doing this indicates that the player has decided exactly where she wants the zoid to land.
5.2.2.4. Principal Component Analysis discussion. The 3 components found in the PCA explain just under 50% of the variance among all observed episode-level behavioral measures of our Tetris data. The component explaining the most variance, “disarray”, is a general assessment of the disorder in the pile. A close 2nd in terms of variance accounted for is “4-line planning” which relates to building orderly piles (i.e., structures). The importance of this component is unsurprising, as orderliness is a defining feature of the Tetris game in that one must neatly fill rows to perform well. The third component, “decide-move-placed”, is our quick but not rushed, working at a deliberate speed but not hasty component (see our Background section discussion in Section 2.2) and is an outstanding example of the mind’s hand working in close coordination with the mind’s eye. Together, these are the principal components, the elements of expertise, in Tetris.

5.2.3. Part 5: Principal Component Regression

Can we separate our novices from our stars by the moves they make or only by the points they score? Are our three principal components really “components of expertise” which allow us to understand as well as predict performance?

To answer these questions we construct three separate, multiple linear regression models. First, the low time pressure (level 1) model uses all 236 players who completed level 1. With data from 236 players, this is our “model of the masses.” Hence, our first question is whether, when all players are playing an easy level, can our three components capture any differences in expertise?

Second, the level 9 model shifts the focus to only those 27 players who are skilled enough to survive at higher game speeds, highlighting what distinguishes the very best players from one another; among these 27 players can we distinguish the “best of the best” from the “benchwarmers”?

Third, our final multiple regression model provides a twist on the first two. For this one, we construct a level 1 model that includes only our best 27 players; that is, those who made it through level 9 or beyond. As we will show, we know from our level 9 model that these players differ from each other at the high-speeds demanded by level 9, but can we separate our “best of the best” from our “benchwarmers” even when both are at an easy and relaxed level of play?

5.2.3.1. Data aggregation. We summarize our players’ performance data by taking the mean of their component scores for each difficulty level for which they had data. By reference to Fig. 5, for player 3117, we took the mean of games 2, 4, 5, and 6 for level 1. With data from 236 players, this is our “model of the masses.” Hence, our first question is whether, when all players are playing an easy level, can our three components capture any differences in expertise?

Second, the level 9 model shifts the focus to only those 27 players who are skilled enough to survive at higher game speeds, highlighting what distinguishes the very best players from one another; among these 27 players can we distinguish the “best of the best” from the “benchwarmers”?

Third, our final multiple regression model provides a twist on the first two. For this one, we construct a level 1 model that includes only our best 27 players; that is, those who made it through level 9 or beyond. As we will show, we know from our level 9 model that these players differ from each other at the high-speeds demanded by level 9, but can we separate our “best of the best” from our “benchwarmers” even when both are at an easy and relaxed level of play?

5.2.3.2. Multiple linear regression results. As shown in Table 6, we calculated three multiple linear regressions to predict player expertise (i.e., criterion score), two with level 1 data and one with level 9 data. Each model was then run through a bidirectional stepwise model selection process based on minimum Akaike Information Criterion (AIC) to eliminate superfluous components in the model.10

---

This excludes our 3 poorest players who never made it to the end of level 1, see Table 5.

The AIC is a measure of model quality designed to assess which of a given set of models is the most likely given the data. In the case of stepwise model selection, factors are added or removed and the resulting model is compared. The resulting model is the most likely model from among the set of permutations of the factors initially included.
Table 6
Component coefficients of three multiple linear regression models constructed using data from difficulty level 1 and 9 using the 3 principal component scores. The model in the left column includes level 1 data (only) for all 236 players who completed level 1. The model in the right column includes level 9 data (only) for each of the 27 players who completed level 9. The model in the middle column includes the same 27 players included in the right column model, but only uses their data collected at level 1. All 3 models attempt to predict the players’ criterion score (i.e., player expertise).

<table>
<thead>
<tr>
<th>Components</th>
<th>Level 1 – all (236)</th>
<th>Level 1 – best (27)</th>
<th>Level 9 – best (27)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coeff.</td>
<td>sign.</td>
<td>coeff.</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>5.01</td>
<td>&lt; 0.0001***</td>
<td>5.59</td>
</tr>
<tr>
<td>disarray</td>
<td>−0.12</td>
<td>&lt; 0.0001***</td>
<td>−</td>
</tr>
<tr>
<td>4-line planning</td>
<td>0.22</td>
<td>&lt; 0.0001***</td>
<td>0.09</td>
</tr>
<tr>
<td>decide-move-placed</td>
<td>0.44</td>
<td>&lt; 0.0001***</td>
<td>0.51</td>
</tr>
</tbody>
</table>

![Fig. 8. Fitted values of principal component multiple regression models for the two game levels examined.](image)

For the Level 1–all model (leftmost model in Table 6), regression was significant \[F(3, 232) = 144.9, p < 0.0001\] with an adjusted \(R^2\) of 0.648. Players’ predicted expertise level (criterion score) is equal to 5.01−0.12 (disarray score) + 0.22 (4-line planning score) + 0.44 (decide-move-placed score). All three components were significant predictors of player expertise at level 1. None of the components were eliminated during model selection.

For the Level 9–best model (rightmost model in Table 6), regression was significant \[F(2, 24) = 12.5, p = 0.001\] with an adjusted \(R^2\) of 0.470. Players’ predicted expertise level (criterion score) is equal to 5.65 + 0.13 (“disarray” score) + 1.12 (“decide-move-placed” score). Only the “decide-move-placed” component was a significant predictor of player expertise at level 9, while the “disarray” component showed near-alpha significance \((p = 0.058)\). The “4-line planning” component was eliminated from this model.

To investigate whether the behavior of the “best of the best” players’ differed among themselves at low time pressure, we ran a multiple regression model using level 1 data for only the 27 best, Level 1–best (i.e., those who successfully completed level 9 at least once), shown as the middle model in Table 6. The regression was significant \[F(2, 24) = 24.47, p < 0.0001\] with an adjusted \(R^2\) of 0.644. Players’ predicted expertise level (criterion score) is equal to 5.65 + 0.13 (“disarray” score) + 1.12 (“decide-move-placed” score). Both the “4-line planning” and “decide-move-placed” components were significant predictors of player expertise at level 1 for the 27 best players. The “disarray” component was eliminated from this model. Fig. 8 illustrates the fit of each model’s predictions to its training data.

5.3. Discussion: laboratory study

5.3.1. Level 1 – All 236

For the low time pressure level 1 model with 236 players, the component scores successfully predict differences in player expertise. Even though level 1 is not challenging, the component analysis differentiates between low and high skilled players. The
component that most separates players by skill is the decide-move-placed component. This is our human-execution component whose
11 features fall into three categories: decision efficiency (choice-reaction-time and drop probability), efficient movements (mini-
mimizing the number of rotations and transpositions used to move the zoid to the targeted location), and goodness of the zoid’s
placement (i.e., placement of the zoid so as to maximize the number of edges matched to the piece and to lower the overall height of
the board).

5.3.2. Level 9 – Best 27
The high time pressure level 9 model (27 players) differentiates among the relatively small number of players with data at this
high difficulty level. Importantly, the decide-move-placed component is the only significant predictor of player skill at this level.

5.3.3. Level 1 – Best 27
We ran this model to determine whether we could differentiate among our 27 best players based on their performance at level 1.
After all, a naive perspective would be that although this group could not help being better on our metrics at level 1 than our lesser
skilled players, the skill differences among this top group would not show through at this supposedly relaxed level of play. However,
our discovery of skill differences between members of this elite group of players at level 1 play suggests two conflicting hypotheses:
(a) the best players are intentionally rushed or hasty, versus (b) the best players are quick but not rushed – that is, they are working at
a deliberate speed but not hasty. The former suggests that players are executing movements as fast as possible perhaps to warm up or
practice. The latter suggests that from the players’ perspective, they are not feeling rushed or hasty but are simply making decisions
with their acquired efficiency and precision that results in faster decisions, more efficient movements, and more effective placements
(as per our discussion in Background 2.2).

5.3.4. Components across difficulty levels
To better understand the nature of our three components, we can also examine Fig. 9, which illustrates the components’ dynamics
across the breadth of difficulty levels achieved by our players. We have broken out both the top 27 players (those who successfully
completed difficulty level 9) as well as the 27 worst players to give an impression of how these dynamics shift with player skill. These
groupings for our top and bottom players creates a large middle group of 182 players.

Component 1, “disarray” (the left-most plot of Fig. 9), the higher this score, the worse a player is doing. However, disarray
appears to be something that all players struggle with for the duration of the game. The very best players do not differ much from the
mid-range players, with the exception that the best players survive longer, managing to better cope with the naturally increasing
disarray score. The very worst players, on the other hand, appear unable to manage this dimension of the game and are eliminated
early on.

Component 2, “4-line planning”, seems to differ greatly across the three groups, with the best players displaying much higher
scores than the mid-range players, and the very worst showing still lower scores. All players, regardless of skill, show a decrease in
this score as time pressure increases. Part of what component 2 captures can be seen in Fig. 10. This figure, borrowed from Sibert and
Gray (2018), shows the number of Tetrises (i.e., 4-Line clears) per level for each of 67 players as their level of play increases. As the figure suggests, for even the best players, the number of Tetrises per level drops as the rate of falling increases from level 0 to level 16. However, as Appendix B1 shows, this component does not count the number of Tetris but, rather, is measuring changes in board configurations as the game speeds up and players become unwilling to build the high structures required for 4-Line clears; such human concerns are captured by the “4-line planning” component as something about board height, jaggedness, and other features of board construction which vary across player skill as well as across level-of-play.

Component 3, “decide-move-placed”, shows a very interesting trend – most players increase their “decide-move-placed” behaviors (i.e., those behaviors associated with speed, accuracy, and efficiency) as task demands increase, but the very best players maintain a nearly constant level of the “decide-move-placed” component across all difficulty levels, maintaining nearly the same speed and accuracy at level 1 as at level 9 and beyond. Although we list the features of our decide-move-placed component in the serial order of (a) decide, (b) move, and (c) placed, we believe this component is more complex than that ordering suggests. For example, with increasing expertise we believe that the ability to spot the best placement comes first. Hence, the end of the “decide” phase (which is marked by any keypress for a transposition, rotation, or drop) signals that the player has decided not simply to “move” the zoid but fully “where best to place the zoid”. This location decision also implies that the player has evaluated the board to determine that an open path exists to that location.

Fig. 10. Number of 4-line clears (i.e., Tetrises) per level for each of 67 players as their level of play increases. As the figure suggests, for even the best players, the number of Tetrises per level drops as the rate of falling increases from level 0 to level 16. However, as Appendix B1 shows, this component does not count the number of Tetris but, rather, is measuring changes in board configurations as the game speeds up and players become unwilling to build the high structures required for 4-Line clears; such human concerns are captured by the “4-line planning” component as something about board height, jaggedness, and other features of board construction which vary across player skill as well as across level-of-play.

Gray (2018), shows the number of Tetrises (i.e., 4-Line clears) per level for each of 67 players as their level of play increases. As the figure suggests, for even the best players, the number of Tetrises per level drops as the rate of falling increases from level 0 to level 16. However, as Appendix B1 shows, this component does not count the number of Tetris but, rather, is measuring changes in board configurations as the game speeds up and players become unwilling to build the high structures required for 4-Line clears; such human concerns are captured by the “4-line planning” component as something about board height, jaggedness, and other features of board construction which vary across player skill as well as across level-of-play.

11 The 5 levels of expertise used by Sibert and Gray, roughly correspond to the five levels used in this paper but note that the Sibert and Gray figure is based on the 67 players used in Sibert et al. (2017) which is a subset of the 239 players used here.
5.3.5. Summary

These analyses show that the components that emerged from our Principal Component Analysis of players are important to different degrees for discriminating among players at different levels of expertise. Under low time pressure (level 1 play), scores on these three components correlate significantly with expertise among all 236 players (see Table 6). Interestingly, under high time pressure (level 9 play), only our decide-move-placed component remains discriminative among the 27 players who completed level 9. Isolating these 27 players to look at their level 1 performance (see Level 1 - best (27) of Table 6) shows that this decide-move-placed component continues to discriminate among these 27 even when that time pressure is removed. We also see that, of the three components, only the “decide-move-placed” component remains invariant across skill levels for our very best players. Implications of these findings are discussed in the general discussion section below.

6. Field study: validating the PCA model

Typically to validate our statistical models of player expertise we would employ methods such as leave one out (LOO) cross-validation or splitting the data into training and test sets. While such paths of analysis are well traveled and more than adequate, we are in a unique position to perform a more unusual method of model validation which increases the external validity (Gray & Salzman, 1998, p. 217) [see also, (Cook & Campbell, 1979)] of our PCA model by generalizing our conclusions across persons, settings, and times; namely, (a) to a new dataset of players (persons), (b) who were playing in the “qualifying rounds” of a tournament and not “laboratory” conditions (places), and (c) in some cases, were playing years before or after our laboratory data collection (times).
6.1. Tournament procedure

We hosted Tetris tournaments using our Meta-T software (Lindstedt & Gray, 2015) at Rensselaer Polytechnic Institute’s annual student-run, Genericon conventions from 2014 to 2017. Genericon’s attendees consist of a mix of RPI students, local area residents, and some superfans who fly in from longer distances. Entrance into each Tetris tournament was free, such that even low-skill players could be encouraged to participate and contribute (after all, “it’s for science!”). To attract skilled players, we awarded cash prizes of $300, $200, and $150 for 1st, 2nd, and 3rd place winners respectively. (Data collection was approved by the Rensselaer IRB in a proposal written specifically for the tournaments.)

Each tournament consisted of a qualifying round and a single elimination tournament. Across the four years of tournaments, 124 hopeful players competed in our qualifying rounds with those 8 who scored the highest in one of their two qualifying round games going on to compete in that year’s tournament.

During the qualifying round, all contestants played two games, using the same two random seeds. This meant that the order in which the sequence of zoids appeared in each qualifying game was fixed such that all players would experience the same two zoid lists, though the order in which the two games were played was randomly selected.

As each year’s tournament consisted of only 8 players, we do not have enough tournament data to conduct a statistically powerful analysis of the tournaments proper. However, the 124 hopeful players who completed the two qualifying round trials did provide us with enough data to validate our laboratory results.

6.2. Qualifying round data selection

In common with the laboratory analyses reported above, we excluded the final incomplete level of each player’s game and then extracted level 1 data from each tournament entrant’s two qualifying round games. Then all three component scores were computed, and all values were averaged per player such that each player was represented by one three-dimensional behavioral data point for the purposes of estimating their expertise using our level 1 PCR model from the laboratory study. Fig. 11 illustrates the available data for a sample tournament entrant.

6.3. Qualifying round estimates of expertise

During the qualifying round, each contestant played two complete games. Hence, we cannot compute the same, “mean of the highest four games”, criterion score as with the laboratory data. Instead, we estimate player expertise by the same ruler used to determine their inclusion in the tournament: the highest qualifying score (HQS). We then apply the sixth-root transformation used in the criterion score used for the laboratory study.

6.3.1. Qualifying round results: PCR skill estimate predicting qualifying round scores

We predicted each player’s individual expertise level (criterion score) based only on their level-1 behavioral components, which we will refer to as a player’s PCR skill estimate. We then ran a simple linear regression examining whether the PCR skill estimates correlated with the players’ HQS. The resulting model had an adjusted $R^2$ of 0.525 ($F(1, 122) = 136.8, p < 0.0001$), showing a relationship between the laboratory model’s skill estimates and players’ game performance during the qualifying round. Fig. 12 illustrates this relationship. This finding compares well to our in laboratory test reported above which found an adjusted $R^2$ of 0.652.

6.4. Tournament prediction discussion

The PCR model did well in predicting qualifying round performance (adjusted $R^2$ of 0.525), only somewhat less than on the laboratory data on which it was trained (adjusted $R^2$ of 0.648). Notably, this successful prediction involves (a) a new dataset of players (persons), (b) who were playing under “tournament” not “laboratory” conditions (places), and (c) in some cases, were playing years before or after our laboratory data collection (times). These successful predictions increase the effect construct validity of our findings. The model’s success is especially exciting as these predictions are based only on data collected during a short section of early gameplay (level 1 alone), long before it is clear what the final scores for those games will be – our model can “watch” only a handful of episodes and make reasonable estimates about that player’s success in a tournament setting.

7. Summary & conclusions

Our work has focused on “taking one task” – a faithful re-creation of the 1985 NES Tetris – and “doing it all” – examining a plethora of measures of the task, both in and outside the laboratory, to try to find telling trends. While we do not believe that we have done all of it, we do believe that this work, together with our recent (Sibert et al., 2017; Sibert & Gray, 2018) and planned work is making progress in that direction. The outcome of the current work can be summarized as six findings.

---

12 We did attempt some “prognostication”, attempting to see if our model could predict the winners of tournament matches just from “watching” players’ level 1 qualifying round performance, but the low N across many years of tournaments made this an underpowered endeavor, perhaps to be revisited in the future.
First, under low time pressure, our three behavioral components capture some of the elements of expertise which distinguish player expertise across a wide spectrum of skill.

Second, when the pressure is on (at level 9), players with the highest expertise can only be distinguished by the “decide-move-placed” component (see Fig. 9).

Third, under low time pressure (level 1), when the stakes are low, our 27 “best of the best” players can be distinguished from each other by behaviors of the “decide-move-placed” and the “4-line planning” components.

Fourth, the best players show “decide-move-placed” scores on par with their best performance even at lower levels, while the same score for other players starts low and increases to keep pace with the game’s increasing difficulty.

Fifth, all players, even the very best, show decreasing “4-line planning” behavior as the game’s time pressure increases. As Appendix B shows, the 4-line planning component reflects board structures and is not a simple count of lines cleared. The best players stay high on the 4-line planning component at a higher rate, and for longer periods as task demands increase, but eventually everyone abandons this high-risk, high-reward strategy in favor of behaviors more sustainable under the increased task demands.

Finally, our predictions are validated outside the laboratory in a field study, doing almost as well at predicting skill on fewer data per subject, with less time, and under more uncontrolled conditions.

### 7.1. Interpretations

The three components reflect different scopes. We interpret the “decide-move-placed” component as our strongest measure of immediate cognition; namely, the interactions of cognition, perception, and action required by dynamic decision-making. This component captures elements of both fast and accurate decision-making, and it seems to reveal a player’s level of skill across the full spectrum of expertise represented in our data set, as well as across the full breadth of the game’s increasing levels of time pressure.

In contrast, the focus of “4-line planning” seems less on immediate behavior and more on the nature of the board structures being build. This component provides an element of higher-order cognition which guides the “on board” structure as it is being built, maintained, and repaired. However, the parts of these subtasks – that is, the steps required to build, maintain, and repair under Tetris time pressure – are the domain of the decide-move-placed component.

In contrast to the above two components, we read “disarray” more as a snapshot of the state of a building construction site; albeit, one in which as soon as the successful structures are built, parts of them vanish or collapse, and the process must start again. To a large degree, the other two components work together to control or eliminate the level of disarray that results from the random selection of zoids.

The picture painted by the nature and interactions of these three components presents Tetris as a dynamic, decision-making task, entailing situation assessment (the disarray component), planning (the 4-Line planning component), and dynamic decision-making and action (the decide-move-placed component). Together and separately each component is mediated by the player’s skill and knowledge.

### 7.2. What’s the hurry?

It is somewhat curious to see our very best players exhibiting their “fast and accurate” behaviors even under the minimal temporal demands of level 0 or 1 – after all, at low levels there is no cost to taking time to do things slowly and carefully. It seems that by setting their pace during low game levels to the breakneck speeds required by high game levels that players are taking needless risks! Are they?

In studies of StarCraft players, Huang, Yan, Cheung, Nagapan, and Zimmermann (2017) suggest that there is a perceived value in “staying warm”; that is, players will press control sequences at a rapid rate that adds little or nothing at the beginning of the game, but which become critical for the high pressure, rapid action situations later in the game. Are our best Tetris players simply staying warm? Perhaps so, but we suspect there is more to the puzzle: it is possible these players simply cannot go slow. As discussed in Section 2.2, Beilock et al. (2004, p. 379) show that “haste does not always make waste”, as the best golfers perform far worse when they slow down to focus explicitly on sub-components of the task, then when they are distracted or performing at speed. As in Wulf and Prinz (2001), Beilock and colleagues’ findings suggest that slowing down to examine these well-established routines disrupts their execution.

In the case of Tetris, at least some of this speedup must be due to the increasing automation of event structures, driven primarily by the same simple practice effects as were observed by Shiffrin and Schneider (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) over spans of 10 or so hours. By this view, rapid performance is simply a consequence of dozens if not hundreds of hours of practice; hence, players are not intentionally working fast; rather, “working fast” has become their new normal. Alternatively, it might be the case that the players who do become Tetris experts have deliberately practiced being fast at low levels so as to achieve the speed needed at high levels of play. Champion Tetris players report practicing 5 or more hours per week and many report having done so for decades. It may be that through long hours over long time periods of either simple or deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993) players become faster; hence, at early levels of play working fast is their new normal.

---


14 Based on unpublished data collected during interviews with players at the 2016, 2017, and 2018 Classic Tetris World Championships.
7.3. Balancing Newell’s challenge

We have taken up Newell’s (1973) challenge– to “take one task and do all of it”, as well as address Wulf and Shea’s (2002) complaints regarding the need to examine cognition in complex tasks, beyond the simplicity of most laboratory studies. To achieve this, however, we have sacrificed something intrinsic to most of cognitive psychology research: experimental control. Just as the laboratory differs greatly from the wilderness of the outside world, so too does a complex task such as Tetris differ from streamlined laboratory experiments designed to tease apart the inner workings of cognitive mechanisms.

For Tetris, even in expert competitions, in which two players battle each other by playing the same random seed at the same time, the decisions made by each player are very different and the board configurations they build are very different. Likewise, across games the random selection of zoids (with replacement) almost guarantees that the sequence of zoids will never be one that a given player has seen before. When these factors are considered along with the individual variability between players and the noise inherent to the performance of a complex, dynamic task, we are pleased to have captured 65% of variance in the lab (with our level 1 laboratory model) and 53% of the variance in the field study.

Our work has benefited from the fact that Tetris is a task that, although somewhat rough in its distribution of data, still inherently provides a quasi-experimental task environment. Indeed, this property of many games has led us to argue for “Game-XP”; that is, games as experimental paradigms (Gray, 2017), the view that action games may be used as vehicles to advance the study of human cognition. By this view, games provide experimental psychologists with a motivated workforce which may either be experienced already in the task or may be induced to perform the task in the laboratory for longer than one hour. In either case, there are considerable advantages to the field from studying the behavior of people who are skilled as opposed to unskilled in our laboratory tasks. We hasten to add that we see this research direction not as a replacement to carefully focused laboratory experimentation, but as complementary in the pursuit of understanding how our basic cognitive mechanisms coordinate and interact to explain the complexity of behavior beyond the laboratory.

7.4. Above novices and beyond Tetris

The present work focuses on a subset of players with a limited range of skill, and only in the singular domain of Tetris. We think this approach is promising beyond both of those limitations.

7.4.1. Looking to higher skill

We know that greater skill exists in the version of Tetris examined here. The Classic Tetris World Championship (CTWC) is a growing annual competition that attracts players from across the world to compete in the Nintendo Entertainment System (NES) version of Tetris. The competitors in this event demonstrate extremely fast and accurate gameplay, reach difficulty levels far beyond that of the players in our studies, and enjoy discussing maneuvers and nuances of the task that emerge only at the very fastest speeds. Essentially, these champions begin where our players stop – they often begin their games at difficulty level 18, while our players seldom reach level 16 (see Table 5). We anticipate future studies which will focus on our collaboration with this group and our findings about their extreme expertise in our chosen domain.

7.4.2. Beyond Tetris

We believe this work sets the stage for examining expertise in other manageably complex tasks, both in and outside the laboratory. Ultimately, our principal component regression approach requires only that a task can be broken into a meaningful event structure, that its behavioral dynamics can be measured in multiple dimensions, and that skill is reasonably quantifiable. Most complex tasks have enough constituent moving parts to lend themselves well to these requirements, and the results of the current study have convinced us that principal component regressions are a powerful lens through which to examine complex, dynamic behavior.

7.5. Conclusions

We examined how expertise manifests in Tetris. A principle component analysis of 39 task features yielded three components which differed across levels of player skill for both the rigorous setting of the laboratory and for the “wilderness” of locally held Tetris tournaments. These three components represent elements of Tetris expertise.

The disarray component represents the natural disarray of the task that all players face but which the very worst players find difficult to contend. The 4-line planning component captures the board structures required for the high-risk, high-reward “Tetris” maneuver, for which the game is named. Our analyses show that the ability to build and maintain the needed structures varies greatly with player skill, but all players eventually abandon it as the task demands increase. The decide-move-placed component captures the players’ interactive skill within the task – both by hand and by eye – the ability of the best players to rapidly process the environment,
select good moves, and execute them quickly and efficiently. Our plots (see Fig. 8, right) show that average players, increasing engage in these activities as task demands increase. Interesting, the plots also show that the very best players engage in these activities as much at level 0 as they do at level 16 – apparently, once mastered these skills are never disengaged.

Together, these components highlight the different challenges – and, to an extent, the very different games – that players of different skill levels face in the game of Tetris, and the gradual coordination of players’ inherent cognitive abilities with their growing understanding of task dynamics as their skills evolve from novice to expert.

Author contributions

This work is based on Lindstedt’s doctoral thesis. For this paper, Lindstedt and Gray wrote the manuscript, Lindstedt did the modeling and data analyses, Gray and Lindstedt discussed the results and elaborated the theoretical implications of the results. Correspondence should be sent to John K. Lindstedt, Department of Psychological Sciences, Rice University, Houston, TX 77251. Email: john.kenneth.lindstedt@gmail.com. The work was supported, in part, by grant N00014-16-1-2796 to Wayne Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer.

Appendix A. Behavioral features of Tetris

Many of these features appear at first to be tediously similar, but work in the AI community (Thiery & Scherrer, 2009a) has shown that some subtle variations between these different features can be critical to survival for artificial agents; thus, we have included as many of these features in the analysis as was feasible.

A.1. Motor-related features

These features are related to the execution of motor actions, including keypresses, actions performed, latencies, and efficiency of the zoid’s path.

**Actions:** keypresses made by the player.

- **rots** - Rotations. The number of zoid-rotations performed in an episode.
- **trans** - Translations. The number of times the zoid was moved left or right– translated– in an episode.
- **prop_u_drops** - Proportion of user drops. The proportion of top-to-bottom movement that was intentionally dropped by the player in an episode. 1 indicates the player dropped the zoid the full length of the screen, and 0 indicates they never dropped it at all.

**Path efficiency:** comparing the keypresses used to the minimum necessary to complete the episode.

- **min_rots_diff** - Minimum rotations difference. The difference between the number of rotations used and the number needed to achieve the zoid’s final position. Lower values are more efficient.
- **min_trans_diff** - Minimum translations difference. The difference between the number of translations used and the number needed to achieve the zoid’s final position. Lower values are more efficient.

**Latencies:** reaction and response times at various points in an episode.

- **initial_lat** - Initial latency. Time elapsed in milliseconds from the start of the episode until the first keypress.
- **drop_lat** - Drop latency. Time elapsed in milliseconds from the start of the episode until the player first drops the zoid.
- **avg_lat** - Average latency. The mean time between all keypresses in an episode.
- **resp_lat** - Response latency. The time from the start of the episode in milliseconds until *either* the zoid is first dropped or the zoid is locked into its final position. This is a more clear measure of actual response time than drop_lat, as it accounts for episodes in which the zoid was never dropped.

**Task-structure features.** These features are related to the structural information in the task environment available for the player to reason over and, in turn, those that result from the player’s gameplay decisions and mistakes.

**Column heights:** measures of the heights of each column of the game board.

- **mean_ht, max_ht, min_ht** - Mean, Maximum, and Minimum height. The mean, maximum, and minimum height among all 10 columns in the pile. Note that only maximum height needs to exceed 20 to result in a game over.
- **cd_1 - cd_9** - Column difference (0,1)–(8,9). 9 features representing the difference in height between each successive pair of
columns. Positive values imply a raise in height from left to right, while negative implies the opposite.

**max_diffs** - Maximum difference. Maximum difference in heights among cd_1 through cd_9.

**Pits:** the unworkable covered holes that prevent line-clears.

**pits** - Pits. The number of empty cells in the pile that are covered from above. As a player must fill an entire row to clear it, pits must have the cells above them cleared before their row can be cleared.

**pit_depth** - Pit depth. The sum of all pits weighted by the number of filled cells above them in a column. This score gives more weight to pits buried deeper in the pile.

**pit_rows** - Pit rows. The number of rows containing pits. This measure considers any number of pits in one row to be equivalent, as clearing everything above such a row would uncover all of its pits simultaneously.

**lumped_pits** - Lumped pits. A measure of pits considering all adjacent groups of pits to be identical. Thus four isolated pits in the pile would have more weight than one 2 × 2 cluster of pits.

**Wells:** the low-height columns surrounded by higher columns on either side which are relevant for fitting certain zoid shapes.

**wells** - Wells. The number of empty, uncovered cells with a filled cell on either side. The deeper a well is, the harder it is to work with. Yet deep wells are also associated with the highest scoring 4-line-clear Tetris maneuver.

**deep_wells** - Deep wells. The number of consecutive well segments of depth 3 or more. These are unique in the game in that they can only be filled by an I-zoid without creating one or more pits.

**cumul_wells** - Cumulative wells. Similar to wells, but weighing each segment of the well heavier as it goes deeper. A well of depth 1 evaluates as 1 (1); a well of depth 2 evaluates to 3 (2 + 1); a well of depth 3 evaluates to 6 (3 + 2 + 1); and so on.

**max_well** - Maximum well. The depth of the deepest well.

**Pile orderliness:** measures of the overall “randomness” or “order” of the pile.

**jaggedness** - Jaggedness. The perimeter of the top of the pile. A lower value implies a flatter pile, while a higher value implies a more craggy surface.

**col_trans, row_trans** - Column and row transitions. The number of times a cell changes from open to closed along either columns or rows. This generally measures the “randomness” of the pile; a tall pile with no pits would rate low, as would a completely empty board, whereas a checkerboard pattern or completely random pile (i.e., riddled with pits and overhangs) would rate high.

**pattern_div** - Pattern diversity. This measure compares the pattern of empty and filled cells in each column, and the same for each row. Lower scores implies similar patterns across the pile, whereas a higher score implies more variability in the patterns created.

**weighted_cells** - Weighted cells. A count of the total number of filled cells in all columns, each weighted by its own height. The same number of total cells filled arranged flatly would rate lower than a those same cells stacked entirely along one wall, as more cells would be weighted higher due to their height.

**Zoid-placement:** local measures of the current zoid’s position at the end of the episode.

**landing_height** - Landing height. The height of the bottom of the zoid’s final position.

**matches** - Matches. The number of edges of the zoid (in its final position) that border a filled cell. A low number implies precarious positioning of the zoid, whereas a higher number implies a zoid fitting more “snugly” into the surrounding pile.

**d_max_ht** - Delta maximum height. The change in the max_height score after placing the zoid and clearing any filled lines. A negative value implies line clears, while a zero-value implies the zoid was not placed at the top, and positive values imply the zoid pushed the top of the pile higher (and closer to failure).

**d_pits** - Delta pits. The change in the pits score after placing the zoid and clearing any filled lines. Negative values imply lines were cleared and pits were successfully opened up, while positive values imply the zoid placement created one or more new pits.

**Appendix B. Principal component loadings**

See Table B1.
Appendix C. Exploratory factor analysis loadings

Table C1 makes explicit the factor loadings produced by an alternative analysis to the PCA reported in the paper, an exploratory factor analysis. Though some weights are different, many of the factors found either bear a striking resemblance to those found in the PCA or are so subtle and esoteric that they defy description. For simplicity’s sake (in an already very complex analysis) we decided to err on the side of the straightforward and report the PCA results (see Table C1).

### Table B1

Feature loadings from the principal component analysis. 5 features are omitted (cd_2–cd_8) as they did not contribute to any of the selected components.

<table>
<thead>
<tr>
<th>Component</th>
<th>Label</th>
<th>% variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>disarray</td>
<td>25.6%</td>
</tr>
<tr>
<td>2</td>
<td>4-line planning</td>
<td>14.0%</td>
</tr>
<tr>
<td>3</td>
<td>decide-move-placed</td>
<td>9.6%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>rots</td>
<td>0.955</td>
<td>0.109</td>
<td>0.116</td>
</tr>
<tr>
<td>trans</td>
<td>0.106</td>
<td>0.901</td>
<td></td>
</tr>
<tr>
<td>prop_u_drops</td>
<td>−0.277</td>
<td>−0.468</td>
<td>−0.193</td>
</tr>
<tr>
<td>min_rots_diff</td>
<td></td>
<td></td>
<td>−0.257</td>
</tr>
<tr>
<td>min_trans_diff</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial_lat</td>
<td></td>
<td></td>
<td>−0.180</td>
</tr>
<tr>
<td>drop_lat</td>
<td></td>
<td></td>
<td>−0.390</td>
</tr>
<tr>
<td>avg_lat</td>
<td></td>
<td></td>
<td>−0.193</td>
</tr>
<tr>
<td>resp_lat</td>
<td></td>
<td></td>
<td>−0.452</td>
</tr>
<tr>
<td>mean ht</td>
<td>0.305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>min ht</td>
<td>0.255</td>
<td></td>
<td>−0.218</td>
</tr>
<tr>
<td>max ht</td>
<td>0.290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cd_1</td>
<td></td>
<td>0.192</td>
<td></td>
</tr>
<tr>
<td>cd_9</td>
<td></td>
<td></td>
<td>−0.201</td>
</tr>
<tr>
<td>max_diffs</td>
<td></td>
<td></td>
<td>0.260</td>
</tr>
<tr>
<td>pits</td>
<td>0.272</td>
<td></td>
<td>−0.177</td>
</tr>
<tr>
<td>pit_depth</td>
<td>0.273</td>
<td></td>
<td>−0.150</td>
</tr>
<tr>
<td>pit_rows</td>
<td>0.278</td>
<td></td>
<td>−0.173</td>
</tr>
<tr>
<td>lumped_pits</td>
<td>0.265</td>
<td></td>
<td>−0.180</td>
</tr>
<tr>
<td>wells</td>
<td>0.119</td>
<td></td>
<td>0.373</td>
</tr>
<tr>
<td>deep_wells</td>
<td>0.115</td>
<td></td>
<td>0.372</td>
</tr>
<tr>
<td>cuml_wells</td>
<td>0.107</td>
<td></td>
<td>0.365</td>
</tr>
<tr>
<td>max_well</td>
<td>0.112</td>
<td></td>
<td>0.367</td>
</tr>
<tr>
<td>jaggedness</td>
<td>0.113</td>
<td></td>
<td>0.309</td>
</tr>
<tr>
<td>col_trans</td>
<td>0.273</td>
<td></td>
<td>−0.161</td>
</tr>
<tr>
<td>row_trans</td>
<td>0.291</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pattern_div</td>
<td>0.199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weighted_cells</td>
<td>0.295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>landing_height</td>
<td>0.285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>matches</td>
<td></td>
<td></td>
<td>0.137</td>
</tr>
<tr>
<td>d_max_ht</td>
<td></td>
<td></td>
<td>−0.133</td>
</tr>
<tr>
<td>d_pits</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix C. Exploratory factor analysis loadings

Table C1 makes explicit the factor loadings produced by an alternative analysis to the PCA reported in the paper, an exploratory factor analysis. Though some weights are different, many of the factors found either bear a striking resemblance to those found in the PCA or are so subtle and esoteric that they defy description. For simplicity’s sake (in an already very complex analysis) we decided to err on the side of the straightforward and report the PCA results (see Table C1).

### Table C1

Loadings of each feature on the twelve factors produced by an exploratory factor analysis (EFA) of the same data set used in the PCA presented in the article.

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>% variance</td>
<td>23.8%</td>
<td>13.5%</td>
<td>5.5%</td>
<td>5.3%</td>
<td>4.6%</td>
<td>3.7%</td>
<td>3.1%</td>
<td>1.6%</td>
<td>1.4%</td>
<td>1.0%</td>
<td>0.8%</td>
<td>0.7%</td>
</tr>
<tr>
<td>rots</td>
<td>0.955</td>
<td>0.109</td>
<td>0.116</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trans</td>
<td>0.106</td>
<td>0.901</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prop_u_drops</td>
<td>−0.277</td>
<td>−0.242</td>
<td>−0.468</td>
<td>−0.193</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min_rots_diff</td>
<td>0.947</td>
<td>0.105</td>
<td>0.110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min_trans_diff</td>
<td>0.145</td>
<td>0.875</td>
<td>0.120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
Table C1 (continued)

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>% variance</td>
<td>23.8%</td>
<td>13.5%</td>
<td>5.5%</td>
<td>5.3%</td>
<td>4.6%</td>
<td>3.7%</td>
<td>3.1%</td>
<td>1.6%</td>
<td>1.4%</td>
<td>1.0%</td>
<td>0.8%</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Features</th>
<th>initial_lat</th>
<th>drop_lat</th>
<th>avg_lat</th>
<th>resp_lat</th>
<th>mean_ht</th>
<th>min_ht</th>
<th>max_ht</th>
<th>cd_1</th>
<th>cd_2</th>
<th>cd_7</th>
<th>cd_8</th>
<th>cd_9</th>
<th>max_diffs</th>
<th>pits</th>
<th>pit_depth</th>
<th>pit_rows</th>
<th>lumped_pits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.091</td>
<td>0.267</td>
<td>0.116</td>
<td>-0.207</td>
<td>0.139</td>
<td>0.103</td>
<td>0.102</td>
<td>0.108</td>
<td>0.136</td>
<td>0.136</td>
<td>0.138</td>
<td>-0.146</td>
<td>0.162</td>
<td>0.162</td>
<td>-0.186</td>
<td>-0.107</td>
<td>-0.267</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.949</td>
<td>0.114</td>
<td>0.121</td>
<td>-0.203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.946</td>
<td>0.112</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.574</td>
<td>0.722</td>
<td>0.108</td>
<td>0.137</td>
<td>0.121</td>
<td>-0.287</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.846</td>
<td>0.277</td>
<td>-0.262</td>
<td>0.197</td>
<td>0.172</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.803</td>
<td>0.269</td>
<td>0.117</td>
<td>0.245</td>
<td>-0.222</td>
<td>0.161</td>
<td>0.146</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.127</td>
<td></td>
<td>-0.155</td>
<td>-0.734</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.105</td>
<td>0.628</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.105</td>
<td>0.203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


