TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 13 (2021) 610–665

ISSN: 1756-8765 online
DOI: 10.1111/tops.12575

This article is part of the topic "Tasks, Tools and Techniques," Wayne D. Gray, François Osiurak and Richard Heersmink (Topic Editors).

Constructing Expertise: Surmounting Performance Plateaus by Tasks, by Tools, and by Techniques

Wayne D. Gray, Sounak Banerjee

Cognitive Science Department, Rensselaer Polytechnic Institute

Received 19 April 2021; received in revised form 26 August 2021; accepted 27 August 2021

Abstract

Acquiring expertise in a task is often thought of as an automatic process that follows inevitably with practice according to the log-log law (aka: power law) of learning. However, as Ericsson, Chase, and Faloon (1980) showed, this is not true for digit-span experts and, as we show, it is certainly not true for Tetris players at any level of expertise. Although some people may simply "twitch" faster than others, the limit to Tetris expertise is not raw keypress time but the *techniques* acquired by players that allow them to use the tools provided by the hardware and software to compensate for the game's relentlessly increasing drop speed. Unfortunately, these increases in drop speed between Tetris levels make performance plateaus very short and quickly followed by *game death*. Hence, a player's success at discovering, exploring, and practicing new techniques for the tasks of board preparation, board maintenance, optimal placement discovery, zoid rotation, lateral movement of zoids, and other tasks important to expertise in Tetris is limited. In this paper, we analyze data collected from 492 Tetris players to reveal the challenges they confronted while *constructing expertise* via the discovery of new techniques for gameplay at increasingly difficult levels of Tetris.

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

Correspondence should be sent to Wayne D. Gray, Cognitive Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180, USA. E-mail: wayne.gray.cogsci@gmail.com

1. Introduction

> The best illustrations of mental functions at their limit of efficiency are to be found among those occupations of work or play [in which] excellence … is sought with great zeal and intelligence. The championship 'records' in typewriting, shorthand, telegraphic sending, golf, billiards, and the like, show approximations to the limits of improvement … (Thorndike, 1913, p. 178)

Our *task* is Tetris™, a real-time, dynamic decision-making task in which "even the hesitation must be decided" (Lec, 1962). We examine the core set of simple *tools* provided by the game's designers as well as how players at different levels of expertise use these tools to execute different *techniques*. As we discovered, even in the age of YouTube™ videos, mastering the techniques of tool use may require days, months, or years of focused practice. Indeed, contrary to the expectations of readers who were weaned on the log-log or power law of learning for individual players, skill acquisition in Tetris is not a smooth ascent but is better described as a series of *Plateaus, Dips, and Leaps* (Gray & Lindstedt, 2017).

Techniques are easier to name than they are to describe and they are easier to describe than they are to master. If the tool is in the hands of a skilled performer, a Eddie Van Halen, Serena Williams, or, perhaps, a Jonas Neubauer, difficult techniques that took months or years to acquire may appear extremely simple. That false sense of simplicity is well captured in this quote from Tetris Master Alex Kerr:

> I found Harry Hong's first max-out video impressive, but still operated under the assumption that … it was a feat of physical ability as much as it was of Tetris prowess. It wasn't until Jonas Neubauer's uploads and the comments he wrote in response to questions that solid information on how to conquer Nintendo Tetris without hypertapping began to surface. (Alex Kerr as quoted in Smith, 2014, p. 2)

Following Alex Kerr, and contrary to the opinions of most non-players and many casual players, Tetris achievement is not limited by a person's *twitch* speed (i.e., simple reaction time). Indeed, in prior work comparing human play with that of machine models (e.g., Sibert & Gray, 2018; Sibert, Gray, & Lindstedt, 2017), we assumed that human response times for Tetris players were about the same as for other individuals. In some cases, it will clearly be the case that players speed up both their movement times and decision-making times as the drop speed increases. However, in other cases, it is equally clear that the better players are the ones who have developed and mastered not faster, but better, techniques.

1.1. An overview of tetris basics

Tetris is structured into 30 levels of play. As explained in Section 2.2, *The Events of Tetris*, except for the rate at which the Tetris *zoids* (i.e., pieces) drop, each level is mostly identical (see Fig. 1). Visually, the biggest differences across levels are in the color schemes of the zoids that, of course, have no effect on gameplay. Conceptually, the biggest difference across games of Tetris is that for each game, the sequence of pieces is determined by a different random seed. Hence, across games, the seven zoid shapes appear, with replacement, in an

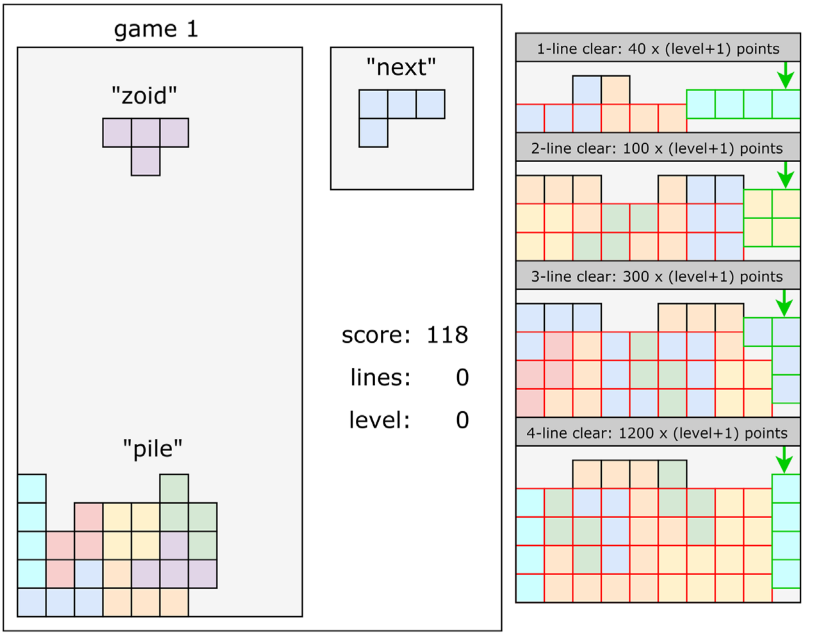


Fig. 1. Left-side: Tetris game screen showing a falling T-zoid (one of seven zoid shapes), the Preview Box (upper-right) holding the "next" zoid that will drop after the currently dropping zoid stops, and the *pile* that reflects the accumulation of zoids that have dropped but have not been "cleared." Beneath the next box, the player sees her current game score, the number of lines cleared (none yet), as well as the level number (in this example, level 0, as this is the beginning of a new game). Right side: example of Tetris boards in which either 1, 2, 3, or 4 lines can be cleared by placing one piece.

unpredictable sequence, and "droughts" are possible in which one of the seven zoid shapes may not reappear until a longer-than-expected-by-the-player series (sometimes 20 or more) of other-shaped zoids have dropped.

Mastering the tools and techniques of Tetris may require months, years, or decades; hence, the study we present in this paper will not be a longitudinal one. Rather, we *sample expertise across players* and attempt to determine the set of techniques which players who make it through say, level n possess that those who died at level $n-1$ did not. We will also limit ourselves to 492 student players. The analyses we present require much data for each comparison and at this time, we have not analyzed enough data from our CTWC players to permit us to draw firm conclusions.

1.2. Changes with increasing expertise

The number of techniques which players must master to succeed at Tetris increases with player expertise. This may seem like an odd statement to make about a simple game in which the same seven pieces continually drop, one at a time, from the top to the bottom of the playboard and in which the only controls available to players are designed to either rotate the piece clockwise or counterclockwise, move it left or right, or drop it so it falls slightly faster than it falls when left to itself. The veracity of this assertion is one of the things that will be demonstrated in this paper.

Tetris requires "learning-by-doing." Skilled performance in Tetris requires knowing where to place a zoid, how to move it to that place, and when to initiate the various rotations, transpositions, and drops that might be required to get it there. That is to say, learning-by-doing requires acquiring the right moves and applying them at the right times. A paradigmatic example for Tetris is provided by Jonas Neubauer's "spin class" in which he shows the YouTube viewer the visual cues and millisecond timing required to master the "turn-and-tuck" technique, whereas different zoids are turned and tucked as they drop so as to fit into spaces that most, if not all, of our 492 student players would deem impossible to fit.

To be clear, learning-by-watching-YouTube explains what to do and how to do it but, by itself, will not produce skilled Tetris performance. Although the techniques of the experts can be mastered at the lower speeds, these techniques are not required at these lower levels. Rather, the lower levels require the mastery of skills that are pre-requisite to playing at the higher levels. Indeed, as the game speeds up, what is needed by our student players, is a learning-by-doing approach that engages the player in a process of active exploration to acquire the predictive processes (Hommel, 1998; Hommel, Musseler, Aschersleben, & Print, 2001) required for coordinating movements with perceptions.

1.3. Whats to come

In Section 2.1, our first background section, *How Does the Perception of Action Affect Action Control?*, we briefly review the history of Ideo-Motor Action, along with recent thought on perceptual learning, *Predictive Processing*, and *Event-Predictive Cognition* (EPCog) to provide the cognitive psychology background for what occurs during the acquisition of dynamic skilled performance. Our second background section (Section 2.2) introduces and discusses the major events which players must master if they are to achieve intermediary levels of expertise in that game.

After these background sections, we move to a detailed discussion of the Methodology (see Section 3) used for collecting our data, deciding which people and games to include, random number seeds, and the various steps of data preparation.

We report our analyses in four sections. Section 4 discusses feature extraction and the six factors we found which account for most of the variance in our data. Section 5 introduces our logistic regression models used to distinguish between beginner, intermediate, and expert players at various levels of gameplay. As not every player who makes it to a given level applies the same tools and techniques in the same sequential order, in Section 6, we apply linear models to determine which factors yield differences among players at the same expertise level. Finally, as we use a limited set of random seeds in our games, in Section 7, we discuss the variation in skill requirements by different seeds and their effect on player performance. Also note that in these three sections, we use the terms of statistics, such as *factors*, rather than terms such as *event*, *tool*, and *technique*.

After discussing the details and highlights of our results in Section 8, we summarize our project (Section 9), and try to clearly state our conclusions (Section 10). Finally, for those who want more details as to what we did and how we did it, we hope that you will find the answers you seek in our six Appendix sections (Appendices A–F).

2. Background

2.1. How does the perception of action affect action control?

> Maintaining fixation rather than moving the eyes can draw upon executive control. An interesting extension of this idea applies to tennis playing. The very best tennis players direct their gaze differently than less skilled players … the best players keep their gaze where the racquet made contact with the ball after the ball was hit. Less skilled tennis players tend to follow the ball with their eyes (Lafont, 2007). Planning to maintain fixation may confer the advantage of focusing attention more fully on hitting the ball in just the right way. (From Rosenbaum, 2010, p. 194)

Champion Tetris players during the Classic Tetris World Championships (CTWC) appear not to move their eyes as they play. Less skilled players, such as the 492 we sampled for our study, definitely move their eyes (Gray et al., 2015a) with the more novice student players moving their eyes in different patterns than the slightly more expert ones.

The study of *Perceptual-Motor Skill* has always been a strong part of the study of human expertise with some of our current theories and controversies tracing their roots back to the early 19th century (as discussed by, Shin, Proctor, & Capaldi, 2010) and with a strong continuing focus in the motor learning community (e.g., Wulf, 2013). In the last two decades, much attention has turned to (a) theories of event coding (Hommel, 1998, p. 143), (b) event structure and event segmentation (Zacks & Swallow, 2007; Zacks & Tversky, 2001), and (c) event-predictive cognition (EPCog) (Baldwin & Kosie, 2021; Butz, Achimova, Bilkey, & Knott, 2021; Kuperberg, 2021; Loschky, Larson, Smith, & Magliano, 2020). Each of these three is important for understanding the nature of skill in Tetris.

2.1.1. Theory of event coding

Our title for Section 2.1, *How Does the Perception of Action Affect Action Control?*, is taken from a question posed by Hommel (1998), but the question itself is puzzling … how could the "event" of *perceiving an action* affect something which comes before it; namely, the action itself? Answering this answer is both the focus and the contribution of the *Theory of Event Coding* (TEC) (Hommel, 2019; Hommel et al., 2001).

The theory's three most general assumptions (Hommel, 2019) are that goals for perception and action are (a) each coded the same way (the common coding assumption), (b) through distributed feature codes, and (c) which refer to distal features of the represented event. The challenge of the "common coding" assumption is the claim that part of the re-afferent signal comes from the movement itself, for example, from a hand movement or a head movement, and that such goal-directed movements can be made without any conscious knowledge about the motor system.

Part 2 of the above, the distributed *feature code* assumption, is that; "The integration of motor patterns with codes of *re-afferent effects* renders the latter effective primes of the former, so that an agent can simply reactivate …the action-effect codes which then tend to reactivate the motor patterns they are integrated with" (Hommel, 2019, p. 240). Although initially counter-intuitive, on further thought the logic is compelling. As Hommel emphasizes, perception and action are always closely linked. Almost anything that we perceive can be attributed to some prior movement/action on our part. Even an eye-saccade to an object or to a location is a movement and that movement is what makes the subsequent perception possible. Hence, by Hommel's account, perception and action are different parts of one two-stage event where the first part executes a particular movement to produce a particular sensory event. When we speak about the perception, we emphasize the produced event. When we speak about how we produced an event, we emphasize the action.

2.1.2. *Event segmentation*

Johansson, Hofsten, and Jansson (1980) moved the term *event perception* away from its older focus on "object motion in a passive perceiver" to "motion and space perception in connection with action" (p. 28). Cutting (1981) followed close behind Johansson with his "Six tenets for event perception;" namely,

(1) events have underlying structure;
(2) two classes of invariants divide event structure: topographic and dynamic;
(3) dynamic invariants divide into those of wholes and those of parts;
(4) dynamic invariants divide according to a minimum principle;
(5) dynamic and topographic invariants yield a center of moment; and
(6) centers of moment are perceptually useful.

All seems to have moved slowly in the study of event perception until two decades later when Zacks and coauthors, in a flurry of four papers published in 2001 (Zacks, 2001; Zacks et al., 2001; Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001) reinvented and reinvigorated the study of Event Perception for the new century. Perhaps, most importantly, for our current narrative, 2001 was the year in which Hommel et al. (2001) published "The Theory of Event Coding (TEC): A framework for perception and action planning" and Zacks published a small commentary on that paper as "Scaling up from atomic to complex events," which asked whether the claims made by TEC with brief events had implications for longer events, especially those longer events that are the focus of goal-directed activity. Hommel et al. (2001) had made two claims for the use of distal features; namely, that "action planning based on proximal features is inefficient" and that the "prediction of future stimulus input is easier with distal features." Zacks concluded his brief essay by saying, "the implications of the TEC view for complex events have been more or less assumed in the literature . . . apparently because the arguments for distal features become overwhelming as they scale up (p. 910)."

Event-Predictive Cognition is the offspring of a happy marriage between the *Theory of Event Coding* and *Event Segmentation*. The surge of interest in EPCog has resulted in an important collection of 16 interdisciplinary papers and commentaries recently published in this journal (see Butz et al., 2021, for the introduction and overview to this recent special issue). In its essence, "EPCog sets out to explore the extent to which event-predictive encodings and processes foster the development of abstract, conceptual, compositionally recombinable structures from sensorimotor experiences. It links event-predictive, conceptual structures to both sensorimotor and language structures" (Butz et al., 2021, p. 12).

2.1.3. The anticipatory behavior of expert performers

As one example of the ubiquity of EPCog and its importance for understanding variations in real-time, perceptual-motor expertise (including the game of Tetris), we focus on *representational momentum* (RM) (Blättler, Ferrari, Didierjean, & Marmèche, 2011; Didierjean, Ferrari, & Blättler, 2014; Hubbard, 2019).

> An observer's memory for the final position of a previously viewed moving target is often displaced forward in the direction of target motion. This displacement has been suggested to reflect the implied momentum of the target and has been referred to as representational momentum (Hubbard, 2005, p. 822).

Expert fighter pilots versus novice pilots RM produces different effects for experts than for novices. Indeed, experts often show the effect and novices often do not. For example, Blättler et al. (2011) found that expert pilots were more likely than novice pilots to make errors in judgment due to RM. Their study compared data from 21 novices (never in a cockpit and never in a flight simulator) with 15 expert pilots (French Air Force pilots).

After familiarization with the task, subjects were shown videos simulating the pilot's view of a landing. They were told that a black screen would briefly interrupt each video and that when the video resumed they would either be jumped ahead in time or back in time. Their task was to press the red key if they thought they had been jumped backward or the blue key if they had been jumped forward. However, they were not told that on some trials, no jumps would occur and, of course, that is the heart of this study.

As shown in Fig. 2, the Experts but not the Novices, misperceived: Experts made errors that were in anticipation of forward shifts of movement. Interestingly, these forward errors made by the experts are in contrast to their performance in estimating backward-shifts that showed no differences when compared to the novices.

The experimenters were surprised that the novices made pretty much no forward displacements and conducted two further experiments with only novices. In their conclusions, they speculate that two types of anticipatory processes may be at work. The expert pilots had available to them high-level semantic and strategic knowledge that could be used to "extrapolate the visual scene continuity." In contrast, without this knowledge, novices relied on sensory information "arriving at the retina when the cut occurs."

Expert versus novice in rugby, tennis, and tetris In a very description title, Anderson, Gottwald, and Lawrence (2019) study, *representative momentum in the expertise context* and find *support for the theory of event coding (TEC) as an explanation for action anticipation* in the game of Rugby. Of course, unlike Tetris, Rugby is a game played with teammates against an intelligent adversary; namely, the members of the opposing team. However, despite these differences, there are similarities between these two games that we find essential. For example, Anderson and colleagues point out that, "supporting evidence for the TEC (Hommel et al., 2001) also suggests that experience in action planning facilitates the tendency to anticipate similar actions of others toward distal effects when these actions are congruent with those previously learned" (Anderson et al., 2019, p.2).

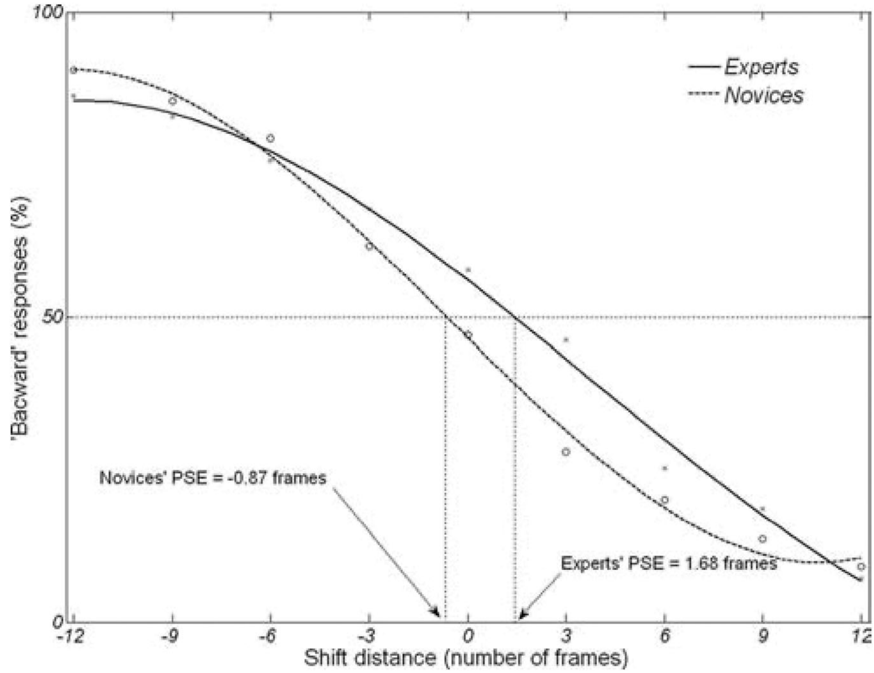


Fig. 2. Percentage of backward responses, by expertise level, shift direction, and shift distance in Experiment 1. PSE = point of subjective equality. (Figure and caption from Blättler et al., 2011.)

The key point for our paper is the commonality of some form of, "anticipatory behavior that is developed as a function of action planning experience." The commonality of *anticipatory behavior* is expected by TEC (Hommel et al., 2001) and expert-novice differences in anticipatory behavior have been found in Beach Volleyball (Cañal-Bruland et al., 2011), as well as Basketball and Volleyball (Vicario, Makris, & Urgesi, 2017), and (as per our discussion of Blättler et al., 2011) in Expert Pilots. Murphy, Jackson, and Williams (2018) put an interesting twist on this question by showing that skilled players can pick up information from the motions of their opponent and/or the movements of the ball.

Later in this paper, we present several statistical analyses that support the notion that, similar to the pilots and the more expert Rugby (Anderson et al., 2019), Beach Volleyball (Vicario et al., 2017), and Basketball and Volleyball (Murphy et al., 2018) players, our best student Tetris players have high-level semantic and strategic knowledge that allows them to think further ahead in the game than the lesser skilled players.

2.1.4. Section summary

In the title to this section we asked, *How Does the Perception of Action Affect Action Control?* and then followed that title with Rosenbaum's anecdote that suggested that tennis players of differing skills vary greatly on where they look *after* the racquet hits the ball. We then discussed the rise of EPCog from its roots in TEC and Event Segmentation. After that, we

Table 1
The events of Tetris

	Event Label	Description	Initiated by
Event 1	Episode	Zoid enters screen, falls until stops	System
Event 2	Rotation	Zoid rotated CW[1] or CCW[2]	Player
Event 3	Translation	Zoid moved left/right	Player
Event 4	Forced drops	Holding the down key increases drop rate	Player
Event 5	Filling	Plugging holes in the pile to clear 1-4 rows	Player
Event 6	Level change	Clearing 10 rows triggers level increase	System
Event 7	Speed ups	Drop rate increases from one level to next	System
Event 8	RNG	Sequence generation	System

[1]CW—clockwise.
[2]CCW—counterclockwise.

shifted focus to Blättler et al. (2011) work on *representational momentum* that, like Rosenbaum's anecdote, strongly suggests that experts and novices look at or perceive different types of information. This discussion brought us to representational momentum in the expertise context (Anderson et al., 2019) that led us to a brief discussion of TEC's expectations and expert-novice differences in anticipatory behavior in Beach Volleyball (Cañal-Bruland et al., 2011), Basketball and Volleyball (Vicario et al., 2017), and movements of opponents and/or the ball in Tennis (Murphy et al., 2018).

Perhaps paradoxically, as a culture we seem to think of jet pilots, racing car drivers, expert tennis, basketball, and volleyball players as people who are "faster than us." However, as our discussion in this section shows and as Alex Kerr warned us, "twitch speed" is not the limit of expertise; rather, human expertise is not a matter of doing the same thing faster but of doing the same thing differently; that is, mastering our tools is a process of acquiring the right techniques.

2.2. The events of tetris

Tetris is a simple game that can be defined by its events. These events or, at least, the ones encountered by our student players, are shown in Table 1. 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* (Cooper, 2021; Zacks, 2020; Zacks, Speer, Swallow, Braver, & Reynolds, 2007; Zacks & Swallow, 2007).

2.2.1. Episodes

Although the order of events in Table 1 is nominal, in many senses, the event type we list first, the *Episode*, is the most basic. 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. In the case of an empty or nearly empty board, a zoid could fall all 20 rows, coming to a stop once it hit the bottom of the

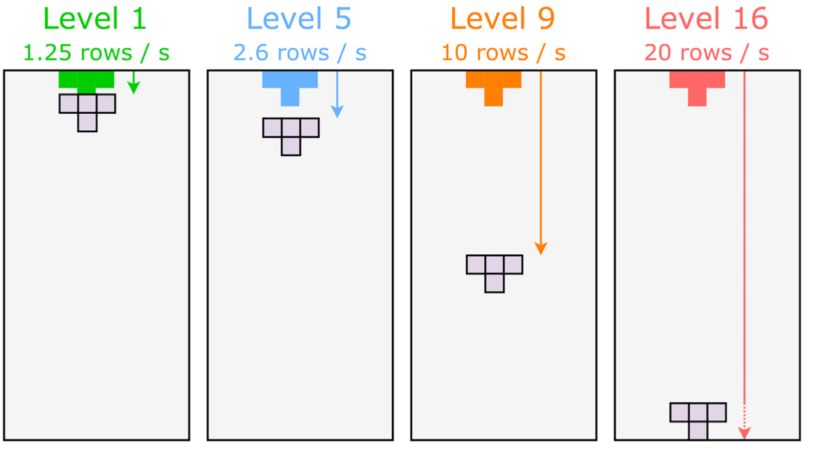


Fig. 3. Time pressure in Tetris: The drop rate increases as the play level increases. The figure shows how far the same zoid will drop in one second at four different levels.

board. However, most zoids stop before they hit bottom, by running into a zoid that is already in the pile.

As shown in Fig. 3, at level 1, the zoid will drop at the rate of 1.25 rows/s and, as there is no zoid accumulation shown in our example board, it will land at the bottom of the board in 16 s (see also Table 2). Not shown on the board and not further discussed in this paper is that from levels 19–28 the same zoid falls from top-to-bottom in 0.67 s, while for level 29 and above, it falls top-to-bottom in 0.33 s. Human performance at speeds like this is common during the annual CTWC but none of our 492 student players came closer than level 14 (see Table 2).

2.2.2. Player initiated movements: Rotations, translations, filling (rows cleared), and forced drops

Events 2–5 of Table 1 form the category of *Player Initiated Movements*. Three of these four movements are illustrated in Fig. 4. Event 2 requires the player to press one of two keys (or to use the keying options on the specially designed NES^{tm} controller, see Fig. 7) that will "translate" the zoid to the left or to the right. Event 3 requires pressing one of two keys that will rotate zoids (other than the square that does not rotate) clockwise or counterclockwise. Event 4 is evoked by holding the "down button" that hastens the rate at which the zoid drops. Event 5, *Filling*, which is not shown in our figure, consists of plugging a hole in the pile to clear one to four rows.

Interestingly and, perhaps, confusingly, in Classic Tetris, the zoids do not rotate as a naive user might think they should. As illustrated in Fig. 5, rather than rotating around the exact middle of the zoid, in Tetris, the rotation is lopsided. This is not a design flaw; indeed, experienced Tetris players, especially those at tournament levels of proficiency, view this as a positive feature. Suffice it to say that many, if not most, non-tournament-level players do not know that clockwise versus counterclockwise rotations may produce asymmetric results. Indeed, non-tournament level players when asked about "rotation direction" often report that

Table 2
The Tetris Drops Table for difficulty levels 1–16. As the level increases the seconds for a zoid to fall from top to bottom decreases as does the number of players who are able to play at that speed. Dashed lines separating levels 10–12 and 13–15 emphasize that there are no speedups between the three levels in these groups. Also, note that none of our 492 players make it beyond level 14

Difficulty Level	Sec to Fall	Speed Up %	Players Left	% Lost fr Prior Lvl	% Cum Total Lost
0	16.0	0	492	0.0	0
1	14.3	10.4	485	1.4	1.4
2	12.7	11.6	455	6.2	7.5
3	11.0	13.2	405	11.0	17.7
4	9.3	15.2	344	15.1	30.1
5	7.7	17.8	277	19.5	43.7
6	6.0	21.8	231	16.6	53.0
7	4.3	27.8	161	30.3	67.3
8	2.7	38.3	68	57.8	86.2
9	2.0	25.1	26	61.8	94.7
10	1.7	16.5	11	57.7	97.8
11	1.7		7	36.4	98.6
12	1.7		5	28.6	99.0
13	1.3	20.4	1	80.0	99.8
14	1.3		1	0.0	100
15	1.3		0		
16–18	1.0	24.8	0		
19–28	0.67	50.0	0		
29–30	0.33	100.0	0		

they try to simplify their life by always rotating all Tetris zoids in the same direction (i.e., either all clockwise or all counterclockwise). (Also see Fig. 6.)

A more experienced player (one who routinely qualifies for a seat in the CTWC playoffs), kindly explained to us that he, too, used to rotate all zoids in the same clockwise direction and it was not until he was good enough to routinely reach and die at levels 15 and 16 that he realized the error in his ways. He also reports that one of the most grueling parts of his life as a Tetris player was the 6 months he spent unlearning strict clockwise rotations and learning to always rotate each piece in the direction that used the least clicks for his planned zoid placements.

2.2.3. *Rows cleared and speed ups*

Each Tetris row is 10 blocks wide. As shown in Table 1, by filling in all of the empty spaces, Event 5 can clear one to four rows. However, the better the player, the more likely she is to attempt to build a solid wall of zoids from, for example, columns 1–9 while holding column 10 open. If a player can hold open one column that spans four contiguous rows, and if an I-beam drops while that column is being held open, then a four row "wall" of zoids will dissolve and the player will score eight times as many points as she would by filling one row, by itself, four times—this maneuver is called a *Tetris* and is what gives the game its name.

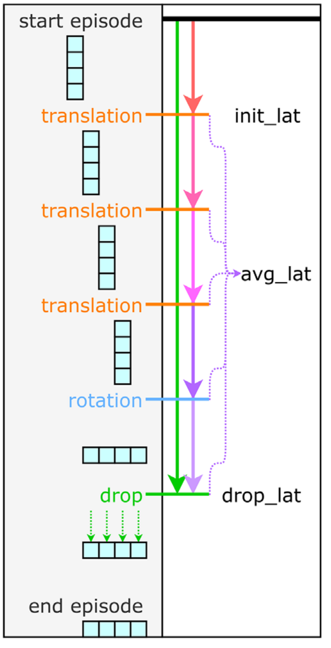


Fig. 4. Player-initiated movements. The figure illustrates left-to-right translation, rotations (clockwise or counter-clockwise), and drops.

The sixth *Event of Tetris* shown in Table 1 is the level change. As Table 2 shows in column 2, completing each of the first 10 levels of Tetris increases the drop speed of the zoids. Hence, at level 6, a zoid would take 6 s to drop from the top to the bottom of an empty board, whereas at level 7, it would take only 4.3 s to drop that same distance.

2.2.4. The tool of tetris

One of two types of tools are usually used for controlling Tetris. At its simplest, a computer keyboard can be used with actions mapped, typically, to the "a" and "d" key (for use by the left hand to translate the zoid left or right), the "k" and "l" key (for rotating a zoid counterclockwise or clockwise, usually using the right hand), and the "s" key (for dropping the piece faster than it would otherwise fall—also using the right hand). However, the tool of choice for those who frequently play Tetris is the NES Game Controller, as shown in Fig. 7. This tool also enables the five movements described above. More recent Game Controllers also exist. These duplicate the functionality described above but have the advantage of a more ergonomically engineered, two-handed grip. Popular in this category is the XBox controller. The hand-held ergonomics of this device seem superior to even the casual player.

2.2.5. Random number generation—The tortoise and hare in tetris

One of the most important but seldom discussed attributes of Classic Tetris is the random number generator (RNG) that creates a randomized stream of numbers 0–6, with each number

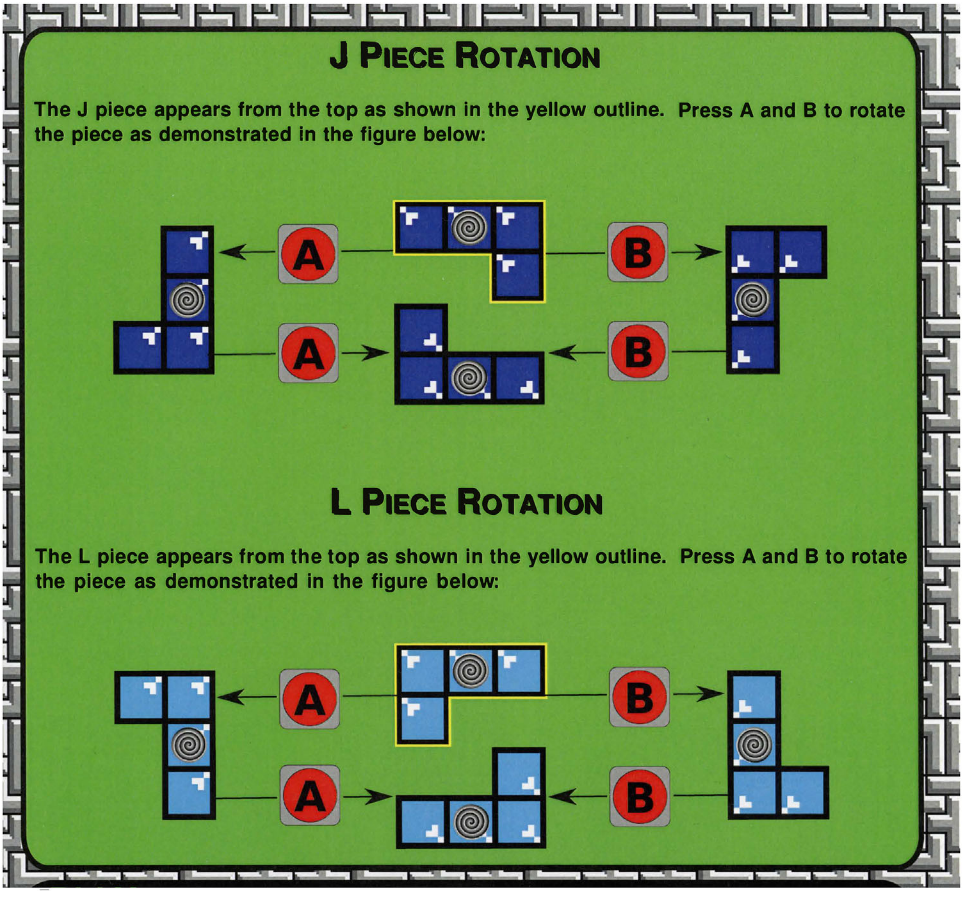


Fig. 5. The J and L piece each have four orientations and, for intermediate players under pressure at the limits of their skill (generally around levels 12–16), each are confusingly similar. The figure shows that three counter-clockwise clicks are needed to rotate the zoid from its top orientation to the rightmost orientations. In contrast, the player could achieve the same results with one clockwise click (see the leftmost J in the top row). (From Smith, 2015, with permission, page 7.)

representing one of the seven types of zoids. The stream of zoids generated by Classic Tetris differs from that generated for many other versions of Tetris in that, for those versions, each number 0–6 appears once in each string of seven numbers. In contrast, for Classic Tetris, the number stream is as random as the RNG can make it. The downside of this for CTWC Tournament play is that the best players always try to keep a deep “well” open (usually in the right-most column) so that if an I-beam appears they can plug the well and score many points. Games can be lost if players cannot maintain a clear “well” for the I-beam or if the well becomes clogged near the top of the screen so that other pieces pile up to the top of the screen. As CTWC players build to maximize points, they are usually trying to build walls of zoids with one column empty so that when a I-beam comes along they can drop it into this column and score big points.

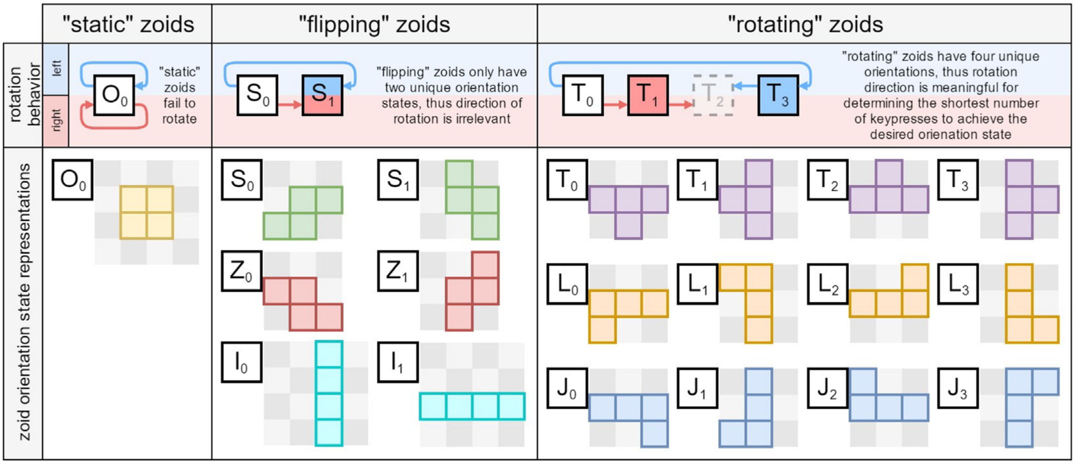


Fig. 6. The O (or square) is a static zoid that does not rotate. The S-, Z-, and I-beam each have two orientations. Pressing left or right to rotate has the same effect. The T, L, and J each have four orientations. At drop speeds higher than level 15, the wrong decision to rotate clockwise or counterclockwise can cost the player the game by increasing the rotation time. See also Fig. 5.

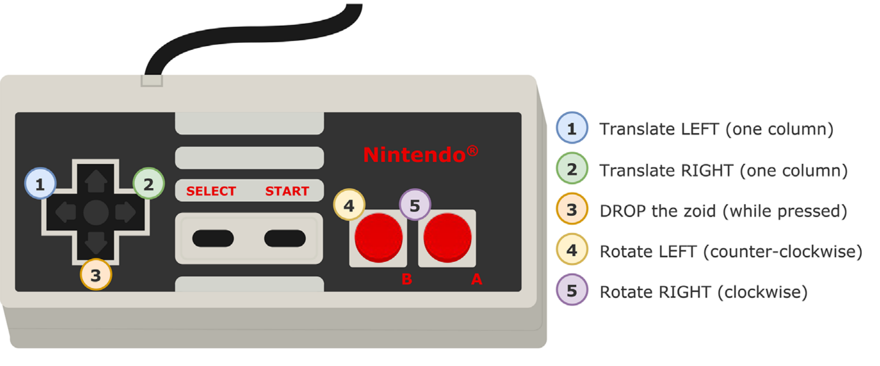


Fig. 7. NES Game Controller—containing a variety of tools for controlling Tetris.

In a study that, in part, examined the influence of the RNG, Sibert and Gray (2018) ran 1,771,561 games of Tetris by systematically varying 11 weights on each of six features for two different game lengths. In the short length condition, all models were run either until they died or until they had played 506 zoids. In the long length condition, all models ran until they died.

After the fact, we named the longest playing model the *Tortoise* model and the highest scoring short model, which played 506 zoids, the *Hare* model. The best Tortoise model scored 34,847,635,400 points by clearing 125,829 lines. It was nearly 10 billion points higher than the next highest scoring model (27,572,380,920) that had cleared 107,934 lines. For our short game (i.e.,"human length") game, the best Hare model scored 240,900 points, clearing 199

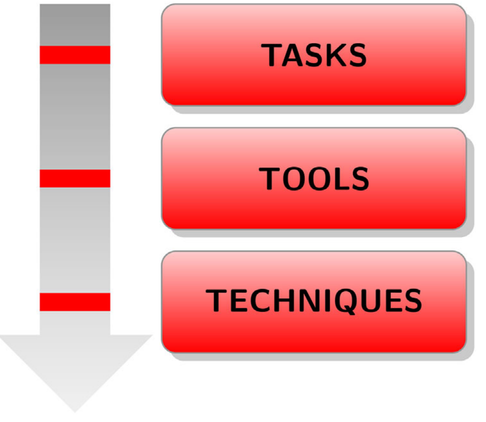


Fig. 8. Tasks, tools, techniques: Tasks require tools and tools may be manipulated by variety of taught or discoverable techniques.

lines. This score is close to the highest human score we have collected in our dataset of 492 Rensselaer students that we discuss in the Methodology and Results sections of this paper.

2.2.6. The events of tetris: Summary and conclusions

As suggested by Fig. 8, in common with many human tasks, Tetris can be considered as a task or set of subtasks, with tools that are used for performing the tasks, and different techniques available (or discoverable) for using the tools for performing those tasks. Within an episode (see, Event 1, Table 1), within each step of each zoid's fall, there can be one instance of one of the four types of player initiated movements. Players can rotate the zoid clockwise or counterclockwise (rotations), move it left or right (translations), force an early drop, or fill in a gap in the row (which briefly creates one to four solid lines of filled-in rows and then dissolves). These first three movements usually occur in combination with each other, and as the last event (filling in one to four rows) stops the zoid from falling, it always ends the episode.

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. 1) and by changes in the color schemes for the seven 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 events is speedups in the drop rate (see Table 2); however, for other than our 17 best players, speedups are synonymous with level changes. That is, for the first 10 levels of play, levels 0–9 (see Table 2) each change in level results in an increase in drop rate (and a color change). But, as Table 2 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 492 college players make it beyond level 9, the intricacies of this fifth level of event structure are largely ignored in this paper. Tetris supports player events other than those shown in Table 1; however, those other events represent Extreme Expert maneuvers that our laboratory players either do not know or cannot execute.

3. Methodology

Tetris data were collected from undergraduate students at Rensselaer Polytechnic Institute between Fall 2014 and Spring 2019 during 50-min gameplay sessions using the Meta-T software (Lindstedt & Gray, 2015). Game state information was logged at 30 Hz and player actions were recorded at 60 Hz. Data from 240 of these players were reported by Lindstedt and Gray (2019). However, in addition to considering different research issues than the Lindstedt and Gray (2019) study, the current study (a) reports, (b) segments, and (c) analyzes data differently than those of Lindstedt and Gray (2019). Hence, in addition to containing more than twice as many players, our current study also includes data that allow us to address new sets of research questions.

3.1. Participants

Players were recruited from the Cognitive Science Department's Undergraduate Subject Pool. All experimental procedures were reviewed and approved by Rensselaer's IRB.

3.2. Task

Games of Tetris are of variable length. Players play each game until they die, and they always die. However, the higher skilled players take longer to die than the lesser skilled ones; hence, in the 50-min play period, lesser-skilled players play more, but shorter, games of Tetris.

All our players used Meta-T (Lindstedt & Gray, 2015), which implements a version of Tetris that is close to the original Nintendo Entertainment System (NES) Tetris. Meta-T is implemented in Python that results in minor visual differences between it and NES Tetris. Software experts at the CTWC have examined Meta-T and confirmed that, behaviorally, it is a faithful version of Classic Tetris up to level 19. At level 19 and above, 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 duplicate. However, these differences can be ignored since, as shown in Fig. 2, very few players from our pool of participants managed to cross level 9, and among those who did, the maximum level reached was level 15.

3.3. Random number seeds

In collecting the data, to ensure that each participant played the exact same set of games in the exact same order, the same ordered sequence of 10 random seeds was used across game sessions for all student players. For players who played more than 10 games, beginning at game 11, Meta-T would cycle back to the first seed and keep cycling through the sequence

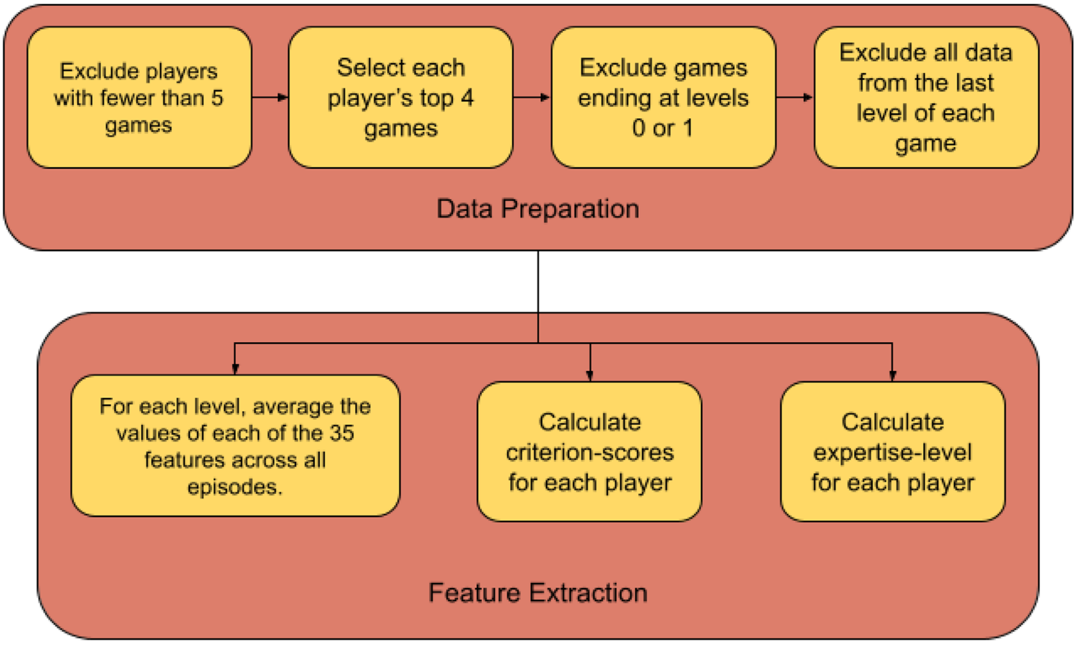


Fig. 9. Flowchart for data preparation and feature extraction.

until the end of the 50-min gameplay session. Finally, to be very clear, each player played as many games during their 50 min gameplay session as they could. The weaker players played more but shorter games, whereas stronger players played fewer, but longer games.

3.4. Gameplay

All games were played in the CogWorksLab's Acoustic Pods that ensured that each player was isolated from any and all lab noises. At the end of each game, players were required to click on an icon to start the next game. They were also encouraged to take breaks, between games, when and if needed. All sessions entailed 50 min of gameplay and all game actions were controlled using an NES controller connected to the computer's USB port through an adapter. Most players were eyetracked; however, eyetracking results will be the subject of a future report. After the game session, players completed a brief exit survey and were debriefed.

3.5. Data preparation and feature extraction

The steps in our data preparation are shown in the top row of Fig. 9 and the steps in feature extraction are shown in the bottom row. Before data preparation, our complete corpus of Meta-T Tetris contained 2772 games collected from 499 players.

3.5.1. Data preparation steps

(1) Tetris is an unforgiving game and most players, even those who play in the annual CTWC, make slips that create play sequences that spiral out of control and quickly end the game. To exclude such *early death* games, our analyses only consider the top four games played during game sessions that were, at least, five games long. This resulted in the elimination of seven players.

(2) Ranking players: Level-based and score-based ranking
  (a) To meet the needs of our logistic regression models, we established a *level-based ranking measure* based on the levels reached in each player's top four games (as discussed in Section 4.2).
  (b) To meet the needs of our linear regression models, we used a *score-based ranking system* that took the mean score of each player's top four games as their *criterion score* (as discussed in Section 6.1).

(3) Games that ended at Tetris level 0 or 1 were excluded, since they were deemed to contribute more noise than useful information to the analysis.

(4) Data for the last level of gameplay (the level at which the player died) were removed to ensure that we were only looking at stable performance data.

Data preparation and extraction ended with 1962 games played across 494 sessions for 492 players. The reader will note that we had two more sessions than players. For these players, the software running the session somehow failed but was restarted by the experimenter while the player remained in the Acoustic Pod. For our analysis, we merged the data from the two session files corresponding to each player and treated the merged data as a single session (which it was) and performed steps 1 through 4 for each merged session.

3.5.2. Gameplay features

In Table 1, we introduced eight high level events that can be used to describe the behavior of the Tetris system and players during the game. In contrast, our statistical analyses of Tetris play is based on 35 features that can be combined and analyzed to describe the various states which these events may assume during the game. All readers are encouraged to turn to Appendix A to glance over our list of features; however, no reader should feel compelled to read this list or to read any other of our appendices unless they desire a deep dive into both data and features.

Many features were either adapted or wholly adopted from Lindstedt and Gray (2019), whereas most of the remainder are based on Smith (2014) excellent guide, *Tricks of the Classic NES Tetris Masters*. All features fall into two broad categories: (a) Board state: during gameplay, the Tetris board changes dynamically. A few dimensions by which the board state can vary include; mean board height, empty space (or gaps) that are surrounded by pieces, and construction by the player of empty spaces reserved for certain Tetris pieces, such as leaving open the rightmost column in the hopes of plugging it with an I-beam, and (b) player behavior: as a zoid falls, players can move it laterally, rotate it, and/or force it to drop faster than it would otherwise fall (see also Table 1, *The Events of Tetris*, Events 2–5).

The between-player variations in board state and actions reveal player plans and strategies that enable us to characterize player expertise based on the types of skills players present.

3.5.3. Aggregating gameplay features

The data preparation steps (Section 3.5.1) provided a set of 35 feature values for each episode of gameplay. Although this allows us to access very fine-grained data, Tetris players often make slips and such slips can introduce unwanted noise for certain episodes. Hence, to mitigate this noise, for each game of each player, we averaged the values of our 35 features across each difficulty level (see the leftmost column of Table 2 and the bottom nodes of Fig. 9). The resulting *level-averaged features* were used to perform our analysis. Appendix A lists all 35 level-averaged features along with their descriptions, and information about how they were calculated.

3.6. Section summary

The methodology reported in this section has been a stable feature of our laboratory since prior to the publication of Lindstedt and Gray (2015). We refer to this initial hour of Tetris play as our *Population Study*. After playing Tetris for an hour, many of our players go on to a second or third session in which we use the Population Study data to calibrate each player's level of Tetris skill. By obtaining performance data during the Population study, we have been able to assign players of approximately equal Tetris skill to different conditions in those other studies.

Some of these data have formed the basis of more specialized studies that compare human performance with machine learning models (e.g., Sibert, 2015, 2019; Sibert & Gray, 2020; Sibert, Gray, & Lindstedt, 2015; Sibert, Lindstedt, & Gray, 2014; Sibert, Speicher, & Gray, 2019), whereas others have focused on the role of eye movements during Tetris play (Gray, Hope, Lindstedt, & Destefano, 2014; Gray, Hope, Lindstedt, & Sangster, 2015b; Gray et al., 2015a, 2018). Destefano et al. (2011) revealed that epistemic action is, at least for the game of Tetris, a novice, not an expert ploy as had been assumed (Kirsh & Maglio, 1994). Also, many studies resulted in unpublished undergraduate theses. The Lindstedt and Gray (2019) paper, mentioned above, is the immediate predecessor of the current study.

4. Establishing the basis of expertise differences between beginner, intermediate, and expert players

To identify the latent factors (a linear combination of the features) that account for most of the variance in the data, EFA was applied to the set of feature values averaged by level. Each selected factor is associated with a specific type of skill based on the features that constitute it. After these factors are extracted, to verify whether they are useful for explaining differences in player expertise, two types of multivariate regression models were used. First, we applied linear models to identify factors that discriminate among players belonging to a specific expertise group. Second, after the linear models, we applied logistic regression models to identify factors that might explain differences between expertise groups. (As discussed in Section 3.5.1, above, the expertise level for each player was calculated by taking the mean of the final level of gameplay for their top-four games.)

The best linear models were derived using a bidirectional step-wise model selector based on the Akaike information criterion (AIC). AIC entails an iterative process of input variable selection based on the significance of the information each variable contributes to model fit. Finally, further analyses were performed to determine the influence of random seeds on gameplay (these analyses will be discussed in Section 7).

4.1. Exploratory factor analysis

Fig. 10 shows the correlation matrix constructed from the level-averaged feature values in the data. The heat-map for the correlation matrix is shown on the left side of the figure with the numbers for each entry shown on the right side. These values provide the input to our EFA. Appendix A lists all 35 level-averaged features along with their descriptions, and information about how they were calculated.

Factor analysis finds sets of correlated features and uses these sets to form individual factors. The method used here for identifying latent factors (Costello and Osborne, 2005) is *principal component analysis* (PCA). PCA finds linear combinations of features in the original data, called components. The weight/contribution of each feature for a component is its loading value (see Figure 11).

The first component captures the highest amount of variance in the distribution of the data, the second component captures the second highest variance in the data, and so on (Wold, Esbensen, & Geladi, 1987). By default, these components are orthogonal to each other, which means that there is no collinearity present among the components.

In general, it can be difficult to clearly determine the type of information each component carries. However, rotation of components solves this problem, as the components become factors that represent linear combinations of subsets of the original features. The loadings of other less important features are awarded near-zero values, which can then be ignored. By examining the features that constitute each rotated factor, it is possible to specify the kind of information the factor carries. For our rotations, we used varimax-rotation, which is one of the most commonly used forms of orthogonal-rotation (Jackson, 2005).

Our PCA used level-averaged features (explained in Section 3.5.3). One of the commonly recommended methods for selecting the number of to-be-retained factors is the Kaiser rule, that is, select all factors whose eigenvalue is greater than 1 (Kaiser, 1960). However, Costello and Osborne (2005) warn that this method often leads to suboptimal results (because analysts end up retaining too many factors) and suggest other methods for the selection process. Interestingly, the human eye is generally considered at least as accurate as an algorithm for this process so that the most common method entails plotting the data to then look for the inflection point (as per the Fig. 12 plot for our current data set). In such plots, the horizontal line represents an eigenvalue of 1 (serves as a reference line, factors below that should not be selected for analysis) and the vertical line is the point at which the slope

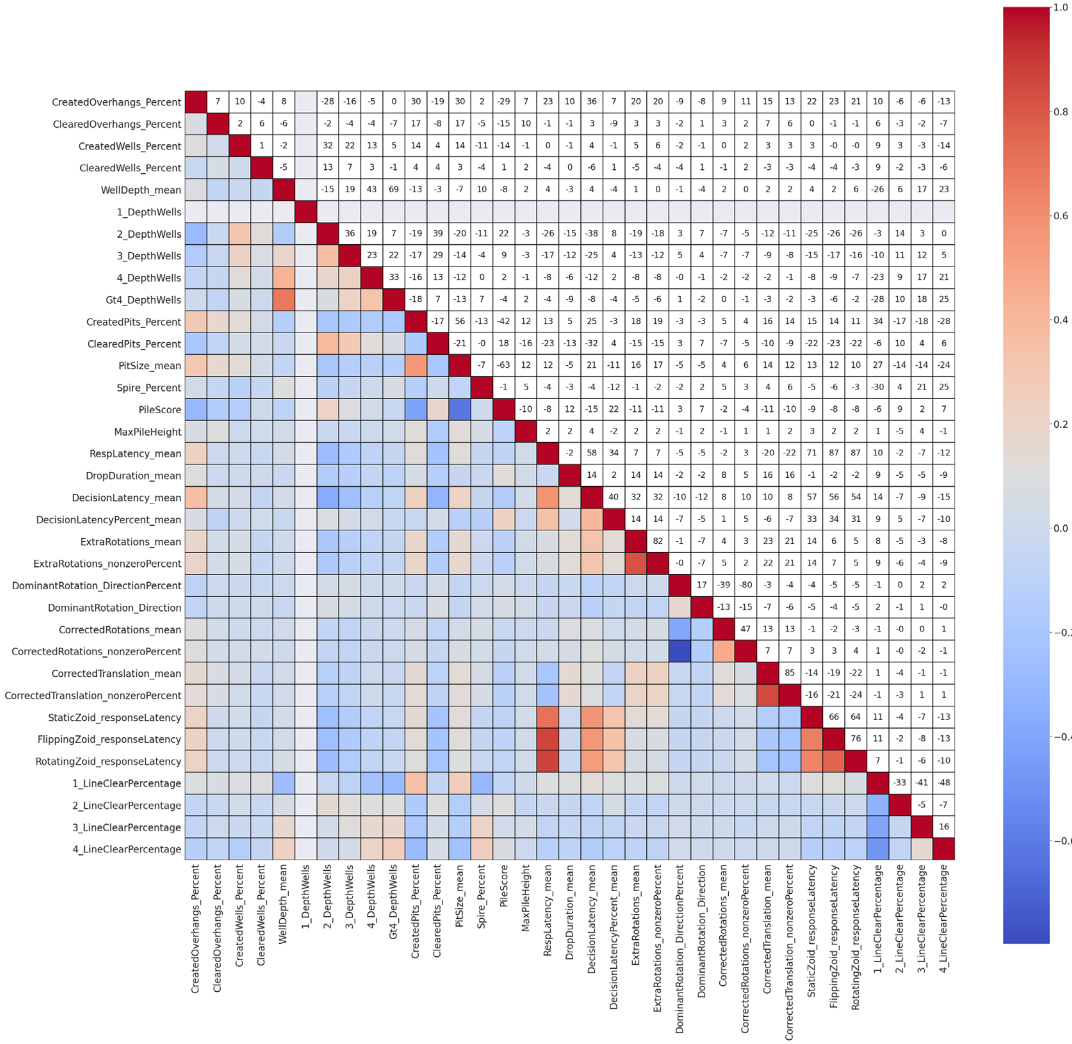


Fig. 10. Combined heat-map and numeric weighting of the correlation matrix (Kendall's Tau) for all 35 derived features. The numeric weightings provide the correlation values (shifted by two decimal places) for the heat-map. A quick scan of the heat-map reveals several small clusters of positively correlated features and a small number of (strong) negative correlations. Some combination of features belonging to each of these clusters correspond to the higher level behaviors of our players. The exploratory factor analysis (EFA) defines composite factors based on linear combinations of the correlated features.

of the curve inflects. For these data, the inflection point supports a decision to retain the first six factors and discard the rest. In our case, these six factors explain a total of 52.6% of the total variance, with each explaining 12.1%, 10.4%, 9%, 7.4%, 7%, and 6.7% of the variance.

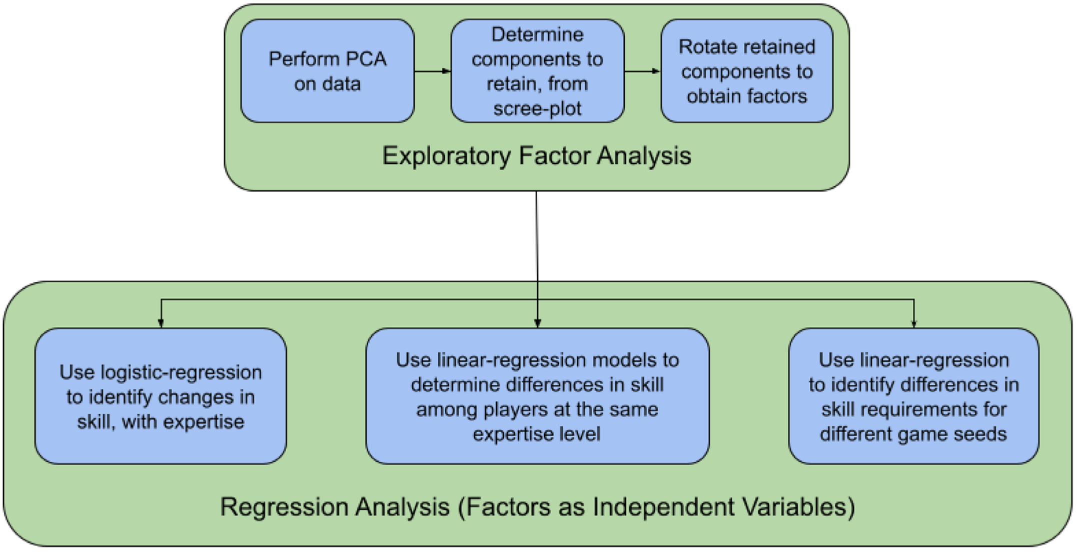


Fig. 11. Flowchart showing the steps in our exploratory factors analysis (EFA) (top line) and the various types and functions of regression analyses used (bottom line).

The loadings for our 35 features for each of the six factors are presented in Appendix B. From these loadings, we conclude that the factors contain the following information:

(1) Factor 1 (planning-efficiency): How fast players can decide the best placement position for a zoid and react by taking the necessary actions. Lower values of this factor are indicative of faster planning and action, whereas higher values indicate slower performance.
(2) Factor 2 (pile-management): How well can the player manage the pile of zoids. Bad pile management includes too many holes, deep crevices, a central spire or hanging structures, and greater pile heights. Such messy piles result from bad zoid placements and make line-clears difficult. Higher values on this factor are associated with bad pile configurations.
(3) Factor 3 (zoid-control): For each zoid placement, there is a minimum number of rotations and translations that are needed to move the zoid to its final position. A high value for this factor indicates that the player is performing more than the minimum rotations and/or horizontal movements needed to move the zoid to its destination.
(4) Factor 4 (pile-uniformity): The shape of the top of the pile (depressions and spikes) is a very important part of Tetris gameplay. Piles that are “too flat” make it difficult to place zoids, especially the asymmetric zoids (i.e., J, L, Z, and S). Concurrently, piles with deep wells tend be hard to manage, because such wells can be difficult to fill up. Piles that are smooth at the top are indicated by a lower score for this factor, while higher scores imply more jaggedness.

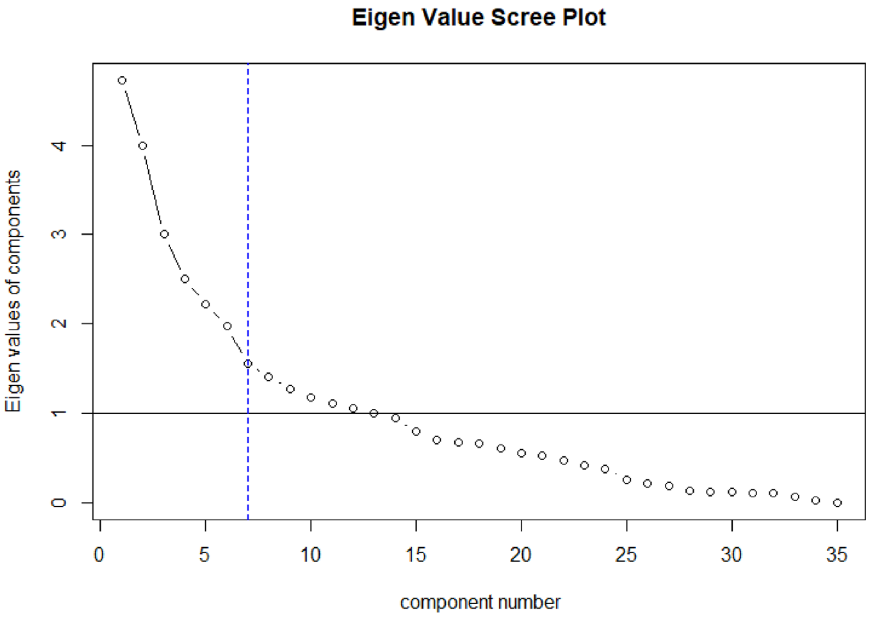


Fig. 12. Scree plot of eigenvalues of all 35 components of the PCA.

(5) Factor 5 (minimum-line-clears): This factor tells us the extent to which players make more one or two line clears as opposed to three or four line clears. Fewer line clears generate less score and are mostly seen when players struggle to maintain their pile. For this factor, a higher value indicates more single or double line clears.

(6) Factor 6 (rotation-corrections): Of the seven Tetris zoids, the Square neither flips nor rotates, three (the S, Z, and I) flip, and three (the T, L, J) rotate both clockwise and counterclockwise (see Fig. 6). If a slip is made so that the zoid overrotates, the rotation-correction factor penalizes the player for having to make extra corrective rotations to achieve the desired orientation. A high value on this factor indicates fewer unnecessary rotations.

4.2. Defining player classes

Our logistic regression model allows us to determine differences in skill among groups of players. To rank player expertise, we averaged the final level of Tetris gameplay for the player's top four games, rounded to the nearest integer. For example, if the top four games of a player ended at levels 8, 9, 10, and 10, their expertise level would be rated at 9.

Fig. 13 shows the distribution of expertise levels among our 492 players. The distribution is right skewed because there were very few players who were able to survive at level 9 or higher. Indeed, only two of our players were rated higher than level 10. (See also the *Difficulty Level* and *Players Left* column of Table 2.)

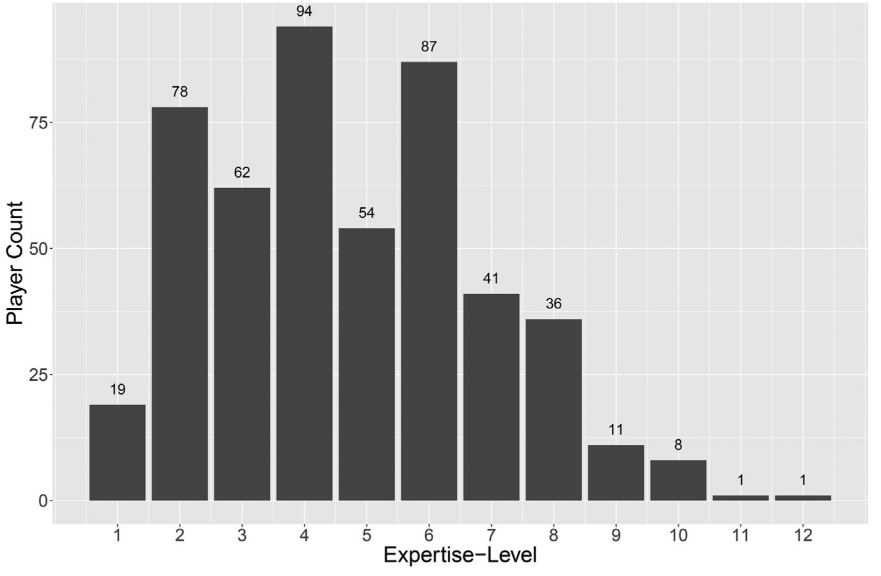


Fig. 13. Distribution of expertise levels in the data. Beginners players correspond to expertise level 3 (62 players), intermediate players are at expertise level 6 (87 players), and our experts are everyone who belong to expertise level 9 or higher (22 players).

We used a clustering algorithm to define distinct groups of players based on expertise, for comparison of skills. *Clustering* is a collection of unsupervised classification algorithms that divide any given data into groups of similar data-points, based on dimension(s) of variation in the data. (See Appendix F for more details about our clustering process.)

In this case, we use the k-means clustering algorithm, which divides the data into k clusters (value of k supplied by the analyst). A challenge posed by this algorithm is defining the correct value of k. A favorite choice for selecting an optimal value for k is the elbow method (Marutho, Hendra Handaka, Wijaya, & Muljono, 2018); which we also used for factor selection in PCA (above).

For cluster analysis, the y-axis represents the sum of squared error (SSE) for the data as a function of the number of clusters (x-axis). The data were divided into at least 2 to a maximum of 10 clusters, and the SSE was calculated in each case. An elbow was observed at $k = 3$ clusters, so we divided our data into three clusters. The results were also subjected to other verification processes to confirm we indeed had the optimal number of clusters (discussed in detail in Appendix F). The verification process supported 3 as the optimal number of clusters for our data. From the results of the clustering process, we determined three player groups for comparison (see Fig. 14):

(1) Beginners: Players with an expertise level of 3.
(2) Intermediates: Players with an expertise level of 6.
(3) Experts: Players with an expertise-level of 9 or higher.

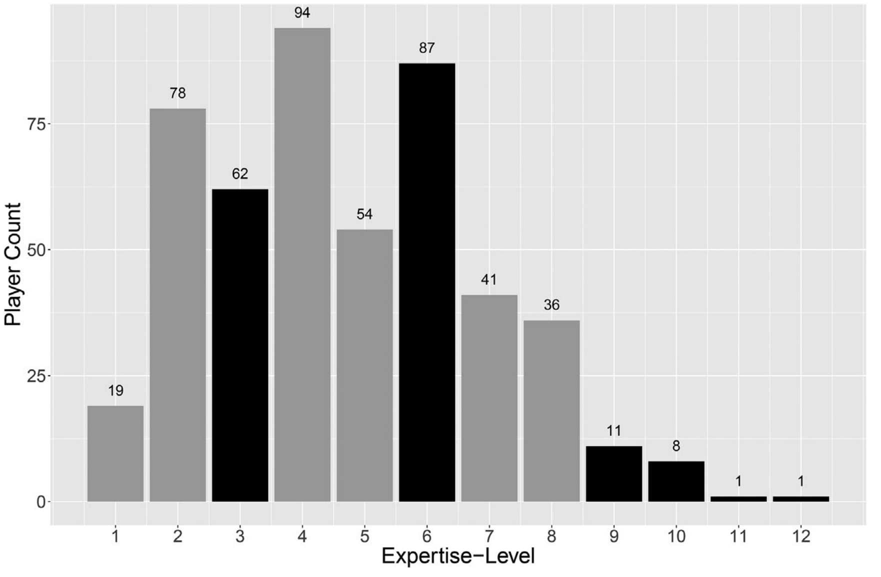


Fig. 14. Black bars represent expertise levels that were selected for the analysis.

Table 3
Distribution (mean and standard deviation) of the number of games played (during each session) and final game level (for each game) for players from each expertise level

Expertise-Level	Mean (Game Count)	Std. Dev. (Game Count)	Mean (Final Level)	Std. Dev. (Final Level)
Beginner	11.27	4.71	1.64	1.48
Intermediate	8.27	1.57	4.06	2.37
Expert	7	1.28	6.96	2.91

Expertise groups were defined on specific expertise levels to widen the gap between the groups, for example, players from expertise level 4 and 5 were purposefully left out (from beginner or intermediate groups), because, in all likelihood, it would be very difficult for statistical models to differentiate between skills of expertise level 4 and expertise level 5 players (for a detailed explanation, refer to Appendix F.) Readers should also note that Fig. 14, which shows the player groups selected for analysis, is derived from Fig. 13. For our sample, there are 62 beginners (level 3), 87 intermediate (level 6), and 21 expert players (the sum of expertise levels 9–12). Finally, Table 3 presents the distribution for number of games played by players from each expertise level and the length of the games (expressed as the highest level of gameplay for each game). In general, beginners play more but shorter games, whereas experts play fewer but longer ones.

5. Two at a time: Using logistic regression to compare player groups

Logistic regression models (also called generalized linear models or logit models) are often used for classifying categorical binary data. As we are trying to compare two groups of players with each model, logistic regression is perfect for our use-case. These models first perform a linear combination of input variables and then feed the output into an activation function, which converts the values into binary outputs. The models are first trained to fit the data, and then the goodness-of-fit is determined by calculating the mean squared error.

5.1. Overview of analyses

We fed our regression models the six factor values (defined in Section 4) as input and trained each model to classify players as belonging to one of two classes. (For example, for model 1, factor 1, the best fit for planning-efficiency was −0.225. Also, note that the sign, positive or negative, is not important; the best model is the one with the greatest absolute size.) The best version of each model was determined through bidirectional, step-wise model selection based on AIC. Four logistic regression models were trained:

- Model 1: Trained to distinguish between beginner and intermediate players, from factor values corresponding to gameplay at level 0.
- Model 2: Trained to distinguish between beginner and intermediate players, from factor values corresponding to gameplay at level 2 (last level of stable gameplay for beginner players, on average).
- Model 3: Trained to distinguish between intermediate and expert players, from factor values corresponding to gameplay at level 0.
- Model 4: Trained to distinguish between intermediate and expert players, from factor values corresponding to gameplay at level 5 (last level of stable gameplay for intermediate players, on average).

It should be noted that as we move from model 1 to 2 and model 3 to 4 (models that are fit to the same population but at different levels of gameplay), we still retain the same sample of players but the number of games per player reduces. This is because beginners do not survive level 2 for all of their games and intermediate players do not always make it beyond level 5. In general, only a subset of the games considered by models 1 and 3 (at level 0) is also considered for models 2 and 4 (levels 2 and 5).

5.2. Results

The model fits were evaluated using 10-fold cross-validation for predictive performance (mean squared error). Table 4 lists the results for the model fits.

Model 1 uses data collected from Beginner and Intermediate players at level 0. Even at level 0, the pattern of performance of these two levels of players differ on planning efficiency (Factor 1), zoid control (Factor 3), pile uniformity (Factor 4), and minimum lines cleared (Factor 5).

Table 4
Beg versus intermediate: At Game Level 0 and, again, at Game Level 2. Intermediate versus expert: At Game Level 0 and, again, at Game Level 5

Model Information				Factor Information					
Model	Population	Game Level	MSE	Factor 1 plann-effic	Factor 2 pile-mmt	Factor 3 zoid-cntrl	Factor 4 pile-unif	Factor 5 min-l-clears	Factor 6 rot-crrctns
Model 1	B versus I	0	0.194	-0.225^{***}	–	-0.233^{***}	0.076^{*}	0.124^{**}	–
Model 2	B versus I	2	0.148	-0.277^{***}	-0.270^{***}	-0.162^{***}	0.107^{*}	–	–
Model 3	I versus E	0	0.096	-0.240^{***}	–	-0.308^{***}	0.256^{***}	0.373^{***}	0.330^{***}
Model 4	I versus E	5	0.149	–	–	–	0.151^{**}	0.103	0.093

Significance Codes: $p < .001$ ***; $p < .01$ **; $p < .05$ *

Model 2 was collected from the same population as Model 1 but looked at data collected during performance at level 2. Three of the same factors are significant here as for level 0. However, the Intermediate players are superior at pile management—a factor that becomes more important as the game speeds up. Interestingly, for Model 2, the minimum number of lines cleared (Factor 5), no longer differentiates the Beginners from the Intermediates, perhaps indicating that the stress of the faster drop rates at level 2 (16 s to fall at level 0 vs. 12.7 s to fall at level 2) is enough to diminish the small advantage that the Intermediate players had over the beginners.

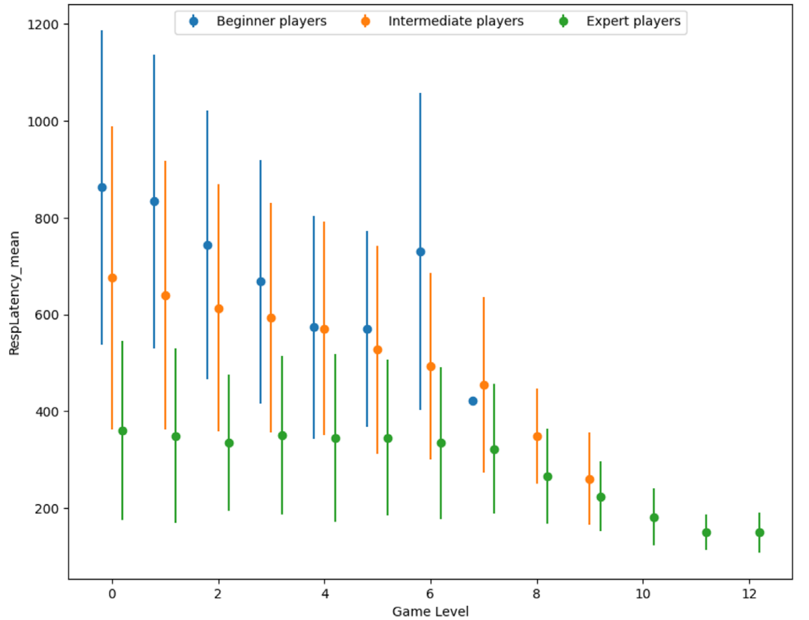
Model 3 compares the Intermediate against the Expert players at level 0. Skill differences between these expertise levels seem to be determined by all factors except Factor 2 (pile-management). The lack of a significant difference for pile-management between Intermediate and Expert players suggests that pile-management is a skill which Intermediate players have mastered.

Model 4 compares Intermediate with Expert players at level 5. This is a complex comparison that suggests that, with the exception of *pile-uniformity* (Factor 4), the other five factors have lost the power to discriminate between expert and intermediate players. For pile-uniformity, the four largest contributing Game Features are WellDepth_mean (0.823), Gt4_DepthWells (0.650), 4_DepthWells (0.633), and 3_DepthWells (0.584) (see Appendices A and B). The significant pile-uniformity factor suggests that experts have learned more about curating the board to avoid or remove gaps or holes, and more about setting up the board so that they can remove one, two, three, or four lines at a time.

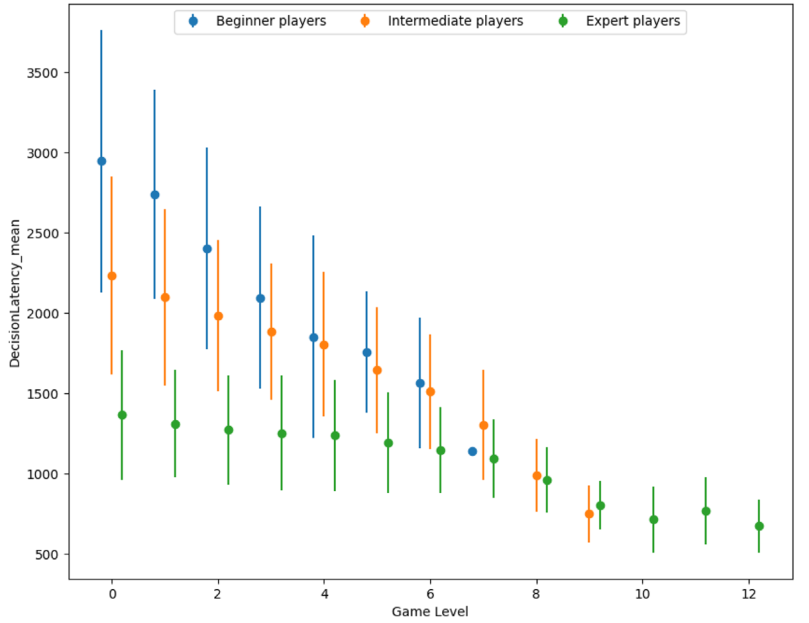
Fig. 15 demonstrates how players adapt their behaviors to changing task demands, which explains why differences in skill between Intermediate (orange) and Expert players (green) (which are relevant at level 0, Model 3) disappear at level 5 in Model 4. For two of the most important features in planning-efficiency (i.e., ResponseLatency and DecisionLatency), the plot shows changing group behavior with increasing game difficulty; however, this factor becomes irrelevant in Model 4 (i.e., around game level 5 of Fig. 15).

The overlap of error bars (between groups) might be thought to undermine the significance of the differences in mean values. However, that would be an incorrect assumption, since partial overlap of standard deviation bars should not be interpreted as evidence against significance of group differences (Krzywinski & Altman, 2013). Also, our purpose for plotting standard deviations is to show the distribution of the values for each group; that is, we do not intend it as a tool for measuring significance of differences between groups.

As difficulty increases, the group means allow us to understand the overall trend in behavioral changes for each group. The trends indicate that both beginners and intermediate players adapt with increasing task demands, in this case speeding up their response and decision times. However, based on the results of our regression models, intermediate players can change their behavior to match the skills of experts, while beginners are unable to adapt to the extent required to close their gap with intermediate players. This is revealed by the large number of factors that lose discriminatory power between models 3 and 4 (comparing intermediates and experts) as opposed to only one factor from model 1 becoming irrelevant in model 2 (comparing beginners and intermediates). Combining all this information leads us to



(a) *Response Latency*



(b) *Decision Latency*

Fig. 15. Changes in response latency (top) and decision latency (bottom) with game level. Means and SDs of response (a) and decision (b) latencies (ms) of three-player categories across game difficulty levels (Y-labels use the feature names described in Appendix A). Trends in mean values show adaptive behavior for all players. For detailed analyses, see Section 5.2.

conclude that intermediate players are capable of playing with near expert skills when forced to, but it is not their default behavior.

Patterns of individual skills also reveal interesting trends. Planning-efficiency, zoid-control, and pile-uniformity remain significant discriminators for our first three models, perhaps because expertise in each of these skills vary over a broad spectrum. Pile-management does not contribute to group differences at level 0 (Models 1 and 3), between any of the player groups (see Table 4). Implying, when players are not pressed for time, anyone can build a clean pile (without gaps or holes). At level 2 (Model 2), however, pile-management becomes significant, likely because the increased time pressure is enough to push beginners into survival mode that results in poor pile states. The time pressure at this level also seems to be enough to disrupt the intermediate player's ability to perform higher line clears compared to beginners (indicated by the change in significance of factor 5 from Model 1 to 2).

Pile uniformity remains a significant factor across all models, possibly because it is a difficult skill to master, but players start using it even at very early stages of expertise. The ideal configuration for the top of a pile is somewhere between a very jagged and perfectly smooth pile, with slots that accommodate various zoid types without degrading the pile configuration. Finally, rotation corrections only seem to be useful when comparing intermediate players to experts. It is currently difficult to reach any conclusions about this observation, as the distributions of this factor at level 0 seem to converge with increasing expertise (see Appendix C for plot).

These changes in the significance of factors across the four models are also reflected in the model plots shown in Appendix C, in which they are manifested as differences in the average factor values among player categories. (NB, the Appendix C plots represent factor values for games played by the top three players across each of our three-player categories.)

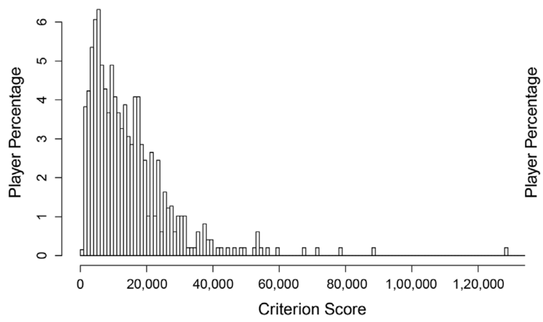
6. Comparing expertise levels across game levels: A factor distribution approach

For our logistic regression models in Section 5, the expertise level of each player was used as the outcome variable. In contrast, for Section 6, we use our linear models to determine factors that contribute to differences among players at the same expertise level.

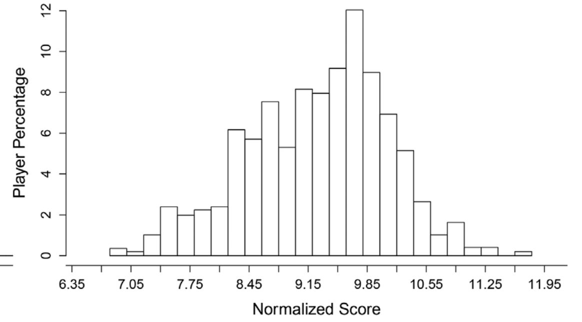
6.1. Criterion scores

Up to now, we used expertise level to differentiate among groups. However, we now examine differences among players within each expertise level. To do so, we return to the *criterion score* measure favored by Lindstedt and Gray (2019) and compute the mean of each player's score on their four highest scoring games.

This decision raises a new difficulty as scores in Tetris are skewed from level to level. That is, the distribution of games by game score is not normal, but is skewed due to the occurrence of many short, low-scoring games and fewer long, high-scoring games. This skewness is clearly apparent in Figure 16(a).



(a) *Non-Normalized Criterion Scores*



(b) *Normalized Criterion Scores*

Fig. 16. Distribution of criterion scores (a) and normalized (natural log) criterion scores (b) for all players.

We solve our problem and normalize the game score distribution by taking the natural log of data plotted in Figure 16(a) that produces the relatively normal curve shown in Figure 16(b). This transformation normalizes criterion scores and makes it a reasonable outcome variable for our linear regression model.

6.2. Linear regression model

As per our logistic regression models, the six retained factors were fed as predictor variables to these models as well. In this case, however, to ensure that the models were able to distinguish between individual players (within each group), criterion scores were used as outcome variables for model fit (since groups are defined on expertise levels). Stepwise model selection using AIC was then used to determine the best set of factors for each model. The models were used to evaluate which factors are necessary to *differentiate between players who belong to the same expertise level*. To this end, the following models were developed:

- Model 1: Fit to the data for *all players at level 0* of gameplay.
- Model 2: Fit to the data for *expert players at level 0* of gameplay.
- Model 3: Fit to the data for *expert players at level 8* of gameplay (last level of stable gameplay for expert players, on an average).
- Model 4: Fit to the data for *intermediate players at level 0* of gameplay.
- Model 5: Fit to the data for *intermediate players at level 5* of gameplay (last level of stable gameplay for intermediate players, on an average).
- Model 6: Fit to the data for *beginner players at level 0* of gameplay.
- Model 7: Fit to the data for *beginner players at level 2* of gameplay (last level of stable gameplay for beginner players, on an average).

6.3. Results

The goodness-of-fit for each model, after going through the model selection process (AIC based), is reported as adjusted R^2. The results for the model fits are presented in Table 5 (the Normal Q-Q plots for the fitted models are available in Appendix D).

Table 5 shows distinct behaviors within player groups. The R^2 values of model fits are the lowest for intermediate players (Models 4 and 5) perhaps because there is a large amount of variance in how intermediate players use the limited number of skills that do capture the variation in this group. Also, the right skew observed earlier (Figure 16) could not be completely resolved for intermediate and expert populations. This is indicated by the deviations of data points from the reference line, seen toward the right ends of the Q-Q plots in Appendix D, for models of intermediate and expert populations. While the model fits for the beginner population (Models 6 and 7) are higher than our intermediate population, expertise for these players varies across a broader range of skills. Also, the R^2 values for the expert models (Model 2 and Model 3) are the highest of the three-player categories, which could imply a convergence of skills for expert players.

Interestingly, at higher game levels, model fits get worse for expert and intermediate players, while the reverse is true for beginners. We speculate that beginners are likely to start off by exploring many different strategies (which might explain the poor fit for Model 6) but when the game becomes difficult for them (at level 2, Model 7), their strategies converge (higher R^2). As the intermediate and expert players have already explored and discarded many possible strategies, the converse is true for them. Indeed, perhaps, "strategy" is too grand a word to be applied to what the beginners are doing. Perhaps, a more fitting phrase would be something like *exploration of the state space* that, of course, is a very basic strategy. Indeed, the tendency of beginners in a complex task to explore various options to see what they do has been recently noted by Rahman and Gray (2020) as well as Anderson, Bettsa, Bothella, and Lebiere (2021).

Rotation-correction (Factor 6) is not a useful discriminator for any of our groups, except for beginners at level 0 (likely experimenting with new strategies for rotation). In contrast, planning-efficiency (Factors 1) and min-line-clears (Factor 5) remain relevant predictors in most models, suggesting a greater variability in the development of these skills at all three stages of expertise. Zoid-control and pile-uniformity (Factors 3 and 4) seem to be mostly useful for determining skill variation among beginners (Models 6 and 7).

Finally, the effects of time pressure are clearly observed in case of experts as pile-management and minimum-line-clears (Factors 2 and 5) no longer account for any significant variation when the experts are trying to survive at level 8 (Model 3) compared to level 0 (Model 2). Such effects are least apparent for beginners as most factors remain significant at both game levels 0 and 2 (Models 6 and 7), with the exception of pile-management. The pile-management exception is interesting since, under time pressure, it becomes a priority for beginners at level 2 but is not a priority in level 0. While the reverse is true for experts to whom it loses priority at higher levels. Perhaps, this reversal implies that basic pile management skills of experts are "good enough" for survival. Perhaps, when time permits them, experts attempt these highly skilled moves but, when pressed for time, revert back to their

Table 5
Information about model fits, coefficients, and significance of each factor corresponding to all linear regression models for each combination of player expertise and game levels. Higher R^2 value indicates a better fit

Model Information					Factor Information					
Model	F-Statistics	Signif.	R^2	Expertise Lvl	Factor 1 plann-effic	Factor 2 pile-mmt	Factor 3 zoid-cntrl	Factor 4 pile-unif	Factor 5 min-l-clears	Factor 6 rot-crrctns
Model 1	$F(5,1956) = 306.1$	***	0.438	L0 All	-0.075^{***}	-0.044^{***}	-0.064^{***}	0.018^{**}	-0.038^{***}	–
Model 2	$F(3,80) = 60.23$	***	0.682	L0 Expert	-0.051^{***}	0.039^{**}	–	–	-0.073^{***}	–
Model 3	F(3,69) = 24.4	***	0.494	L8 Expert	-0.164^{***}	–	0.044^{*}	–	−0.033	–
Model 4	$F(3,344) = 32.63$	***	0.215	L0 Intermed	–	−0.010	–	0.018^{**}	-0.041^{***}	–
Model 5	$F(4,200) = 6.368$	***	0.095	L5 Intermed	-0.022^{***}	-0.011^{*}	–	0.017^{*}	−0.014	–
Model 6	$F(5,242) = 19.45$	***	0.272	L0 Beg	-0.010^{*}	–	-0.022^{***}	0.015^{**}	-0.039^{***}	0.012^{*}
Model 7	$F(5,150) = 26.05$	***	0.447	L2 Beg	-0.012^{**}	-0.038^{***}	-0.015^{*}	0.029^{***}	-0.054^{***}	–

Significance codes: $p < 0.001$ ***; $p < .01$ **; $p < .05$ *; $p < .1$ "."; $p < 1$ "N"
Factors: F1: planning efficiency; F2: pile management; F3: zoid control; F4: pile uniformity; F5: min-line-clears; F6: rotation corrections

basic skills for quick decisions. In contrast, for beginners, surviving under pressure requires them to step up their pile-management skills.

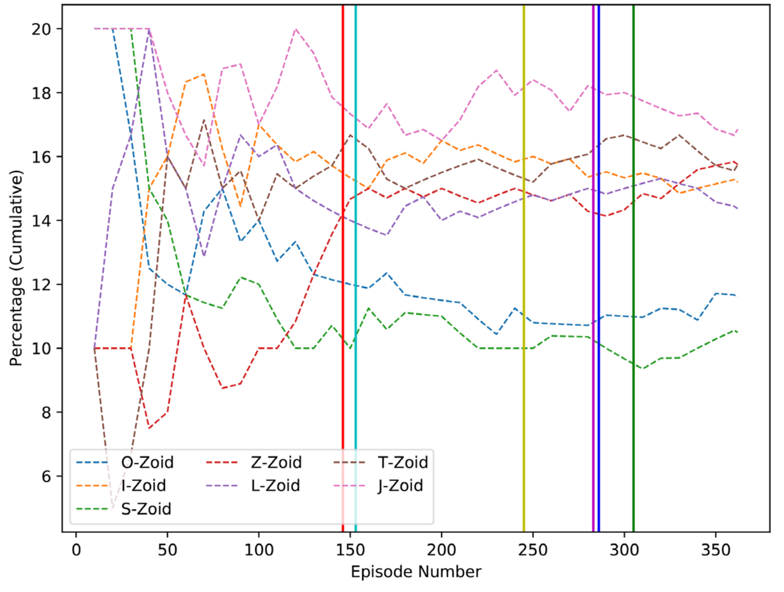
7. Randomness does not seem random: Analyzing random seeds

Each game has a random seed that dictates the sequence of zoids for that game. Figure 17 shows the distribution (cumulative percentage) of various zoid types (by episode) for seeds 111 and 666. The use of the cumulative percentage allows us to demonstrate two properties of the RNG. First, over the short term, say between the first zoid in a game and the 50th, one zoid type can be very frequent, while another can be very infrequent. Second, over the long term, for example, episode 300, as in Figure 17(b) for seed 666, the cumulative frequency each zoid type settles into what we would expect; namely, each appears in approximately 15% of the episodes by episode 300. However, in other cases, as for seed 111, the cumulative variations in individual zoids by trial 350 are still unevenly ranging from approximately 17% for the J-zoid to about 11% for the S-Zoid.

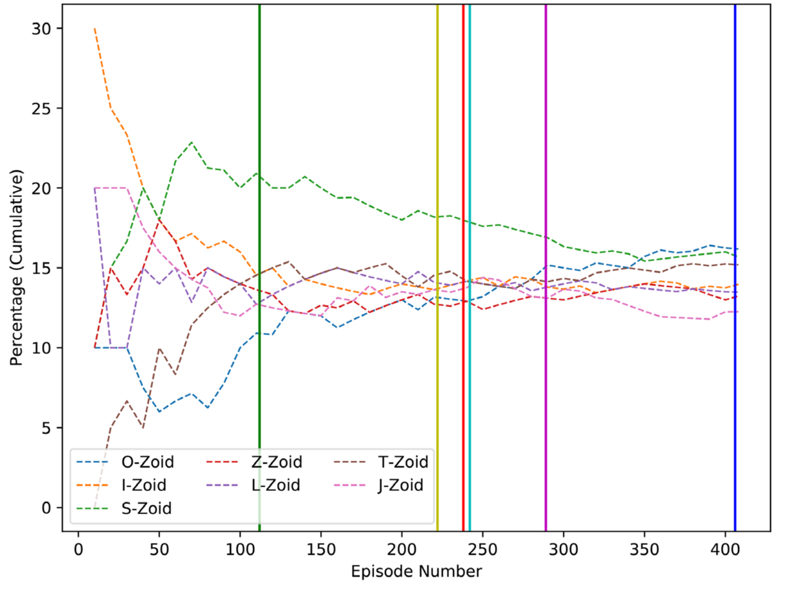
The colored vertical lines in the two plots in Figure 17 represent the episodes at which each of our top six players (based on criterion score) die. The color-coding remains constant across the two plots; for example, the reddish-orange vertical line in both plots represents the same player. A careful observer might notice that these vertical lines tend to cluster around certain episodes. A likely explanation for this is, for each seed, zoid distributions preceding certain episodes lead to challenging game conditions that sometimes overwhelm even our best players to the point of failure. Not much can be done by players if a key piece simply does not come. Indeed, during CTWC tournaments, “waiting for an I-beam,” sometimes adds to the drama of the game, so much so, that the organizers keep count, in real time, of the number of “other” zoids played since the last I-beam dropped. It is not unusual for this count to go into the 20’s (i.e., more than 20 other than I-beam zoids drop) before a new I-beam finally appears.

Comparing relevance of each of our six factors from Table 5 to games created by different seeds reveals the differences in skills needed to survive each game. Linear regression models were fit to level 0 gameplay data for all players, as before. However, in this case, each model represents the best fit for data corresponding to a specific seed. The step-wise model selector (based on AIC) was used to select the best set of factors for each model. The results for the model fitting process are presented in Table 6. Only the first six seeds were analyzed, since sufficient data were not available from our student experts for games beyond the sixth. The following models were trained:

- Model 1: Fit to the data (corresponding to seed 111) for all players at level 0 of gameplay.
- Model 2: Fit to the data (corresponding to seed 222) for all players at level 0 of gameplay.
- Model 3: Fit to the data (corresponding to seed 333) for all players at level 0 of gameplay.



(a) *Seed 111: Cumulative Distribution of Zoids across episodes*



(b) *Seed 666: Cumulative Distribution of Zoids across episodes*

Fig. 17. Distribution (cumulative percentage) of the seven Tetris zoids for seeds 111 (top) and 666 (bottom). The value of a zoid-type at any point represents the percentage of episodes in which the zoid is encountered up to that episode. For example, the value for the z-zoid at 50 episodes (for seed 111) is 8%, which means that the z-zoid showed up in 4 of the first 50 episodes. (Note that the vertical lines running through the plots each represent the episode at which one of our top six players [based on criterion score] died.)

Table 6
Information about model fits, coefficients, and significance of each factor corresponding to linear regression models for each seed. Higher R-squared (R-Sq.) Value indicates a better fit

Model Information				Factor Information					
Model	F-stat	Signif.	R^2	Factor 1 plann-effic	Factor 2 pile-mmt	Factor 3 zoid-cntrl	Factor 4 pile-unif	Factor 5 min-l-clears	Factor 6 rot-crrctns
Model 1	$F(5,220) = 16.57$	***	0.257	−0.034***	−0.107***	−0.029*	0.046*	−0.050*	–
Model 2	$F(4,256) = 50.58$	***	0.433	−0.067***	−0.054***	−0.065***	–	−0.045**	–
Model 3	$F(4,196) = 51.49$	***	0.502	−0.086***	–	−0.090***	–	−0.058**	−0.045
Model 4	$F(5,303) = 70.93$	***	0.532	−0.084***	−0.036**	−0.053***	–	−0.039**	−0.017 N
Model 5	$F(3,132) = 68.15$	***	0.599	−0.010***	–	−0.102***	–	−0.037	–
Model 6	$F(5,226) = 56.07$	***	0.544	−0.082***	0.029	−0.091***	–	−0.056***	−0.038

Significance codes: $p < 0.001$ "***"; $p < .01$ "**"; $p < .05$ "*"; $p < .1$ "."; $p < 1$ "N"

- Model 4: Fit to the data (corresponding to seed 444) for all players at level 0 of gameplay.
- Model 5: Fit to the data (corresponding to seed 555) for all players at level 0 of gameplay.
- Model 6: Fit to the data (corresponding to seed 666) for all players at level 0 of gameplay.

As per Table 6, the three factors that remain relevant across all seeds (with the exception of model 5) are planning-efficiency, zoid-control, and minimum-line-clears. The significance of our other factors change with seeds. This suggests that certain seeds create gameplay conditions that force players to rely more heavily on one set of skills or another.

A deeper analysis of the zoid distribution for each seed and the effect it has on player decisions is likely to generate interesting conclusions and raise more interesting questions about human behavior. Although such an analysis is beyond the scope of the current study, it might be addressed in future research.

8. Discussion

In Section 2.2, we defined the eight events of Tetris with which our non-tournament, student players grappled. Knowledge of these events guided our considerations in identifying and naming the features found in our three analyses. In this section, we briefly review and summarize the major findings for each of these analyses.

8.1. Establishing the basis of expertise differences: Feature extraction (Section 4)

In Section 4, we perform principal component analysis (PCA) on 35 features of Tetris gameplay to identify various dimensions (the components of a PCA process) of player skill. The top six dimensions were retained for the analyses. After rotation, we named the new dimensions for qualities suggested by their dominant features; namely, planning-efficiency, pile-management, zoid-control, pile-uniformity, minimum-line-clears, and rotation-corrections.

Player groups were defined on expertise level (EL), a grade awarded to each player by averaging the last level of gameplay for their top four games. Three player groups were defined for the study: beginners (EL 3), intermediates (EL 6) and experts (>= EL 9). Players belonging to other expertise levels were left out so as to ensure gaps in expertise between adjacent groups. We believe that these gaps have helped to emphasize the differences in skill between our three analyzed groups of players (EL 3, EL 6, and EL >= 9).

The six skills established by our feature extraction were used as metrics to distinguish between groups (Section 5) and among individuals within groups (Section 6).

8.2. Finding important differences among player groups: Logistic regression models (Section 5)

Based on the six skills established in Section 4, Section 5 focused on group differences at various levels of gameplay. Logistic regression (logit) models were trained to perform binary classification on adjacent pairs of player groups (beginners vs. intermediate and intermediate vs. expert). Four logit models were trained, two for the beginner and intermediate populations (gameplay at level 0 [Model 1] and level 2 [Model 2]), and two more for the intermediate and expert populations (gameplay at level 0 [Model 3] and level 5 [Model 4]).

The results yield significant skill differences between the beginner and intermediate populations at level 0 and at level 2 and also between the intermediate and experts populations at level 0 and at level 5. Our findings suggest that all players adapt with the changing game demands of the higher difficulty levels. However, unlike intermediate players, who, when forced, are able to perform nearly at expert levels of skill, beginners are unable to close their gap with intermediate players through adaptation. The implication we draw is that the leap in skill needed for players who routinely survive level 3 to be able to survive level 6 is far greater than the leap needed for those who survive level 6 to also survive at level 9.

Certain skills, when investigated independently, also lead us to interesting conclusions. Pile-uniformity remains a significant discriminator across all models, which could mean that it is a difficult skill to master. Learning the optimal amount of pile jaggedness takes time and practice, and even if the players do manage to figure that out, as additional zoids keep raining down, they need to have high levels of foresight and planning to incorporate these zoids into their pile configuration.

On the other side of the spectrum, pile-management does not seem to be a skill that helps differentiate any of our player groups at level 0. Implying that, given enough time, even our worst players are able to build clean piles.

Finally, rotation-corrections only seem to be useful when comparing intermediate players to experts at level 0. Without more information, we can only speculate that at least some of our players begin experimenting with bidirectional rotation (see Figure 5 and Figure 6). This flirting with rotation may be important as the players who participate at CTWC demonstrate advanced execution of rotation skills. Perhaps, some of our student players begin to acquire rotation skill by correcting over (i.e., extra) rotations.

8.3. Looking for skill differences within each expertise group: Factor distribution (Section 6)

After identifying skill differences between groups, we turned our attention to within-group variations in skill. For this analysis, we trained seven linear regression models, on gameplay data for: (a) All players at level 0, (b) One for each of the three-player groups at level 0, (c) Beginners at level 2, (d) Intermediate players at level 5, and (e) Experts at level 8. The models were trained to predict the criterion score of a player based on the six factor values.

In brief, we found that beginners have the widest variation in skill; five of the six identified skills remain significant across both levels (level 0 and level 2) of gameplay; and model fits for beginners improve at higher levels of gameplay, whereas they get worse for the other two groups. These findings signal a tendency towards exploratory behavior among players in this category. While beginners vary across a broad spectrum of skills, intermediate players vary with greater magnitudes in each skill type but across fewer dimensions. This variation manifests as poor model-fits for intermediate players.

Experts seem to have achieved pile-management skills where their base skill is enough for them to survive even under pressure. For beginner and intermediate players, these skills start to diverge with growing time pressure. Experts also likely use the *extra time* they have at early levels of gameplay, to explore more advanced pile-management strategies.

Finally, an inspection of individual skills reveals that planning-efficiency and minimum-line-clears remain significant in almost all cases implying that, compared to the other skills, these skills exist on a broader spectrum of development across various stages of expertise.

8.4. Controlled randomness and its effect on player performance (Section 7)

In the final section of analysis, we introduced the concept of random seed as a control factor for the dynamic nature of the Tetris environment. All players who use the same random seeds are exposed to the same set of dynamic test environments in the same serial order, a form of pseudorandomization. Random seeds are responsible for randomization of the sequence of zoids that players receive over the course of a game. Advanced players who plan their actions based on the possibility of getting specific zoids to execute special moves are particularly affected by extreme distributions (too much or too little availability of a zoid type).

Linear regression models were trained on data corresponding to various seeds. The models were trained to predict the criterion score for each player (based on their six factor values) for gameplay at level 0 across all players. The results reveal that each seed creates unique game conditions where some skills may become more important than others. Future research can dive deeper into the data to identify how seeds affect player performance, and how specific seeds compare with each other.

8.5. Converging analyses: Comparing player groups and across game levels

Combining results from our analyses in Sections 5 and 6 help us understand the distributions of the six skills, both across and within groups. Planning-efficiency, pile-uniformity, and minimum-line-clears are the three most reliable predictors of expertise for players within the same group. Planning-efficiency shows similar predictive power when comparing players from different groups. In contrast, minimum-line-clears loses discriminatory power for between group comparisons at higher levels of gameplay.

Pile-uniformity remains a significant predictor of performance across the board for both between-group and within-group comparisons, with the exception of within-group differences for expert players. This suggests that expert players have already converged on this skill, while both beginner and intermediate players continue to refine the skill as they get better at playing the game. Zoid-control remains a significant factor for our between-group comparisons, but fails to distinguish between players within the same group.

Finally, based on the trends for the pile-management factor in both analyses, experts seem to have mastered the skill to the point that they show optimal behavior in the skill even when they are close to death, at level 8. Also, they might be utilizing the extra time they have at lower levels for experimenting with more advanced strategies.

9. Conclusions: Revisiting the anticipatory behavior of expert performers

We began our paper with a quote from Thorndike (1913) in which he recommends that those interested in studying the limits "of efficiency" of mental functions should examine, "those occupations of work or play" in which "excellence … is sought with great zeal and intelligence." Of course, our candidate for that type of study was the computer game *Tetris*.

Overcoming plateaus, dips, and leaps (Gray & Lindstedt, 2017) in human performance is more difficult for Tetris than in many other tasks as all Tetris games end in failure and all necessitate restarting the game from its beginning. That statement is true of every Tetris player whether she is a Jonas Neubauer (the seven times *Classic Tetris World Champion)* or a first time player.

If Tetris were "merely a twitch" game, then movements would occur in response to a change in game state. Although this seems like a reasonable statement, the reality is that such a system would be too slow to interact with a dynamic world. For Blättler et al. (2011) French Air Force pilots (discussed in Section 2.1, page 7), in a situation where the world literally stops moving, the experts make "forward errors"—the pilots indicated a shift in target location *forward* to where it would have appeared if the world had continued to move. These expert pilots were trained to look (and presumably aim and fire) at where their targets would be in a few milliseconds, not where they are now. This type of predictive processing (Hommel, 1998, 2019) and EPCog (Baldwin & Kosie, 2021; Butz et al., 2021; Cooper, 2021; Kuperberg, 2021) seems emergent in our data. We also drew on a variety of studies that compared expert with novice performance in team games such as Rugby, Tennis, Beach Volleyball, and Basketball.

Without having to take a position on how these differences arise, it is clear that each of our three classes of Tetris players (beginners, intermediates, and expert) perform the *same* tasks differently from the others. For example, our logistic regression models (see Section 5, Table 4) show differences in the factor information for models trained to distinguish between beginner and intermediary players at level 0, and those trained to distinguish between the play of the same set of players at level 2. Likewise, our linear regression models from Section 6 (see Table 5) show us that, at level 0, our Experts, Intermediates, and Beginners not only have different weights on some of the same features but also have different combinations of features (e.g., compare Model 2 with Model 6 or Model 4 with both Model 2 and Model 6). These are complex differences and, without having to understand (let alone to explain) how these differences arise, we are happy to conclude that expert performance in Tetris is not simply a matter of innate twitch speed or of any other single factor, but arises, somehow, from a complex combination of factors that are caught like ancient flies in the amber of our factor analyses.

10. Summary: League-stepping habits

> Now, the ability to take league steps in receiving telegraphic messages, in reading, in addition, in mathematical reasoning and in many other fields, plainly depends upon the acquisition of league-stepping habits. No possible proficiency and rapidity in elementary processes will serve. (Bryan & Harter, 1899, p. 375)

Does success as a Tetris player depend on a person's *twitch speed*? That is, is it simply a matter of how fast they can move their fingers? Although our closing quote is over 121 years old, the myth of superior performance as due to innate individual differences in speed, in ignorance of the acquisition of league-stepping habits, stubbornly persists.

10.1. The data

In this paper, we have taken up the story of 492 college students, all of whom had some prior familiarity with the game of Tetris. As they are young and, generally, in good physical condition, if superior *twitch speed* were the secret to Tetris success then surely this would be a successful group. However, the key to their play lies not in simple response time but in the way in which the various components of EPCog load onto the 35 factors listed in our Appendix A (and detailed in Appendix Table B, PCA Loadings).

Even our beginner players are doing things other than responding fast. They quickly realize that while clearing one line scores points, points escalate if two or more lines are simultaneously cleared. Hence, rather than just clearing the bottom line as soon as it fills, they are beginning to build solid, multiple line, walls with a vertical gap someplace that can be plugged to remove two, three, or four lines at once (depending on their success at *wall building*). Our intermediates and experts start filling in their walls in a fashion that allows them to handle awkward zoids. For example, once a wall starts to rise, deciding where to place a square zoid becomes not simply an immediate decision but an event with consequences that must be handled correctly, so the square zoid fits in well with the rest of the pile right now but, also, so its placement can prepare the board to best fit the next zoid as well as the next-next zoid, and so on. If Tetris were a twitch game these sorts of factors would not be considerations. Indeed, there would be little difference in strategy, planning, or skill acquisition between Tetris and extended training on, for example, the Hick-Hyman task (Hick, 1952; Hyman, 1953).

10.2. Tasks, tools, and techniques

As suggested by Figure 8, in common with many human tasks, Tetris can be considered as a task or set of subtasks, with tools that are used for performing the tasks, and different techniques available (or discoverable) for using those tools. Within an episode (see, Event 1, Table 1), within each step of each zoid's fall, there can be an instance of one of the four types of player initiated movements. Players can rotate the zoid clockwise or counterclockwise (rotations), move it left or right (translations), force an early drop, or fill in a gap in the row (which, if done with some forethought and planning, may briefly creates one to four solid lines of filled-in rows and then dissolve). These first three movements usually occur in combination with each other, and as the last event (filling in one to four rows) stops the zoid from falling, it always ends the episode.

The techniques our players develop for using their tools were revealed by the various analyses performed in Sections 4– 7. Those techniques have a direct correspondence to the tools available in Tetris. Interestingly, techniques continue to change, develop, and be invented even (or perhaps, especially) at the higher levels of Tetris play. Indeed, during the 2020 CTWC competition, the champions of the Tetris world were divided by their choice of one of two types of techniques for bringing a new zoid onto the board; namely, *Hypertapping* versus *DAS*. Each of these is a maneuver favored by different groups of tournament level players. We are safe in saying that none of our 492 players used either of these two techniques.

10.3. What's next? Above beginners and beyond tetris

Why do players die? Specifically, in Tetris, players play until something happens and they lose control of their boards. An orderly board enables growth and control. But at some point, whether due to player slips, "bad luck" with their random seed, or relentlessly increasing drop speed, all players die. In this, as in prior reports, we avoided any attempt to analyze the death level and only focused on levels which the player completed. However, the question of how they lose control and how or whether our differently skilled players; that is, our experts, intermediates, and beginners lose control in different ways is an outstanding question that we plan to pursue.

What about the good players? The really good players? The ones who make it into the annual CTWC?? As unbelievable as this may sound to all 492 of our student players, Tetris as played at CTWC starts at level 18 where a zoid will drop from top-to-bottom in 1 s and then it gets fast—at level 19, the same zoid will drop from top-to-bottom in 2/3's of a second (see Table 2). Although we mention these incredible CTWC players often in this report, until recently, we have been able to collect only modest amounts of data from them and we have never amassed much data, from any of them, beyond level 18. But, that was then and this is now, and perhaps, there really is always a good side to everything, even the pandemic. The 2020 CTWC was played remotely and streamed over Twitch (https://www.twitch.tv/directory/game/Tetris). For a new research effort, we are busily translating Tetris play files collected by the CTWC organizers into file structures similar to those used in the raw data for this report.

10.4. Concluding thoughts: The control of anticipated action

There is much going on in our study of the simple game of Tetris and many words have been spent in this paper explaining details of the game's events and of the analyses that take us beyond events into the mixture of factors that change across levels of play as well as across the acquisition of expertise. Behaviorally, it is probably not surprising that most (if not all) Tetris players reach points in the game where they become very quiet with their full attention locked on the screen of this "simple twitch game."

It is, indeed, difficult to describe complex tasks like Tetris in terms of what we observe in the psychology laboratory where simple tasks, which are not all that dissimilar from classic tasks such as Hick–Hyman (Hick, 1952; Hyman, 1953), tend to rule the 1-h per subject experimental psychology lab. Such tasks have provided the basis of our modern understanding of dynamic decision-making and perceptual-motor behavior. Time and again, going back into the lab to work out what we think we see happening in more complex tasks has proven to be a vital strategy. Indeed, our work has directly benefited from the thoughts, theories, and conclusions of studies such as those by Henry and Rogers (1960), Hommel (1998), Kunde, Koch, and Hoffmann (2004), Zacks and Swallow (2007), Anderson and Fincham (2014), Hommel (2019), Butz et al. (2021), Cooper (2021), Kuperberg (2021), and many others. Similarly, our research strategy profits from advice given by the late Allen Newell in his seminal paper, "You can't play 20 questions with nature and win" (1973). Three of his several suggestions we have taken to heart are, (a) to know the method your subject is using to perform the experimental task, (b) never average over methods, and (c) to accept a single complex task and do all of it.

Acknowledgments

The work was supported, in part, by grant N00014-16-1-2796 to Wayne D. Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer. Wayne Gray conceived and initiated this paper. Sounak Banerjee is primarily responsible for the results and analyses sections. Both authors contributed to the Discussion, Summary, and Conclusions sections. We dedicate this paper to the memory of Jonas Neubauer, the 7 time *Classic Tetris World Champion* (CTWC). Jonas died from a sudden medical emergency on January 4th, 2021. Across at least 5 years of CTWC, Jonas supported our research by playing Tetris on our machines during the qualifying rounds which precede the annual tournament. This was a contribution to science as it enabled us to collect his point-of-gaze and response time data with millisecond accuracy for analyses similar to those used in this paper. The data that Jonas, and many other CTWC Tetris Wizards, contributed will be the subject of future research papers. Finally, please note that data from Lindstedt and Gray's 2019 paper is archived at OSF as https://osf.io/78ebg/. The data on which this paper is based will be archived at OSF as a private archive pending planned future analyses of the dataset. The archive will be made available to the research community after publication of a planned set of future studies and/or within 3 years after the publication of this paper—whichever event comes sooner.

Notes

1 Auch zum Zögern muss man sich entschließen.
2 A famed guitar player.
3 The winner of the most matches won at Grand Slams by any male or female player in tennis history.
4 7-time Classic Tetris World Champion. Jonas also created a series of YouTube videos which demonstrate advanced Tetris techniques.

5 Harry Hong was the first player to document "maxing-out" Tetris by scoring 999,999 points.

6 "Hypertapping" is a technique for the repetitive pressing of keys that is much faster than the normal technique of keypressing. It is not used by players in our sample of 492 student players but is used very frequently by tournament-level players.

7 Although this statement is true of NES Tetris, as we discuss later, in our *Meta-T* version of Tetris, we used the same set of random number seeds for each of our 492 players.

8 Note that the statement, "with replacement," is true of the classic NES Tetris but is not true for most other Tetris variants.

9 https://www.youtube.com/watch?v=c4l6g8_rei0

10 As we discuss later, in our sample of 492 student players, even our very best student player would be unable to qualify for CTWC.

11 Re-afferent: sensory signals that occur as a result of the movement of the sensory organ. This re-afferent signal of motion is compared by the brain to that which would be expected as a result of the intended movement, and adjustments are made as necessary. https://dictionary.apa.org/reafference

12 Italics added.

13 Subjectively, "reactivation" may be what happens when we "think of," "imagine," or "simulate" the outcome of an action.

14 Also, for anyone who has been lead to believe that mental rotation plays a role in Tetris, please see Destefano et al. (2011).

15 I-beams are very important in CTWC play as most players who qualify spend most of their time building solid walls of zoids that are at least four zoids deep and then waiting until an I-beam comes along so that they can plug the column to clear all four rows at once; thereby scoring a Tetris (and gaining eight times as many points as they would by clearing one line four times).

16 Also see the right-side of Fig. 1.

17 At the time this study was run, 506 zoids constituted the longest Tetris game played by any undergraduate in our lab. Hence, we took this to represent a strong student player.

18 The exit survey was in support of undergraduate projects and is not further discussed in this paper.

19 We recognize that these *death-level* games could be interesting in their own right; however, we reserve their analysis for a future paper.

20 Each of these two players remained in the lab when the experimenter restarted the game. Hence, our curated SQL files for these data include this as one session that was treated the same as the other of 492 player sessions used for these analyses.

21 (An eigenvalue of 1 indicates that a factor explains exactly as much variance as any one of the original 35 features.)

22 For more information on Logistic Regression models, please see: https://simple.wikipedia.org/wiki/Logistic_regression

23 This statement comes with one caveat; namely, that in tournament competition, Tetris play begins at level 18 (see Table 2). However, the spirit of this comment is preserved

as when the CTWC games are reset, they are reset to the beginning level for tournament play; which is, of course, level 18.

24 See Macdonald (n.d.) excellent book on all known Tetris moves and strategies.

25 All we will say here is that DAS is conceptually similar to how the repeat key works on a keyboard.

References

Anderson, J. R., Bettsa, S., Bothella, D., & Lebiere, C. (2021). Discovering skill. *Cognitive Psychology*, *129*, 1–23, https://doi.org/10.1016/j.cogpsych.2021.101410

Anderson, J. R., & Fincham, J. M. (2014). Discovering the sequential structure of thought. *Cognitive Science*, *38*(2), 322–352. https://doi.org/10.1111/cogs.12068

Anderson, D. N., Gottwald, V. M., & Lawrence, G. P. (2019). Representational momentum in the expertise context: Support for the Theory of Event Coding as an explanation for action anticipation. *Frontiers in Psychology*, *10*, 1–9. https://doi.org/10.3389/fpsyg.2019.01838

Baldwin, D. A., & Kosie, J. E. (2021). How does the mind render streaming experience as events? *Topics in Cognitive Science*, *13*, 79–105. https://doi.org/10.1111/tops.12502

Blättler, C., Ferrari, V., Didierjean, A., & Marmèche, E. (2011). Representational momentum in aviation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1569–1577. https://doi.org/10.1037/a0023512

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, *6*(4), 345–375. https://doi.org/10.1037/h0073117

Butz, M. V., Achimova, A., Bilkey, D., & Knott, A. (2021). Event-predictive cognition: A root for conceptual human thought. *Topics in Cognitive Science*, *12*, 10–24.

Cañal-Bruland, R., Mooren, M., & Savelsbergh, G. J. P. (2011). Differentiating experts' anticipatory skills in beach volleyball. *Research Quarterly for Exercise and Sport*, *82*(4), 667–674. https://doi.org/10.1080/02701367.2011.10599803

Cooper, R. P. (2021). Action production and event perception as routine sequential behaviors. *Topics in Cognitive Science*, *13*, 63–78. https://doi.org/10.1111/tops.12462

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7), 1–9. ISSN 1531-7714

Cutting, J. E. (1981). Six tenets for event perception. *Cognition*, *10*(1), 71–78. https://doi.org/10.1016/0010-0277(81)90027-5

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

Didierjean, A., Ferrari, V., & Blättler, C. (2014). Role of knowledge in motion extrapolation: The relevance of an approach contrasting experts and novices. *Psychology of Learning and Motivation*, *61*, 215–235. https://doi.org/10.1016/B978-0-12-800283-4.00006-X

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*(4448), 1181–1182. https://doi.org/10.1126/science.7375930

Gray, W. D., Hope, R. M., Lindstedt, J. K., & Destefano, M. (2014). Elements of extreme expertise. *Cognitive Processing*, *15*(1), S7–S7.

Gray, W. D., Hope, R. M., Lindstedt, J. K., & Sangster, M.-D. (2015a). Expert and novice differences in gaze transitions during play of a dynamic video game. *Journal of Eye Movement Research*, *8*(4), 273. https://doi.org/10.16910/jemr.8.4.1.

Gray, W. D., Hope, R. M., Lindstedt, J. K. & Sangster, M.-D. (2015b). Gaze transitions as clues to expert-novice strategy differences in a dynamic video game. In ETVIS 2015 - First Workshop on Eye Tracking and Visualization.

Gray, W. D., & Lindstedt, J. K. (2017). Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance. *Cognitive Science*, *41*(7), 1838–1870. https://doi.org/10.1111/cogs.12412

Gray, W. D., Lindstedt, J. K., Sibert, C. L., Sangster, M.-D. D., Rahman, R., Denga, R., & Destefano, M. (2018). The essence of interaction in boundedly complex, dynamic task environments. In K. A. Gluck & J. E. Laird (Eds.), *Interactive task learning: Humans, robots, and agents acquiring new tasks through natural interactions*. *26* Cambridge, Massachusetts: Strungmann Forum Reports.147–165.

Henry, F. M., & Rogers, D. E. (1960). Increased response latency for complicated movements and a memory drum theory of neuromotor reaction. *Research Quarterly*, *31*(3), 448–458. https://doi.org/10.1080/10671188.1960.10762052

Hick, W. E. (1952). On the rate of gain of information. *The Quarterly Journal of Experimental Psychology*, *4*, 11–26. https://doi.org/10.1080/17470215208416600

Hommel, B. (1998). Perceiving one's own action–and what it leads to. In Jordan, J. S. (Ed.), *Systems theories and a priori aspects of perception*, (pp. 143–179). North-Holland.

Hommel, B. (2019). Theory of event coding (TEC) v2.0: Representing and controlling perception and action. *Attention, Perception, & Psychophysics*, *81*(7), 2139–2154. https://doi.org/10.3758/s13414-019-01779-4

Hommel, B., Musseler, J., Aschersleben, G., & Print, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/S0140525X01000103

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, *12*(5), 822–851. https://doi.org/10.3758/BF03196775

Hubbard, T. L. (2019). Momentum-like effects and the dynamics of perception, cognition, and action. *Attention, Perception, & Psychophysics*, *81*(7), 2155–2170. https://doi.org/10.3758/s13414-019-01770-z

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*(3), 188–196.

Jackson, J. E. (2005). Varimax rotation. In *Encyclopedia of biostatistics*. American Cancer Society. https://doi.org/10.1002/0470011815.b2a13091

Johansson, G., Hofsten, C. V. H., & Jansson, G. (1980). Event perception [PMID: 7362214]. *Annual Review of Psychology*, *31*(1), 27–63. https://doi.org/10.1146/annurev.ps.31.020180.000331

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

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

Krzywinski, M., & Altman, N. (2013). Points of significance: Error bars. *Nature Methods*, *10*(10), 921–922. https://doi.org/10.1038/nmeth.2659

Kunde, W., Koch, I., & Hoffmann, J. (2004). Anticipated action effects affect the selection, initiation, and execution of actions. *Quarterly Journal of Experimental Psychology: Section A*, *57*(1), 87–106.

Kuperberg, G. R. (2021). Tea with milk? A hierarchical generative framework of sequential event comprehension. *Topics in Cognitive Science*, *13*, 256–298. https://doi.org/10.1111/tops.12518

Lec, S. J. (1962). *Unfrisierte Gedanken (Unkempt thoughts)*. St. Martin's Press.

Lindstedt, J. K., & Gray, W. D. (2015). MetaT: Tetris as an experimental paradigm for cognitive skills research. *Behavior Research Methods*, *47*(4), 945–965. https://doi.org/10.3758/s13428-014-0547-y

Lindstedt, J. K., & Gray, W. D. (2019). Distinguishing experts from novices by the mind's hand and mind's eye. *Cognitive Psychology*, *109*, 1–25. https://doi.org/10.1016/j.cogpsych.2018.11.003

Loschky, L. C., Larson, A. M., Smith, T. J., & Magliano, J. P. (2020). The scene perception & event comprehension theory (spect) applied to visual narratives. *Topics in Cognitive Science*, *12*(1), 311–351. https://doi.org/10.1111/tops.12455

Macdonald, D. (April 2021). New NES Tetris technique: Faster than hypertapping!

Marutho, D., Hendra Handaka, S., Wijaya, E. & Muljono (2018). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In 2018 International Seminar on Application for Technology of Information and Communication (pp. 533–538). https://doi.org/10.1109/ISEMANTIC.2018.8549751

Murphy, C. P., Jackson, R. C., & Williams, A. M. (2018). The role of contextual information during skilled anticipation [cited By 14]. *Quarterly Journal of Experimental Psychology*, *71*(10), 2070–2087. https://doi.org/10.1177/1747021817739201

Newell, A. (1973). You can't play 20 questions with nature and win. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). Academic Press. https://doi.org/10.1184/R1/6612977.v1

Rahman, R., & Gray, W. D. (2020). SpotLight on dynamics of individual learning. *Topics in Cognitive Science*, *12*(3), 1–17. https://doi.org/10.1111/tops.12512

Rosenbaum, D. A. (2010). *Human motor control* (2nd ed. [VitalSource Bookshelf version]). Academic Press, Elsevier.

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, *136*(6), 943–974. https://doi.org/10.1037/a0020541

Sibert, C. L. (2015). Improving novice tetris players with feedback from AI model based tutors (Master's thesis). Rensselaer Polytechnic Institute, Troy, NY.

Sibert, C. L. (2019). Unpuzzling tetris: Exploring the mechanisms of expertise in a complex, dynamic task with simple machine learning models (Doctoral dissertation). Rensselaer Polytechnic Institute, Troy, NY.

Sibert, C. L., & Gray, W. D. (2018). The Tortoise and the Hare: Understanding the influence of sequence length and variability on decision making in skilled performance. *Computational Brain & Behavior*, *1*(3–4), 215–227. https://doi.org/10.1007/s42113-018-0014-4

Sibert, C. L. & Gray, W. D. (2020). The need for speed: Effects of human derived time constraints on performance and strategy in machine models of tetris. In International Conference on Cognitive Modeling.

Sibert, C. L., Gray, W. D., & Lindstedt, J. K. (2015). Tetris: Exploring human performance via cross entropy reinforcement learning models. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2188–2193).

Sibert, C. L., Gray, W. D., & Lindstedt, J. K. (2017). Interrogating feature learning models to discover insights into the development of human expertise in a real-time, dynamic decision-making task. *Topics in Cognitive Science*, *9*(2), 374–394. https://doi.org/10.1007/s42113-018-0014-4

Sibert, C. L., Lindstedt, J. K., & Gray, W. D. (2014). Tetris: Exploring human strategies via cross entropy reinforcement learning models. In: P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Sibert, C. L., Speicher, J., & Gray, W. D. (2019). Less is more: Additional information leads to lower performance in Tetris models. In *Proceedings of the International Conference on Cognitive Modeling* (pp. 1–7).

Smith, B. (2014). *Tricks of the classic NES tetris masters*. ixbrian.com.

Thorndike, E. L. (1913). *Educational psychology vol ii: The psychology of learning*. Teachers College, Columbia University.

Vicario, C., Makris, S., & Urgesi, C. (2017). Do experts see it in slow motion? Altered timing of action simulation uncovers domain-specific perceptual processing in expert athletes. [cited By 7]. *Psychological Research*, *81*(6), 1201–1212. https://doi.org/10.1007/s00426-016-0804-z

Wang, H., & Song, M. (2011). Ck means.1d.dp: Optimal k-means clustering in one dimension by dynamic programming. *The R journal*, *3*(2), 29–33.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis [Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists]. *Chemometrics and Intelligent Laboratory Systems*, *2*(1), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Wulf, G. (2013). Attentional focus and motor learning: A review of 15 years. *International Review of Sport and Exercise Psychology*, *6*(1), 77–104. https://doi.org/10.1080/1750984X.2012.723728

Zacks, J. M. (2001). Scaling up from atomic to complex events [cited By 1]. *Behavioral and Brain Sciences*, *24*(5), 909–910. https://doi.org/10.1017/s0140525x01510109

Zacks, J. M. (2020). Event perception and memory. *Annual Review of Psychology*, *71*(1), 165–191. https://doi.org/10.1146/annurev-psych-010419-051101

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*(6), 651–655. https://doi.org/10.1038/88486

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293.

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29

Gray and Banerjee show that the transition from novice to expert performance in complex dynamic tasks is not a smooth ascent. Rather, learning in such tasks involves phase shifts, where individuals acquire a range of skills along the way. Sets of exploratory factor analyses (EFA) provide detailed examinations of differences between players at various levels of expertise, whereas other sets of EFAs examine differences among players within the same level of expertise. Higher performance among players is consistently associated with various forms of anticipatory behavior.

Appendix A: Game features

Counts, percentages and mean values for features are calculated across all episodes corresponding to each level.

Basic Information:

`EpisodeCount`: Number of episodes corresponding to each level of gameplay.

Pile-Specific Information:

`CreatedOverhangs_Percent`: The percentage of episodes where new overhangs were created.

`ClearedOverhangs_Percent`: The number of episodes where existing overhangs were cleared expressed as a percentage of the number of episodes where new overhangs were created for the current level.

`CreatedWells_Percent`: The percentage of episodes where new wells were created.

`ClearedWells_Percent`: The number of episodes where existing wells were cleared expressed as a percentage of the number of episodes where new wells were created for the current level.

`WellDepth_mean`: The mean depth of all wells across all episodes in the level.

`<X>_DepthWells [Ex:3_DepthWells]`: The average count of 'X'-depth wells present in each episode across a level. 'X' can take values 2-4.

`Gt4_DepthWells`: The average count of wells with depths greater than 4 present in each episode for a level.

`CreatedPits_Percent`: The percentage of episodes where new pits were created.

`ClearedPits_Percent`: The number of episodes where existing pits were cleared expressed as a percentage of the number of episodes where new pits were created for the current level.

`PitSize_mean`: The average size of all pits across all episodes for a level.

`Spire_Percent`: Percentage of total episodes in which a central spire was present.

`MaxPileHeight`: The average height of the peak of the pile, averaged across all episodes in each level.

`RightWell_Score`: A custom score for the right well based on well depth and maximum pile-height at each episode, averaged across levels.

`PileScore`: A custom scoring function for the pile orderliness averaged across all episodes in each level. The score at each episode is derived from a combination of the depth of the rightmost well, pile peak height, shape of the top of the pile, and distribution of overhangs and pits.

Action-specific information:

`RespLatency_mean`: The mean time it took (milliseconds) across all episodes (in a level) to click the first button.

`DropDuration_mean`: The mean time (milliseconds) across all episodes (in a level) for how long the zoid was dropped by the player.

`DecisionLatency_mean`: The mean time it took (milliseconds) across all episodes (in a level) to place the zoid in its final orientation and horizontal position. Calculated by subtracting the response latency from the time of the final rotation/translation, for each episode.

`DecisionLatencyPercent_mean`: The mean value of decision latency expressed as a percentage of the total time available for each episode in a level.

`ExtraRotations_mean`: The mean number of extra rotations (varies by zoid-type; Ex: for static zoids any rotation is an extra rotation) made unidirectionally across all episodes (in a level) to get the zoid to its final orientation.

`ExtraRotations_nonzeroPercent`: The percentage of episodes that had extra rotations.

`DominantRotation_DirectionPercent`: The percentage of rotations made in the dominant direction (the direction in which greater number of rotations were made), based on the all rotations made at each level of gameplay. Ex: if 75 rotations were made in one direction and 25 in the other, then the feature value would be 75.

`CorrectedRotations_mean`: The mean number of rotations made in the opposite direction to fix too many rotations made in one direction across all episodes, to get the zoid to its final orientation.

`CorrectedRotations_nonzeroPercent`: The percentage of episodes that had corrected rotations.

`CorrectedTranslation_mean`: The mean number of translation made in the opposite direction to fix too many translations made in one direction across all episodes (in a level), to get the zoid to its final position.

`CorrectedTranslation_nonzeroPercent`: The percentage of episodes that had corrected translations.

`<ZoidType>_responseLatency`: The response latency for each zoid type averaged for the data. 'ZoidType' can be: 'StaticZoid', 'FlippingZoid' or 'RotatingZoid'.

`<X>_LineClearPercentage [Ex: 4_LineClearPercentage]`: The number of ’X’ line clears expressed as a percentage of total number of all line clears for a game level. ’X’ can take values 1-4.

Appendix B: PCA loadings

The table shows the contribution of each feature in the six factors that were obtained after rotation. The loading values for each feature specify the contribution of that feature to the factor.

Information about model fits, coefficients, and significance of each factor corresponding to linear regression models for each seed. Higher R-squared (R-Sq.) Value indicates a better fit

Factor 1:	planning-efficiency	(pl-eff)
Factor 2:	pile-management	(pile-m)
Factor 3:	zoid-control	(zoid-c)
Factor 4:	pile-uniformity	(pile-u)
Factor 5:	min-line-clears	(m-l-c)
Factor 6:	rotation-corrections	(rot-cor)

Features	Fact.1 pl-eff	Fact.2 pile-m	Fact.3 zoid-c	Fact.4 pile-u	Fact.5 m-l-c	Fact.6 rot-cor
EpisodeCount	0.213	0.293	0.336	0.246	−0.310	
CreatedOverhangs_Percent	0.259	0.449	0.338	0.153		
ClearedOverhangs_Percent	0.202	0.452	0.286	0.125		
CreatedWells_Percent		0.183		0.412	0.530	
ClearedWells_Percent				0.333	0.155	
WellDepth_mean				0.823		
2_DepthWells		0.241	0.131	0.372	0.354	
3_DepthWells		0.187		0.584		
4_DepthWells				0.633		
Gt4_DepthWells				0.650	−0.213	
CreatedPits_Percent		0.682	0.176		0.123	
ClearedPits_Percent	−0.158		−0.244		0.358	
PitSize_mean		0.857	−0.117		0.132	
Spire_Percent		0.220			−0.571	
RightWell_Score					−0.576	
PileScore	0.104	−0.842	0.159	−0.172		
MaxPileHeight		0.812		0.195	−0.264	
RespLatency_mean	0.969					
DropDuration_mean	0.148	−0.498	0.422			
DecisionLatency_mean	0.771		0.517			
DecisionLatencyPercent_mean	0.325	0.138	0.100		0.240	
ExtraRotations_mean	0.177		0.711			
ExtraRotations_nonzeroPercent	0.175		0.717			
DominantRotation_DirectionPercent						−0.785

(Continues)

Features	Fact.1 pl-eff	Fact.2 pile-m	Fact.3 zoid-c	Fact.4 pile-u	Fact.5 m-l-c	Fact.6 rot-cor
CorrectedRotations_mean						0.864
CorrectedRotations_nonzeroPercent						0.947
CorrectedTranslation_mean	−0.220		0.772			0.134
CorrectedTranslation_nonzeroPercent	−0.258		0.747			0.128
StaticZoid_responseLatency	0.726					
FlippingZoid_responseLatency	0.899					
RotatingZoid_responseLatency	0.902					
1_LineClearPercentage	0.145	0.260		−0.271	0.564	
2_LineClearPercentage					0.238	
3_LineClearPercentage				0.219	−0.254	
4_LineClearPercentage	−0.120	−0.172		0.139	−0.687	

Appendix C: Factor-distribution plots by expertise levels across game levels

Appendix D: Normal Q-Q plots for data at different expertise levels

Normal Q-Q plots obtained from linear-regression model fits for various combinations of player expertise and game levels. Points well aligned with the line indicate how well the similar the distribution of the data is to a standard normal distribution.

Appendix E: Normal Q-Q plots for seed-split data

Normal Q-Q plots obtained from linear-regression model fits for various seeds at game level 0. Points well aligned with the line indicate how well the similar the distribution of the data is to a standard normal distribution.

Appendix F: Defining player categories through clustering

To obtain any observable differences in skill when comparing different groups of players, players from one group had to be considerably better/worse at playing Tetris than players from other groups. Comparing players belonging to consecutive expertise levels would not be useful, since there would be too much overlap, for example a player with expertise-level 4 might have their top four games end at levels 2, 4, 5, and 6, while a player rated at expertise-level 5 could have their top four games end at levels 2, 4, 6, and 6. Both players in this example would likely have very similar sets of skills. So first, we had to identify expertise levels that when compared would present significant differences in skill.

We used criterion scores (see Section 6.1 for a detailed discussion on criterion scores) as a metric of expertise, to perform the clustering. Criterion scores define expertise at a finer granularity than expertise level, which helps the clustering algorithm form more precise clusters. We performed a univariate k-means clustering (Wang & Song, 2011) with three clusters, on the criterion scores for all 492 players. Dividing the data into three clusters was the optimal choice as suggested by the values we obtained for within-cluster SSE (in Section 4.2).

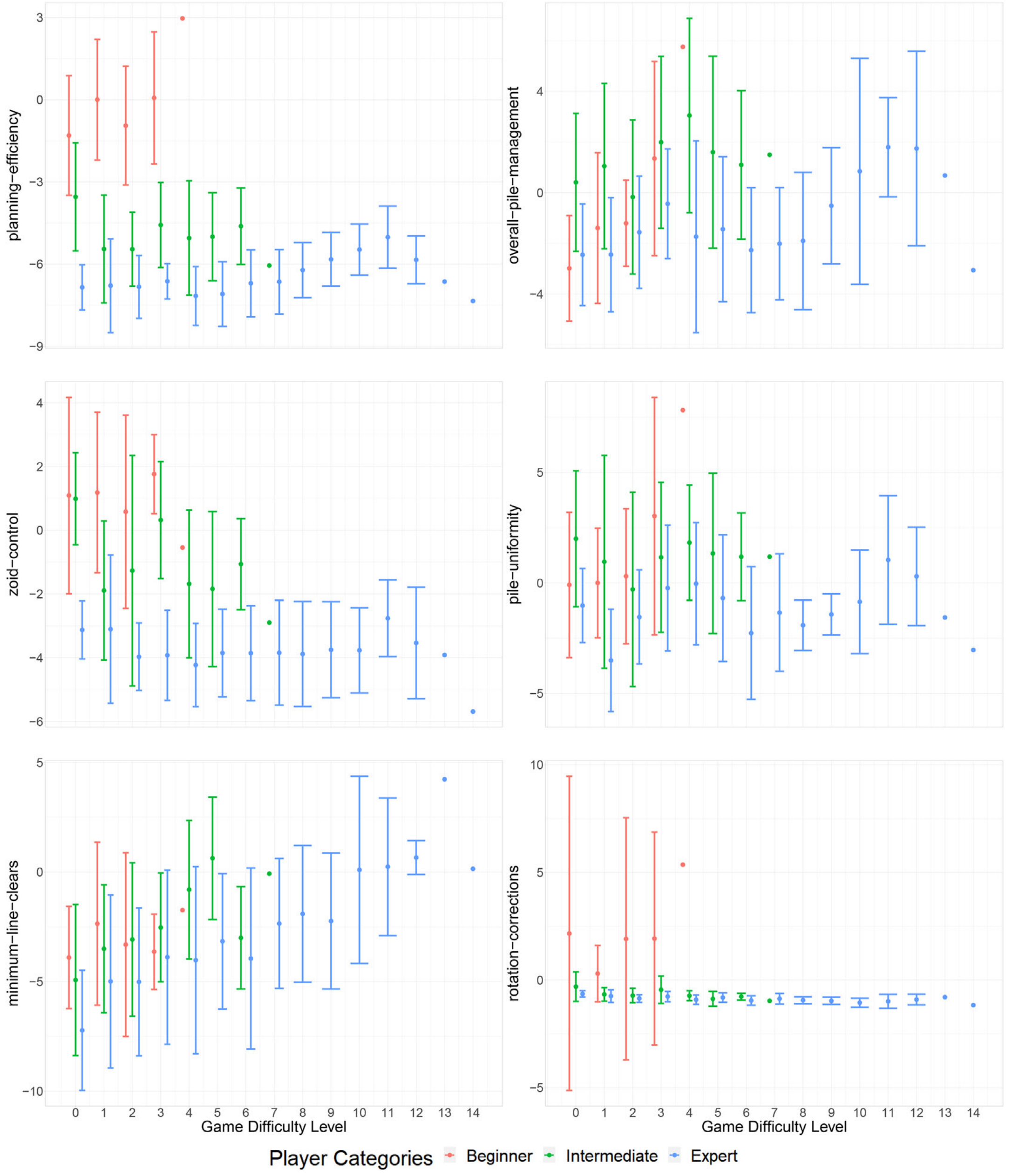
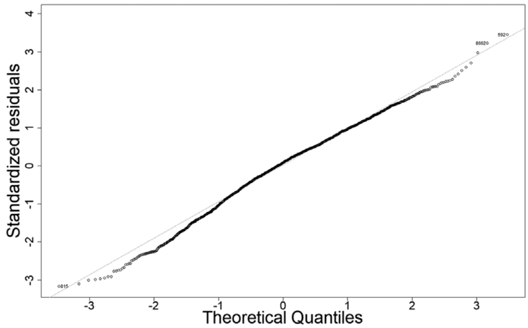
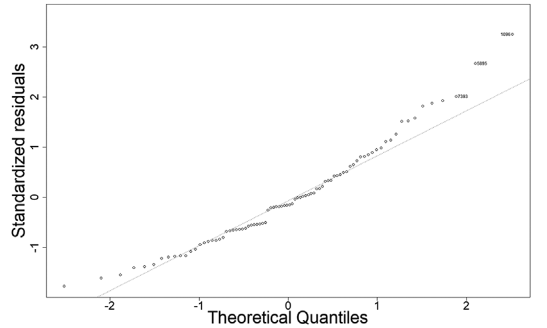


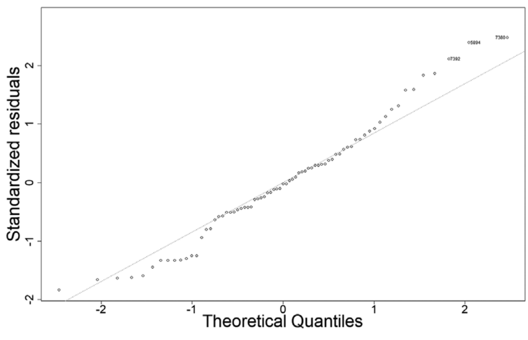
Fig. C.1. Mean and SD for factor values for each rank of players and changes with game difficulty. (Plots the top 3 players of each rank.)



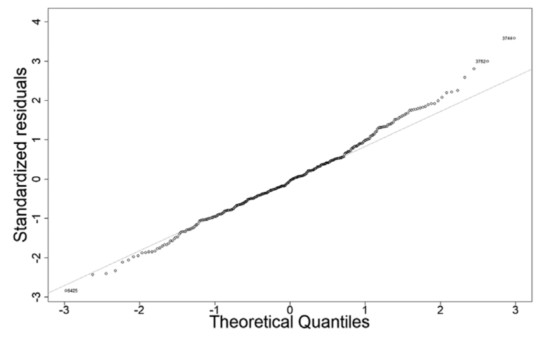
(a) *Model Fit for All Players at Level 0 (Model 1)*



(b) *Model Fit for Expert Players at Level 0 (Model 2)*



(c) *Model Fit for Expert Players at Level 8 (Model 3)*

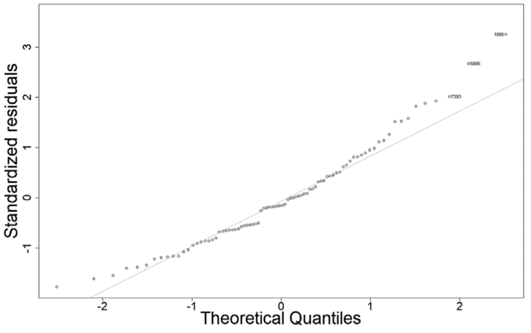


(d) *Model Fit for Intermediate Players at Level 0 (Model 4)*

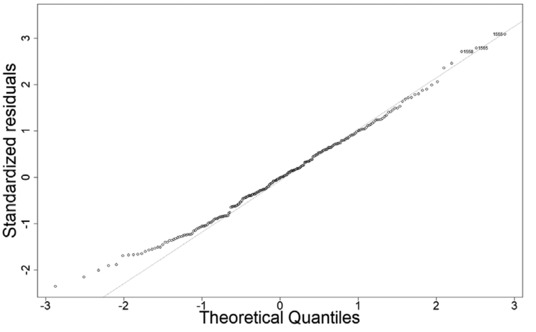
Fig. D.1.

However, we wanted to compare the distribution of data points when choosing three versus four clusters, to verify the results. Figure F.1 is the result of our curiosity. Choosing more than three clusters resulted in distributions with significant overlap among some clusters, especially among the last two clusters.

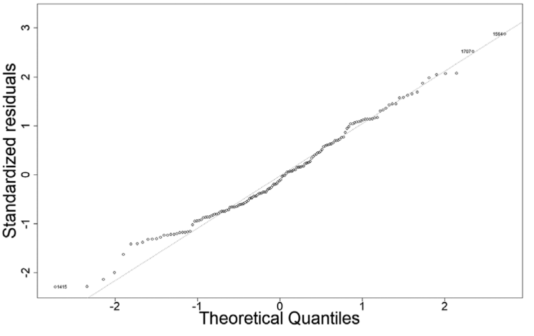
Once the clusters were defined and we knew which players belonged to each cluster, we could now use the expertise levels of all the players in each cluster to calculate the average expertise level (cluster average) for each of the three clusters, which when rounded off resulted in values 3, 6, and 9. This implied that the clusters were centered around players belonging to expertise levels 3, 6, and 9. To alleviate the possibility of overlap (maximize intracluster homogeneity and intercluster heterogeneity), only players belonging to the average expertise levels of each cluster were retained for the analysis, with the exception of expertise levels beyond 9. An exception was made for the higher level players because there were too few players at level 9 and, considering players with a higher expertise level would not lead to an overlap of skill with other groups.



(a) *Model Fit for Intermediate Players at Level 5 (Model 5)*

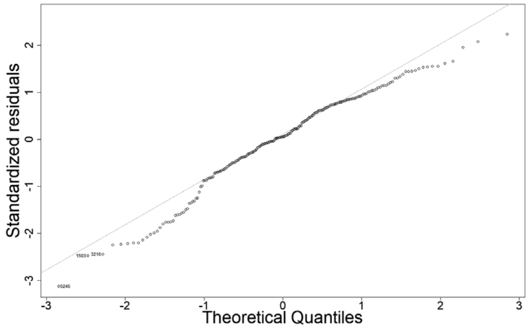


(b) *Model Fit for Beginner Players at Level 0 (Model 6)*

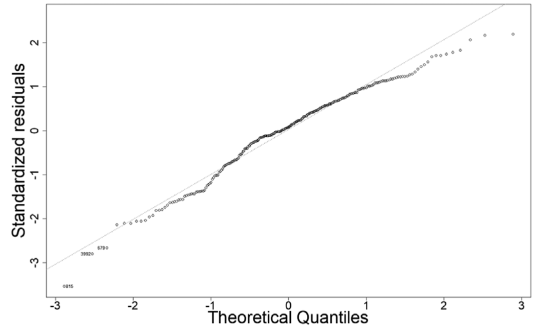


(c) *Model Fit for Beginner Players at Level 2 (Model 7)*

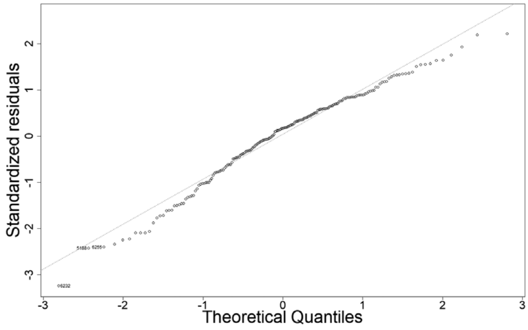
Fig. D.2.



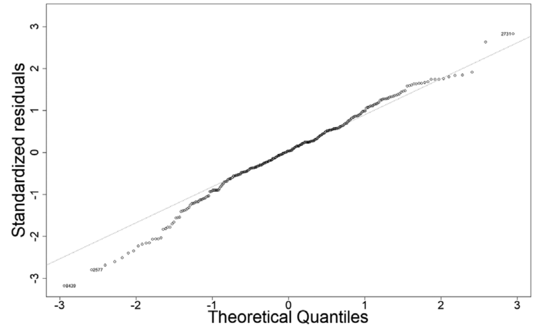
(a) *Seed 111*



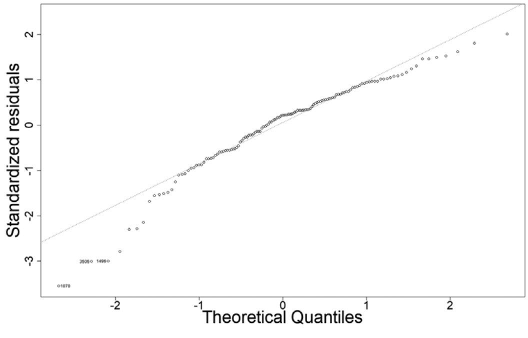
(b) *Seed 222*



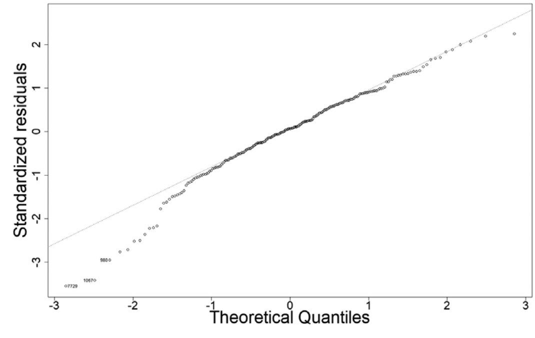
(c) *Seed 333*



(d) *Seed 444*



(e) *Seed 555*



(f) *Seed 666*

Fig. E.1.

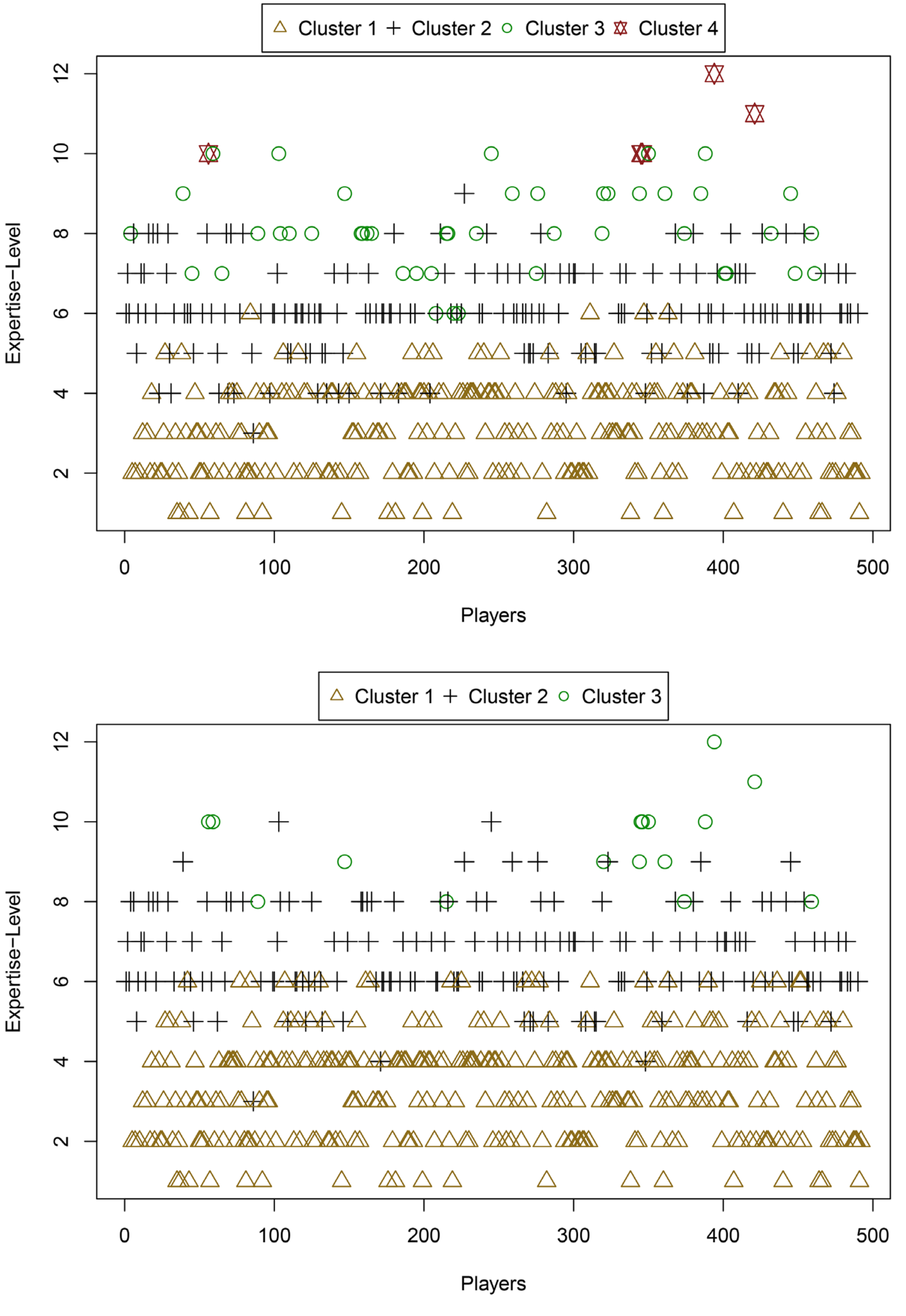


Fig. F.1. Cluster distributions when data are clustered into four (top plot) and three (bottom plot) groups. Each icon in the plot represents a single player and the icon type indicates which cluster the player belongs to. When the data are divided into four clusters, cluster 3 and cluster 4 almost completely overlap.