THE BUILDING BLOCKS OF EXPERTISE: EXAMINING EXTREME EXPERTS IN THE VIDEO GAME "TETRIS"

By
John K. Lindstedt

A Dissertation Submitted to the Graduate
Faculty of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: COGNITIVE SCIENCE

Approved by the
Examining Committee:

Wayne D. Gray
Dissertation Advisor

Michael Schoelles, Member

Brett Fajen, Member

Anders Ericsson, Member

Rensselaer Polytechnic Institute
Troy, New York

July 2017
(For Graduation August 2017)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Thank you, my family, my friends, whose patience and kindness helped me to make this work something I am proud of. I will never forget to pay it forward.

Thanks to my committee. Thanks to Wayne Gray for his constant guidance and friendship. Thanks to Mike Schoelles for his critical eye for programming and statistics. Thanks to Brett Fajen for holding our department's experimental research to a higher standard. Thanks to the venerable Anders Ericsson for his relentless enthusiasm for my work. And a special thanks to Sean Barton for sticking through it all in the neighboring cubicle.

This work was supported, in part, by grants N000140910402 and N000141310252 to Wayne Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer.

ABSTRACT

I present three related studies which examine the features and theoretical structures associated with extreme expertise in the complex, real-time cognitive task of Tetris. To facilitate this research, I developed a platform for experimental research in Tetris, Meta-T, which allows for careful experimenter manipulation and offers high-fidelity behavioral performance logging. My approach "zooms in on and unfolds" high-fidelity, rich human performance datasets to improve the resolution of the portrait of human expertise in complex, real-time tasks in 3 studies. Study 1 used principal component analysis (PCA) to explore the behavioral features of undergraduate Tetris players in the laboratory and identify a set of features associated with a player skill: latent skill component (I named the "decide-move-placed" component), composed primarily of measures of speed and efficiency. This component reliably distinguishes player skill across the spectrum of expertise, under different conditions of time pressure, and even succeeds when examining only a small portion of player data without knowledge of game outcomes. Study 2 validates the model of player skill constructed in Study 1 in two ways: first, by predicting the game scores of Tetris tournament players; and second, by predicting the outcome of individual tournament matches at a rate well above chance. Study 3 examines world champion Tetris player performance in conjunction with local tournament players to compare the novice and expert player strategies, highlight their consequences for performance, and illuminate the cognitive underpinnings of such strategies. I find that even world champion Tetris players remain impacted by Hick's law, indicating there may be some lower bound, beyond which the process of expertise acquisition can no longer adapt to and reduce a task's decision space. These studies together lead me to four conclusions: (1) player skill can be reliably predicted by examining the efficiency of execution; (2) measures of process appear to more reliably predict real-world expertise than measures of individual outcomes; (3) "Hick's law bends with practice, *except when it cannot*"; and (4) "both task and strategy together define the mold into which human expertise may shape itself."

CHAPTER 1

INTRODUCTION

The acquisition of expertise– the ability of some to exceed their peers by leaps and bounds in a domain– is a defining feature of the human species. We humans have leveraged our considerable cognitive capabilities, technology, and culture of education to "evolve in-place" by adapting to any and every niche in the environment. We also celebrate this endeavor: from incredible feats of athleticism in sports arenas or technical prowess in video games, to life-or-death conditions in military operations or on surgical tables, we honor and reward those who exhibit capabilities far beyond those of the general population.

Here I define an expert as an individual who has achieved an extreme level of skill in a domain relative to their peers, usually done so through long term (and deliberate) practice (Ericsson, Krampe, & Tesch-Römer, 1993). The acquisition of expertise is often conflated with the "power law of practice" (Newell & Rosenbloom, 1981) in that acquiring skills follows a pattern of rapid initial improvement followed by gradually diminishing returns for time spent practicing. However, in the realm of more complex tasks in which expertise is of particular interest, it is easy to get stuck in local maxima, or "stable sub-optimal solutions" (Fu & Gray, 2004). It is not always immediately obvious how best to solve complex tasks, and thus a given strategy may initially be attractive but not result in the most effective or efficient solution for the task. This leads to performance plateaus (Bryan & Harter, 1897; Gray & Lindstedt, 2016), wherein returns on practice halt well before the true asymptote for performance in the task– effectively violating the "power law of practice".

How is it, then, that we manage to achieve such heights of expert performance? Through what process does a pilot learn which decisions are good, which will suffice, which are dangerous, and what series of actions and contingencies ought to be adopted to ensure a high degree of performance? How is this process streamlined to reduce deliberation and increase fluid decision-making as the task's demands increase? Ericsson and Kintsch (1995) reason that by compiling piecewise task solutions called "chunks" (perceptual, memory, and motor), an individual develops a robust understanding of the task at hand, thereby reducing time spent deliberating and increasing the ability to make good decisions quickly.

The process of developing these chunks can also be seen as reducing the decision space of the task, or "bending Hick's law" (Logan, Ulrich, & Lindsey, 2016; A. R. Jensen, 1987). Hick's law states that reaction times in a task scale linearly with the log of the number of choices available. When obviously inferior solutions can be eliminated, though, that task space is reduced, thus "bending" the law. But is there a limit to the extent that the human cognitive system can adapt to and reduce the decision space of a given task? Or are further gains in cognitive efficiency always possible, however small?

This investigation seeks to better understand how human expertise manifests in a complex, real-time task environment. Because I cannot reliably expect to examine the behaviors of fighter jet pilots or surgeons in their domains of interest, I turn instead to the realm of video games, and specifically the video game Tetris. The investigation spans a wide array of scientific activity: exploration of wide data, feature analysis, construction and validation of statistical models, and (quasi-)experimental investigation of the causes and consequences of extreme expert performance for the world's best Tetris players.

1.1 BACKGROUND

There are three common ways in which video games are used in cognitive science research (Gray, 2017a); (a) gamification, (b) treatment conditions, and (c) experimental paradigms. The first attempts to identify elements that make games challenging, motivating, and enjoyable with the goal of altering the elements of nongame activities or instruction to make them more challenging, motivating, or enjoyable.

The second treats games as black boxes and examines differences in pre- vs post-game performance on some measures. This is by far the most popular use of Tetris and, in this mode, Tetris has been used for research investigations as diverse as ameliorating sex differences in spatial skills (Linn & Petersen, 1985; Okagaki & Frensch, 1994; Valerie K. Sims, 1996; Terlecki, Newcombe, & Little, 2008), relief from "flashbacks for trauma" (Holmes, James, Coode-Bate, & Deeprose, 2009), improving the abilities of engineering students (Martin-Gutierrez, Luis Saorin, Martin-Dorta, & Contero, 2009), and as a control condition for the effects of First-Person Shooter games on brain capacity (Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Green & Bavelier, 2003, 2006).

The third way games have been used is as experimental paradigms (Game-XPs) which, in common with the extremely simple paradigms used in most behavioral science laboratories (including ours), can focus on the basic processes of cognition, perception, action, dynamic decision making, method discovery, and skill acquisition. However, in contrast to simple paradigms, Game-XPs also focus on how the various cognitive processes act together in the acquisition and performance of manageably complex behavior (see Gray, 2017a, for a discussion of this point).

In the next section, I review prior research that treats Tetris as an experimental paradigm and report the richness of conclusions garnered from this task. I then discuss two distinct paradigms for studying expertise; namely, longitudinal and expertise sampling. In the last background section, I briefly discuss non-Tetris treatments of "response time" or "choice reaction time" ranging from its history as a measure of intelligence to its role in studies of expertise and brain training.

1.1.1 Tetris™ as an experimental paradigm

Amidst the 137 citations found to "Tetris and Game" in a Web of Science search[1], Game-XP is clearly a minority take on the use of Tetris. Within this Game-XP category, perhaps surprisingly, the search reveals that Tetris is not dominated by cognitive science studies but by those from the machine learning community, which I will return to shortly.

Investigations involving the use of Tetris as a Game-XP for cognitive science have focused on expertise in real-time, sequential and dynamic decision-making tasks – for situations in which, "even hesitating requires a decision." For these purposes, Tetris provides a complex, but not too complex, environment within which some people are novices and others, after years of work and practice, become extreme experts.

1.1.1.1 Tetris and complementary action

Kirsh and Maglio (1994) were the first to use Tetris as an experimental paradigm. Their focus was on the tradeoffs people made between performing physical actions in-the-world rather than mental actions in-the-head, also known as "complementary actions". Tetris requires players to rotate pieces (called zoids) and to move them left-to-right. A certain number of rotations and lateral movements are needed to position each piece. Kirsh and Maglio hypothesized that physical movements, beyond the minimum needed to move a Tetris piece to its final location, signal the off-loading of the search for the best place to put the piece from the purely mental to the partially physical. Hence, physical action, rather than simply moving the piece to its final location, provided a faster and easier means of deciding if a given piece would fit in a given location.

[1]Performed on 2017.02.13.

Kirsh and Maglio collected data from two players who started as Tetris novices and played for 20 hrs. Destefano, Lindstedt, and Gray (2011) followed up that study with a dataset of the best of two games played during the qualifying rounds for a Tetris tournament from 59 players.[2] They divided players into five groups based on the scores made during the qualifying round and concluded that by the most lenient of criterion, the number of complementary (or epistemic) actions peaked with the second group and decreased greatly across the top three player skill categories. More stringent criteria showed little evidence of epistemic actions. In conclusion:

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

1.1.1.2 Tetris and machine learning

In the last 10 years, Tetris has been discovered by the machine learning community and used as an experimental paradigm for testing various machine learning algorithms (C. P. Fahey, 2015; Gabillon, Ghavamzadeh, & Scherrer, 2013; Şimşek, Algorta, & Kothiyal, 2016; Szita & Lorincz, 2006; Thiery & Scherrer, 2009b, 2009a). Sibert, Gray, and Lindstedt (2017) borrowed the Cross-Entropy Reinforcement Learning (CERL) model from the machine learners (Szita & Lorincz, 2006; Thiery & Scherrer, 2009b) and varied the objective function to determine whether the models created by different objective functions would differentially predict human performance.

The objective function favored by the machine learning community is simply "survive the longest". Having been trained on this objective function, their best models would clear hundreds of thousands of lines (more on line clearing in the next section). Of course, the version of Tetris played by the models was a mere simulation; that is, these models did not have to deal with increasing game speeds or substantive temporal costs of decision-making. Similarly, the models did not have to worry about the time needed to physically move game pieces to desired locations. Just as clearly, the models could play for extremely long periods of time, as they never became tired, never became bored, never became hungry, and never had to deal with anything that might distract a human.

[2] This was a local tournament open to all without an entrance fee. Hence, those who entered the qualifying rounds ranged from "extremely optimistic novices" to "very good". Also notable is the fact, despite originating from the Cog-Works research laboratory, that the tournaments were played in the 2000's which occurred prior to the development of the Meta-T (Lindstedt & Gray, 2015) used in this work. Hence, these data were collected using a different implementation of Tetris and data collection metrics than used in the current research.

When Sibert et al. (2017) trained their models using different objective functions, they found that the model trained to optimize actual game score, rather than simple survival, provided a better match to the human data with differences in score accounting for approximately 40% of novice human placements and nearly 65% of expert human placements. Unlike the machine learning studies where, during training, the models played until they lost the game, the new models trained for a maximum of 506 game episodes.[3] Although the score-based model cleared fewer lines than did the simple survival model (168 vs 200), its score was notably higher (175,455 versus 103,342). Studying the feature weights learned by the two models, they concluded that the score-based model had developed a higher risk, higher payoff policy whereas the simple survival model's policy reflected a slower rate of gain and an aversion to risk. These findings represent a shift away from merely treating the Tetris task as a robust toy box for machine learners, and into the realm of better understanding human cognition and expertise in real-time tasks.

1.1.1.3 Sequential decision making, simple heuristics, and Tetris

Recent work by Şimşek et al. (2016) transcends the boundary between machine learning and human performance to argue that the nature of decision making in Tetris is such that human and machine learners "can choose well among the available actions without knowing an evaluation function that scores well in the game." Put another way, their interest is not in Tetris behavior and performance per se but in using Tetris as an experimental paradigm "for identifying regularities in sequential-decision environments."

They characterize Tetris as a sequential decision making task in which the placement of the game pieces are largely independent of the placement of the piece immediately prior. This treatment follows Simon's observation that explanations consist of two components, (1) antecedent conditions and (2) general laws to explain how each situation causes the succeeding one, so that "Laws acting on the current state of the system produce a new state–endlessly" (Simon, 1992, p. 152).

In my view, Şimşek's work is notable in showing how simple heuristics can reduce a complex problem (in which there are up to 34 possible locations to place the current game piece) into a much simpler problem. As such, the work represents yet another strong step in the unification of machine learning with human performance under the banner of heuristic decision making (Gigerenzer & Brighton, 2009). However, their conclusions come with a caveat: the Sibert et al.

[3] The span of 506 episodes was picked as, at the time of the Sibert et al. (2017) study, that was the highest span of Tetris played by any player in the laboratory study.

(2017) work that shows that the optimal model for human performance weighs the features of a Tetris board differently than does an optimal model for machine learning. Şimşek and colleagues applied the Thiery and Scherrer (2009b) weights to machine, random, and human zoid placements. I suggest that if the feature weights that Şimşek et al. (2016) had applied to their human data had been optimized for human behavior, that their results for "people" (as in their Figure 2) would be even stronger than reported in their paper.

1.1.1.4 Expertise sampling versus longitudinal studies

Many people play Tetris and many people who do not yet play Tetris would enjoy the game enough to devote time to acquiring Tetris expertise. As Gray argues (2017a), the ability to sample expertise across individuals as well as the ability to collect longitudinal data on the same individuals is an advantage for the Game-XP approach. However, if the threshold for controlled studies of expertise is that all sessions of practice must have occurred in the laboratory under the same conditions of practice, then there are few genuine longitudinal studies of development of expertise. Indeed, the two that come to mind are the carefully controlled studies on digit span conducted by Chase, Ericsson, and colleagues (Ericsson, Chase, & Faloon, 1980; Chase & Ericsson, 1982; Richman, Staszewski, & Simon, 1995) and the mental multiplication study conducted by Staszewski (1988). The former included 220 separate sessions, each collected on a separate day, collected from one subject (called "SF") and 850 sessions collected from a second subject (called "DD"), each collected on a separate day. Staszewski's (1988) studies of mental multiplication consisted of two subjects, JA and GG, who did mental multiplication in the lab for 45 minutes per day, 3-5 days per week. JA did a total of 268 sessions over 3 years (approximately 175 hrs), each collected on a separate day, whereas GG did 618 sessions over 4 years (approximately 300 hrs of practice), each collected on a separate day.

My repetitive use of the phrase "each collected on a separate day", is purposeful and is accompanied by feelings of awe and (at least in my mind) a bow of respect in the direction of the university where this substantive work was done. These are truly longitudinal studies that take the subjects from "normal" to "extreme expertise". However, because all training occurs in the laboratory, the negative side of this otherwise excellent work is that very few subjects are used and that the labor required by the researchers and the dedication required by the subjects makes these studies extremely rare.

A much more common approach is *expertise sampling*, a form of cross-sectional analysis, in which subjects are recruited with different levels of expertise. An excellent example is Logan et al.'s (2016) recent study in which 48 typists were recruited, 24 of whom used the standard keying and 24 of whom used nonstandard keyings. Hence, participants were recruited based on differences in the expertise they had acquired prior to beginning the study. In this work I take a similar approach to sampling Tetris expertise, both that distributed in the Rensselaer undergraduate population, and by seeking out champion players both locally and globally (more on this in the studies that follow).

1.1.2 Speed: Changes in cognitive capacity or efficiency?

Speed, if not speed per se, has always been a contender in discussions of extreme performers whether it be in the context of Telegraph Operator prowess (Bryan & Harter, 1897) or measures of intelligence. [Indeed, the modern study of intelligence started with Galton's attempts to correlate response times with intelligence (Johnson et al., 1985; Arthur R. Jensen, 1982, 1998).] Hence, training that produces, say, a half standard deviation increase in fluid intelligence (as claimed by Jaeggi, Buschkuehl, Jonides, and Perrig (2008) but discredited by Redick et al. (2013)) would present an increase in *cognitive capacity* whereas an increase in typing speeds due to acquiring skill in touch typing would reflect a change in *cognitive efficiency* (see also section 5.1 of Gray & Lindstedt, 2016).

It is, of course, well established that within a given task speed increases as a log-log function of the number of trials of practice (Fitts & Posner, 1967; Newell & Rosenbloom, 1981; Gray, 2017b). Some of these speed increments are due to knowledge compilation processes (John R. Anderson, 1987) which shift performance from controlled to automatic processing (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Some are due to domain-specific alterations in the Hick's (1952) Law coefficient, which predicts response time as a function of the number of choices made. For example, Logan et al. (2016) have shown that Hick's law diminishes with practice such that at extreme levels of practice (such as is the case for many people who are touch typists) the coefficient for choice goes to zero; that is, although several alternatives exist, any given response is made as quickly as if there were no alternates to choose from. Other speedups are due to the discovery or invention of new methods (Gray & Lindstedt, 2016).

Many researchers in the "Game as Treatment Condition" tradition (see above and Gray, 2017a) have reported faster response times in pre- and post-test training studies (e.g., A. F. Anderson &

Bavelier, 2011; Dye, Green, & Bavelier, 2009; Green & Bavelier, 2003) on various simple paradigms. These have sometimes been interpreted as signaling changes in cognitive capacity due to the experience of playing a game. However, the evidence that the post-test changes reflect general changes in cognitive abilities is mixed, with at least one large scale study (Boot et al., 2008) questioning the reliability of these findings. Indeed, it seems to be an open question as to whether changes in post-test performance reflect changes in cognitive capacity or in cognitive efficiency due to the application of methods acquired within a game to post-test performances.

These issues and the questions they raise are complex. However, none of the researchers who favor the "game as treatment condition" approach has ventured over to the Game-XP approach to find within-game evidence as to how expert and novice game players differ. Although the current research will not answer the question as to whether playing games produces general changes in cognitive capacity or task specific changes in cognitive efficiency, it does offer a start on understanding the essence of skilled performance within one complex skill domain, Tetris, in which mastery requires planning, perceptual learning, rapid decision-making, and efficient motor movements.

1.2 RESEARCH OVERVIEW

The present investigation uses high-fidelity real-time behavioral data to examine the nature of extreme expertise in the Tetris video game task. The investigation is composed of three studies: studies 1 and 2 involve identifying and verifying features of expert performance, and together comprise the content of a manuscript submitted for publication at time of writing; study 3 investigates the cognitive underpinnings of discovering and adopting expert strategies in Tetris.

1.2.1 Study 1: Modeling expert performance

In study 1, I explore the behavioral and task feature space of Tetris. Using principal component analysis, I identify the latent elements of task performance which differentiate player performance at different skill levels. In this study, I ask two main questions. First, an empirical question: can a player's level of expertise be predicted from only a small subset of their performance data, lacking *a priori* knowledge of the game's final outcome? Put another way, can we know a player's skill level "just by looking", or do we need to accumulate some critical mass of player performance data to know for sure?

Second, which elements of task performance are most associated with expert ability? Hick's law predicts that measures of decision latency are likely to be associated with increased skill, but can we continue to reduce these latencies indefinitely, or is there limit to such improvements after which only changes in strategy can offer improvement? It may be the case that the majority of decisions made in Tetris become trivially solvable by the human cognitive system as skill is acquired, and thus only the broader elements of strategic decision-making within the task– not reaction times– will account for differences in player skill.

Thus, despite study 1 being primarily exploratory in nature, I ask two questions: (1) can player expertise be predicted using only small subsets of player performance data? (2) which elements of task performance covary with player expertise, and are these elements primarily temporal in nature?

1.2.2 Study 2: Verifying the expert performance model

In study 2, I verify my model of expert Tetris performance developed in study 1 by using it to predict novel performance data collected "in the wild" at locally hosted Tetris tournaments. In this study I perform two kinds of model validation: a classic form of model validation, in which I predict unseen players' skill using unseen performance data, and a more novel approach in which I use my model's skill estimates to predict the winners of real-world prize tournament match-ups. This latter form of validation addresses practical question about my model of expertise: what is a better predictor of player skill, the outcome of a single complete game (the score of which is used to determine the initial tournament ranks), or an assessment based on player performance? Are known outcomes better predictors of skill than features of players' decision-making process?

Thus, in study 2, I ask: (1) does my model successfully generalize to novel players in a real-world setting? And (2) is my model of expertise based on players' decision-making processes a better predictor of tournament winners than single-game scores in the qualifying round of the tournament?

1.2.3 Study 3: Modeling expert memory strategies

In study 3, I drill deeper into a cognitive theoretical account of expert strategies in Tetris. In collaboration with the Classic Tetris World Championship (CTWC) organization, I compare data collected from world champion Tetris players with that of local tournaments to investigate the very edges of human cognitive ability in the task. Zooming in on my primary predictor of skill from the model validated in study 2, I examine millisecond-level measures of task performance to construct an account of both expert and non-expert strategies in the "zoid rotation" subtask, how the two strategies differ cognitively, and what those differences suggest about skill acquisition more generally.

Thus, in study 3, I ask: (1) to what extent do world champion Tetris players differ in their task performance from both novice and regional champions? (2) to what extent do player strategies differ between novices, regional champions, and global champions, and what impact do these differing strategies have on performance? And, (3) what do these findings imply about the memory structures involved and their implications for the process of strategy acquisition more generally?

Figure 1.1 shows an outline of the basic flow of data and analysis presented in this document. Taken together, these three studies comprise a scientific investigation that I believe to be a thorough account of some specific aspects of human expertise in the Tetris task domain. Before examining each study, let me first illuminate my primary research apparatus: the experimental Tetris research platform, Meta-T.

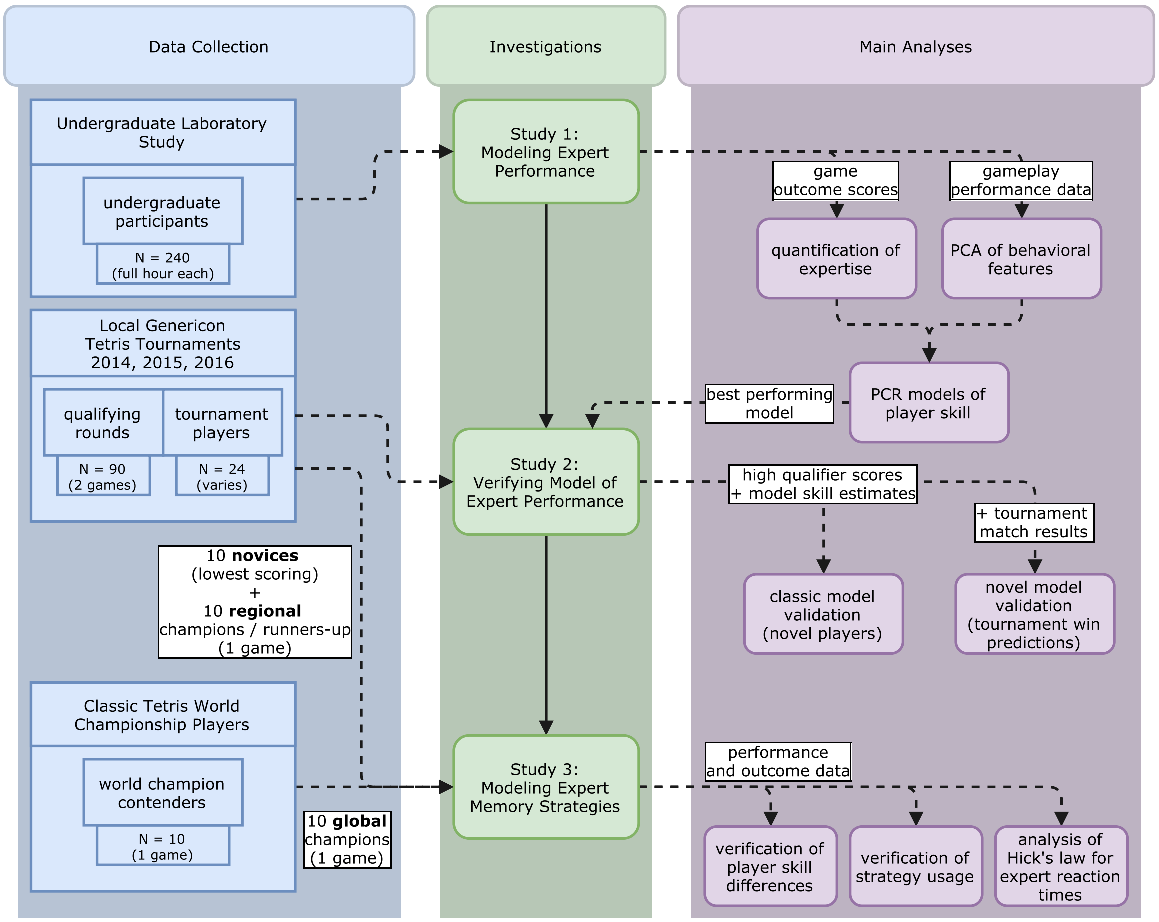


Figure 1.1: A map of the three studies in this document, their associated datasets, and the main analyses performed in each.

CHAPTER 2

META-T: A PLATFORM FOR EXPERIMENTAL INVESTIGATION OF THE TETRIS TASK

As a cognitive task, Tetris hits the sweet spot of being simple to comprehend and tractable to analyze, but complex enough to remain cognitively interesting and rewarding to master. The game space is simple to parse (both by humans and machines), and though a novice can be taught the game's rules in a matter of seconds, to become highly proficient at the task, players must spend time to develop methods that efficiently deploy working memory, mental rotation, strategic planning, prediction, manual dexterity, etc., to negotiate the ever increasing time pressure of the task. It is no simple feat to become a Tetris grand champion [1] (i.e., a highest-scoring player, as determined by a championship tournament), and to that end it is an excellent space in which to examine the phenomenon of human skill acquisition. A bonus attractive feature of Tetris is that the game is inherently motivating (i.e., fun and engaging) and has a rich life outside of academic laboratories: there already exist people who go to great lengths to perform well in the task, due in large part to the fact that the game is fun and engaging.

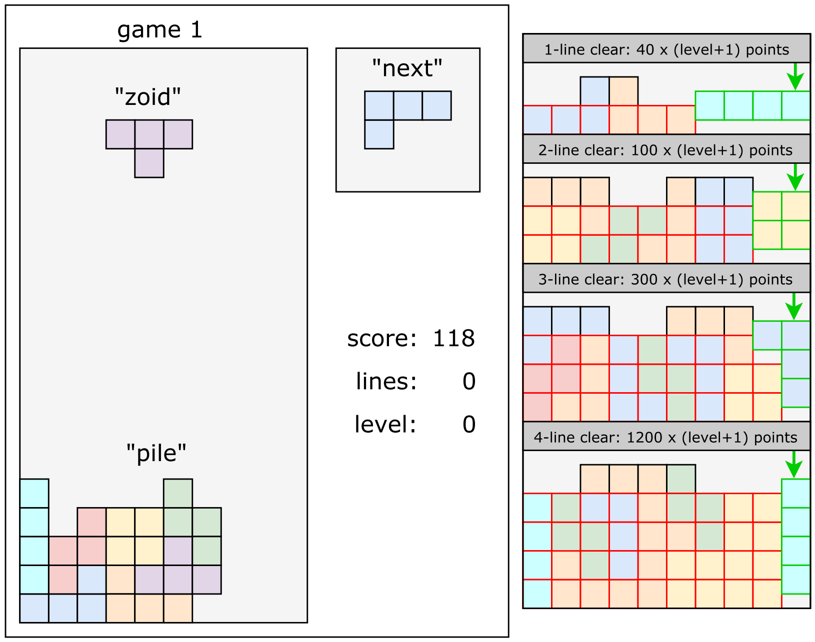
2.1 RULES OF TETRIS

In Tetris players use button-presses to manipulate a series of falling shapes (*zoids*) into an arrangement of previously placed zoids at the bottom of the game space (the *pile*). The player's goal is to maximize their score by filling, and thereby eliminating, rows in the pile. Once a row is filled, it is removed, lowering the pile's height and awarding the player points. Since it is not always possible to clear lines with a given zoid, the pile gradually rises. As players clear lines the game level increases, meaning the drop-rate of the zoid increases, as well as the points awarded. The game ends when the pile rises above the upper threshold of the game space or there is no room for a new zoid to appear. A basic representation of a game board is shown in Figure 2.1. [2]

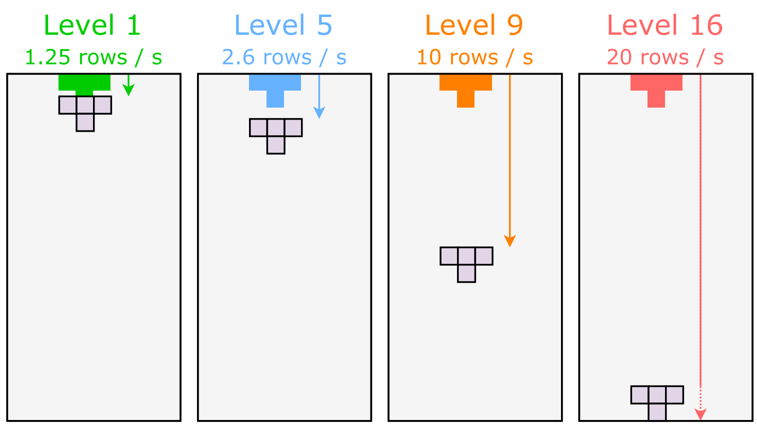
Portions of this chapter previously appeared in: Lindstedt, J. K. & Gray, W. D. (2015). Meta-T: Tetris as an experimental paradigm for cognitive skills research. Behavioral Research Methods. doi:10.3758/s13428-014-0547-y

[1] e.g.,http://www.youtube.com/watch?v=eQOswiAGLU4 and http://www.youtube.com/watch?v=gbyCSFrrcsM

[2] Interestingly, despite its widespread appeal, Tetris is not "winnable"– if played long enough, the player will either succumb to the increasing demands of time pressure, or encounter a random series of zoids for which there is no line-clearing solution (Baccherini & Merlini, 2008; Kendall, Parkes, & Spoerer, 2008; C. Fahey, 2007)!



(a) Tetris task elements.



(b) Time pressure in Tetris.

Figure 2.1: Elements of the game of Tetris. On the left of figure (a) is an example screen from the game with labels added: the “pile” refers to the accumulated block segments at the bottom of the game space; the “zoid” is the active game piece that falls from the top of the screen and can be manipulated by the player; and the “next” area is a preview of the next active zoid. The player’s current score, lines cleared this game, and current difficulty level are displayed, as well as the current game number in the session. On the right of figure (a) are examples of filling and clearing 1 and 4 lines are shown with the score awarded at difficulty level 0. Figure (b) shows the time pressure present at different levels of the game in the form of the amount of vertical space automatically traversed by the zoid in the span of 1 second. Point values awarded at later levels are multiplied by $L + 1$, where $L = thecurrentlevel$.

There are seven unique shapes for zoids, one for each permutation of four contiguous block segments, often denoted by a letter approximating the shape of each zoid: O, I, Z, S, T, L, and J. The events that unfold in the period of time between a zoid appearing and it being placed into the pile defines an *episode*. The player can move the zoid left, right, or rotate it; however, while the player is deciding how to maneuver the zoid, it will continuously and automatically drop down from the top of the game space. At the start of the game this drop rate is very slow, offering players time to deliberate; however, as the player clears lines, the game level increases, and the zoids begin dropping much more quickly, offering very little time for deliberate thought. To add an incentive for achieving higher levels of difficulty, the amount of points awarded increases multiplicatively with each level of difficulty.

An important feature of the game's scoring system is that the points increase when multiple lines are cleared at once. In the original version of Tetris, at level 1 (i.e., the start of the game with a very slow drop rate), a single line cleared is worth 40 points. Clearing two lines at once is worth 100 points, three are worth 300 points, and four yields 1200 points. In contrast, by the time a player reaches level 20, clearing one, two, three, and four lines simultaneously are worth 800, 2000, 6000, 24000 points, respectively. Incidently, clearing four lines at once is known as "scoring a Tetris," the game's namesake, and is notable because of its high payoff and high risk (due to increasing the pile height).

2.2 TETRIS AND BEHAVIORAL RESEARCH

Most college campuses have a population of expert Tetris players as well as novices who can be motivated to put in the hours needed to acquire at least moderate expertise. For example, several researchers have investigated the near- and far-transfer effects of Tetris in reducing gender differences in spatial skills (Okagaki & Frensch, 1994; V. K. Sims & Mayer, 2002; Terlecki et al., 2008) or enhancing the spatial skills of engineering students (Martin-Gutierrez et al., 2009). Others (Haier, Karama, Leyba, & Jung, 2009) have used Tetris to search for general cognitive enhancements from training by examining changes in neural densities associated brain areas in participants who played Tetris. These studies varied hours of Tetris played and looked at how the performance measures provided by the game (such as total score, level of play, and number of lines cleared) correlated with measures of spatial ability.[3]

[3] A *Web of Science* search conducted on 2014.10.13 found 118 papers in fields as diverse as computer science, psychology, engineering, neuroscience, educational research, robotics, psychiatry, linguistics, and rehabilitation that used Tetris or players with experience playing Tetris in their research. My own survey of the literature adds another 15 conference or

As mentioned in the previous section, an approach that focused less on overall Tetris performance and more on details of game play was taken by Kirsh and Maglio (1996, 1994) to advance the general (i.e., not specific to Tetris) claim that the use of complementary actions increased with expertise. The term *complementary action* pertains to physical actions that generate information rather than implementing an immediate goal. For example, when trying to decide if a Tetris zoid (i.e., piece) could fit into an available space, Tetris players could obtain the same information by either mentally (as per Shepard & Metzler, 1971) or physically rotating it. Physical rotation for the purpose of direct comparison in-the-world, rather than mental rotation in-the-head, would be considered a complementary action.

Though Kirsh and Maglio are short on some specifics (Maglio & Kirsh, 1996), they collected data from two novice Tetris players at various periods during 20 hours of Tetris play. It is not clear from their reports whether keystroke and event data were logged to a computer file, recorded by videotape, or tallied by observers in real time– whatever the case, it was likely a substantial effort to retrieve such data from an off-the-shelf version of the Tetris game. In contrast, to obtain the same sort of data that Kirsh and Maglio report, Destefano et al. (2011) wrote a simple, custom version of Tetris using the Flash programming language that logged keystroke and game state events in real-time. This software allowed them to collect detailed response time, system event, and screen object data from 59 experienced Tetris players as they participated in two studies. Their much more detailed data allowed them to algorithmically distinguish between five definitions of epistemic action to conclude that while there was some evidence that weak players exhibited epistemic actions, the frequency of epistemic action decreased with moderate expertise and vanished by extreme expertise.

Destefano's (2011) study using experienced Tetris players in an experimental paradigm convinced me that there were many interesting questions ready to be asked beyond game performance score and beyond the use of Tetris as a treatment condition in studies of spatial skill acquisition (e.g., as per Haier et al., 2009; V. K. Sims & Mayer, 2002; Terlecki et al., 2008). Indeed, though Tetris is more complex than most experimental psychology paradigms, I believe its complexity is manageable for certain research questions of interest to cognitive science. In particular, it provides a mathematically tractable domain in which to study skill acquisition, control of cognition, strategy discovery, time-stressed decision-making, and the interactions among cognition, perception, and action in a dynamic task environment.

journal papers not found in the Web of Science search, which brings the total scholarly papers using Tetris experienced players to a minimum of 133.

Apart from detailed event and state logging, off-the-shelf Tetris *as an experimental paradigm* also lacks the ability to easily manipulate isolated parts of the game environment, which is precisely what I set out to implement in Meta-T, my custom version of Tetris designed top to bottom as an experimental research platform. Take, for example, Figure 2.2 (Haier et al., 1992): this growth curve depicts only one simple measure of performance, the average number of lines cleared. This relatively abstract performance outcome variable leaves much to the imagination; what were these players doing? Were they adopting better strategies as they progressed, or were they simply getting faster at manipulating the game's controls? Was the change in performance due to something general or something more task-specific? If this experiment had employed Meta-T, a detailed picture of human skill acquisition would be available for empirical scrutiny.

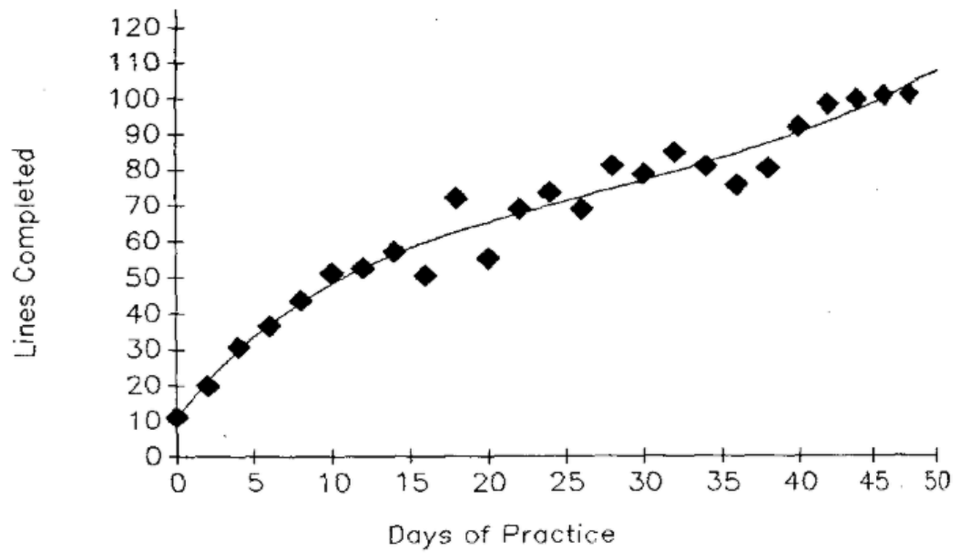


Figure 2.2: Growth curve of players' average lines cleared as practice days increased. Reprinted from Haier et al. (1992).

For the purposes of investigating cognitive phenomena and effects, Tetris may not appear at first to be "deeply cognitive". Tetris does not offer the depth of forethought required by Chess, the broad situational awareness of Go, or perhaps the abstract thinking involved in formal logic. Tetris does, however, require the use of various human cognitive abilities *in real time*. Compared to many classic laboratory paradigms for testing cognitive effects (i.e., N-Back or AX-CPT), Tetris is a rather complex task in which players need to respond to immediate time pressures, executing chains of motor actions while simultaneously planning ahead for known upcoming zoids (as displayed in the "next" box); moreover, if players wish to successfully score high numbers of points, they need to plan many zoids ahead, over uncertainty, to successfully enable many favorable action opportunities, such as the high risk but rewarding "Tetris" maneuver. Though Tetris may not be the "holy grail" of interactive cognitive domains, I suggest that Tetris represents an ideally tractable platform for understanding the complicated choreography between perception, action, and cognition.

2.3 META-T: A CUSTOM EXPERIMENTAL PARADIGM

Meta-T is a custom Tetris variant developed with the support of the CogWorks Laboratory from 2013 to present. It is written in the Python programming language, using the Pygame package for video game development. All elements of the game replicated from the original NES version were made to mimic the specifications present in the community-driven Tetris wiki (http://tetris.wikia.com/). The code is licensed under a GNU General Public License, and is available on request from the CogWorks Laboratory GitHub repository (http://www.github.com /CogWorks/Meta-T). By default, the game is a close replication of the 1985 version of Tetris for the Nintendo Entertainment System (NES). Its true contribution, however, is in two transformative additions to the program: experimental manipulations, and detailed logging.

2.3.1 Experimental manipulations in Meta-T

Meta-T is equipped with a simple text-based interface for altering all desired aspects of the game. All manner of game elements are available for modification. To list a few: the falling speed of zoids can be specified for every level of difficulty in the game; the way line-clears are scored can be altered; the method of randomizing the order of zoids can be specified; the size and orientation of the game space can be changed; the set of zoids presented to the player can be restricted or extended; even the availability of the "next zoid" preview can be switched on and off.

2.3.2 Logging in Meta-T

Meta-T also produces robust logs of the behavioral data during play. Three levels of logs are produced: Game-level, Episode-level, and Interaction-level (called "complete"). Game-level logs contain high-level or aggregate performance data, such as scores, durations, episode counts, etc. Episode-level logs contain information pertaining to each individual episode of the game, including features of the pile and details of the episode itself (configuration of the pile, current and next zoids, etc.). Interaction-level logs contain keypress and game state information at or below millisecond-level. The interaction-level logs are detailed enough to play back entire game sessions in real time based only on the contents of the log.

2.3.3 Meta-T summary

Video games have recently become a popular domain for studying differences between experts and novices, boasting benefits like inherently detailed data, motivated– and large– subject populations, and easy access to large pools of data. To be able to carefully study one of the most prolific video games ever created– Tetris–, I created a detailed and robust experimental paradigm for capturing all the minutia of player performance called Meta-T (Lindstedt & Gray, 2015). Meta-T serves as the primary research platform for the present investigation.

The combination of the system of customization and detailed logging offers unprecedented control over the Tetris task environment and sensitivity to its players' behavior profiles. Only a handful of the experimental parameters will be manipulated in the investigation presented in the subsequent sections, and only a subset of all of the possible analyses of the output data can be feasibly examined in the present work. I hope, however, that the work creating this flexible and robust paradigm will pay off by paving the way for deeper investigations of the Tetris task domain.

2.4 NOTES ON TETRIS AS A COGNITIVE TASK

To re-center the perspective of this project: Tetris is not the primary focus in and of itself, but rather, it is a surrogate for other, more complex task domains that are somewhat more difficult to study. More interesting domain populations, such as laporoscopic surgeons (Keehner et al., 2004; Keehner, 2011), pilots (Hays, Jacobs, Prince, & Salas, 1992; Proctor, Bauer, & Lucario, 2007), submarine operators (Ehret, Gray, & Kirschenbaum, 2000), or even call-center employees (Gray, John, & Atwood, 1993) are, unsurprisingly, not trivial to recruit for study. In this way, Tetris is meant to fill in as a task that consists of cognitive elements being deployed in real-time, similar to other tasks of interest. Though treating Tetris as an interesting transfer domain (Okagaki & Frensch, 1994; V. K. Sims & Mayer, 2002; Terlecki et al., 2008; Martin-Gutierrez et al., 2009) is of theoretical interest, the following work is meant to highlight Tetris primarily as a single, detailed case examining expertise in other, similarly structured tasks.

CHAPTER 3

STUDY 1: LABORATORY EXPLORATION OF TETRIS EXPERTISE

The goal of this research is to identify which behaviors distinguish novice from expert players in a complex, sequential and dynamic decision-making environment – the video game Tetris. Here I address two research questions:

1. What can a detailed factor analysis illuminate about expert behavior in a complex, real-time cognitive task environment?

2. How does time pressure affect player behavior at different levels of expertise? What is relevant in distinguishing highly skilled players as time pressure increases?

3.1 STUDY 1: METHOD

My first analysis requires 3 steps: first, deconstruct the task environment into measurable elements of player performance; second, reduce and combine this set of behavioral factors into maximally informative components of behavior in the task; and finally, run a multiple regression with those components to examine which components distinguish novice, intermediate, and expert performance.

3.1.1 Participants

Over the course of 5 semesters between 2013 and 2015, 240 participants were recruited from the undergraduate populations of General Psychology and Sports Psychology courses. Participation was rewarded with course credit (or hourly compensation of $10 for a session during Summer semesters). Participation in the study was reviewed and approved by the Rensselaer IRB.

3.1.2 Task environment: Meta-T

As mentioned earlier, the task of each subject was to play the Meta-T (Lindstedt & Gray, 2015) implementation of "Classic Tetris" (Wikipedia, 2016). Although there are minor visual differences stemming from my use of Python as an implementation language, the Meta-T version of Tetris is a re-creation of the original NES Tetris used at the annual, "Classic Tetris World Championship" (CTWC). Meta-T has been examined by the software expert of CTWC and found to be a faithful version of Classic Tetris up through level 19. Above level 19 there are software bugs in the original Tetris that are perpetuated in the emulation mode used at CTWC. However, as none of the 240 participants in the laboratory study or in the subsequent tournaments in Study 2 has ever reached level 19, these feature differences were not important to this study.

3.1.3 Procedure

Each participant was seated in front of a desktop computer[1], given a Nintendo Entertainment System (NES) controller retrofitted to connect to the computer via USB, and asked to play 50 min of my custom experimental Tetris software Meta-T (Lindstedt & Gray, 2015). After the game, the participant completed a brief exit survey and was debriefed.

During play, the Meta-T software tracks game states and keypresses to the millisecond level with such fidelity that a perfect replay of each player's performance could be played back at actual (or arbitrary) speed. This high fidelity logging allows analysis of the details of moment-to-moment gameplay that goes far beyond simple analyses of overall game scores used in other Tetris studies (e.g., Holmes et al., 2009; Linn & Petersen, 1985; Martin-Gutierrez et al., 2009; Okagaki & Frensch, 1994; Valerie K. Sims & Mayer, 2002; Terlecki et al., 2008). Hence, Meta-T transforms Tetris from the "game as treatment condition" mode used in these prior studies, to a Game-XP (game as experimental paradigm) (Gray, 2017a).

3.1.4 Data reduction and filtering

As a player progresses through a single game of Tetris, the time pressure rises with the difficulty level until even the best players can no longer control the board and they lose. This final level of play necessarily contains the time at which things "went wrong" and the player lost the game, possibly exhibiting panicked, flailing behavior in the process. While this behavior is likely to be quite interesting to examine on its own, it is likely qualitatively different from whatever constitutes "successful" performance for a given player's level of expertise. Additionally, many players abort games early if all is not going exactly to plan. This behavior is more common during the early levels, when the player has little invested in the current game, than it is when the player's current level approaches their highest level achieved. Although I asked players to avoid aborts when they had the urge to "get a fresh start", it seems it is a difficult habit to break for at

[1]The computer was equipped with an SMI eye tracking system. Eye tracking data were collected, but are not part of the current study.

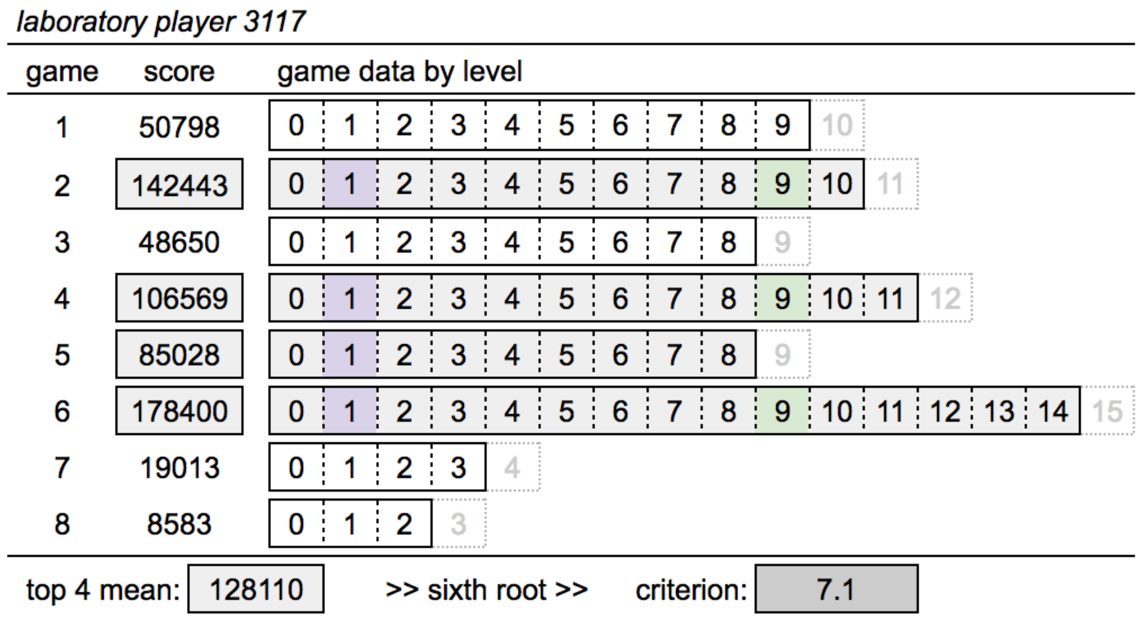


Figure 3.1: An Illustration of the data produced by a single subject (subject ID 3117). Each line represents a single game played, showing both the ultimate score and all game difficulty levels for which data was captured during the course of that game. The four games with the highest scores of the hour-long session are highlighted in gray, and the average of those games' scores and the sixth root "criterion" score are displayed below. Incomplete game levels excluded from analysis are grayed out with dotted borders. Data from game levels 1 and 9 (highlighted) from the best four games are later used for principal component regression models.

least a few players. As a data analysis issue, this aborting behavior is difficult to efficiently and reliably distinguish from both normal play behavior and struggling behavior.

Thus, to side-step this issue, the present investigation uses only data from "completed" game levels, excluding behavior from the final level reached. Before filtering the data, in total the 240 players played 2153 games (mean 9.0 per player), producing a total of 271,843 episodes (mean 1132.7 per player). One player did not ever successfully complete even level 0, so was omitted entirely. After filtering, the resulting data set contains data from 239 players playing 1814 games (mean 7.6 per player) and producing 213,322 episodes (mean 892.6 per player). Refer to Figure 3.1 for an illustration of a given player's data included in the analysis.

3.2 STUDY 1: ANALYSIS

Each participant's gameplay session produced data logs that capture performance from immediate interactions, such as individual keystrokes, to metrics of entire games. The basic unit of analysis for this investigation is data summarized at the level of the *episode*– that is, the time between a zoid appearing, the player deciding where to place it, executing that decision, and ultimately placing the zoid in its final position (whether or not it was the desired destination).

My task was to identify those elements of player behavior that signify the player's expertise. It is worth noting that this initial approach is not confirmatory of any particular theory, but rather to "observe and report" details related to the phenomenon of expert player skill in Tetris. I enter into this analysis trivially expecting to find significant correlations between player behavior and player expertise, but I am most interested in which patterns of behavior differentially signal novice or expert performance.

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

3.2.1 Codifying expertise in Tetris

In the present investigation, "experts" are players who score higher than their peers on performance in the task, and who do so routinely and consistently, not just in one lucky instance. As such, I prefer a measure of expertise derived from an average score to reduce the impact of high outliers. But because Tetris is a sequential task where a series of perfectly executed "good" decisions can ultimately be undone with a single critical error, it is much easier for a skilled player to "crash and burn" than it is for a novice player to secure an abnormally high score. Thus I examine each player's "best consistent performance" by taking the mean of each player's best four game scores achieved in their one hour of laboratory gameplay. This procedure forgives the player's worst games, while considering their four best games as their measure of expertise. Figure 3.1 shows that Player 3117's highest four games scored 142,443, 106,569, 85,028, and 178,400 points. The mean of these four is 128,110 points. Figure 3.3, shows the distribution for each of those "best four" games for each of the 239 players (those who survived difficulty level 0, at minimum). Most notable in the Figure is the pile of overlapping lines and low (under 50,000 points) scoring games at the easier levels of game difficulty (x-axis) with a sparse collection of very good players represented at the higher levels of play.

As players achieve higher levels of game difficulty the speed at which the zoids drop increases and the points received for 1-line clears, 2-line clears, 3-line clears, and 4-line clears (aka "a tetris") are multiplied. The result of this scoring system design is that the distribution of scores is highly positively skewed. In pursuit of a normal distribution of criterion scores measuring player skill, I found the transformation that maximized normality for these criterion scores was a sixth-root transform, with a Shapiro-Wilks test revealing it had the least deviation from normal (W = .992, *p* = .24, see Table 3.1). This transformation is done not with any particular cognitive mechanism in mind, but for statistical convenience. Figure 3.2 shows the distribution of players' criterion scores after a 6th root transform.

Thus, I take the 6th root of the mean of the best 4 games played in 1 hour to be a player's "criterion score", a representation of this player's expertise as comparable to the normally distributed population of their peers. With a measure of player expertise defined, I now examine what measures of behaviors are most distinctly meaningful for the task of Tetris using principal component analysis.

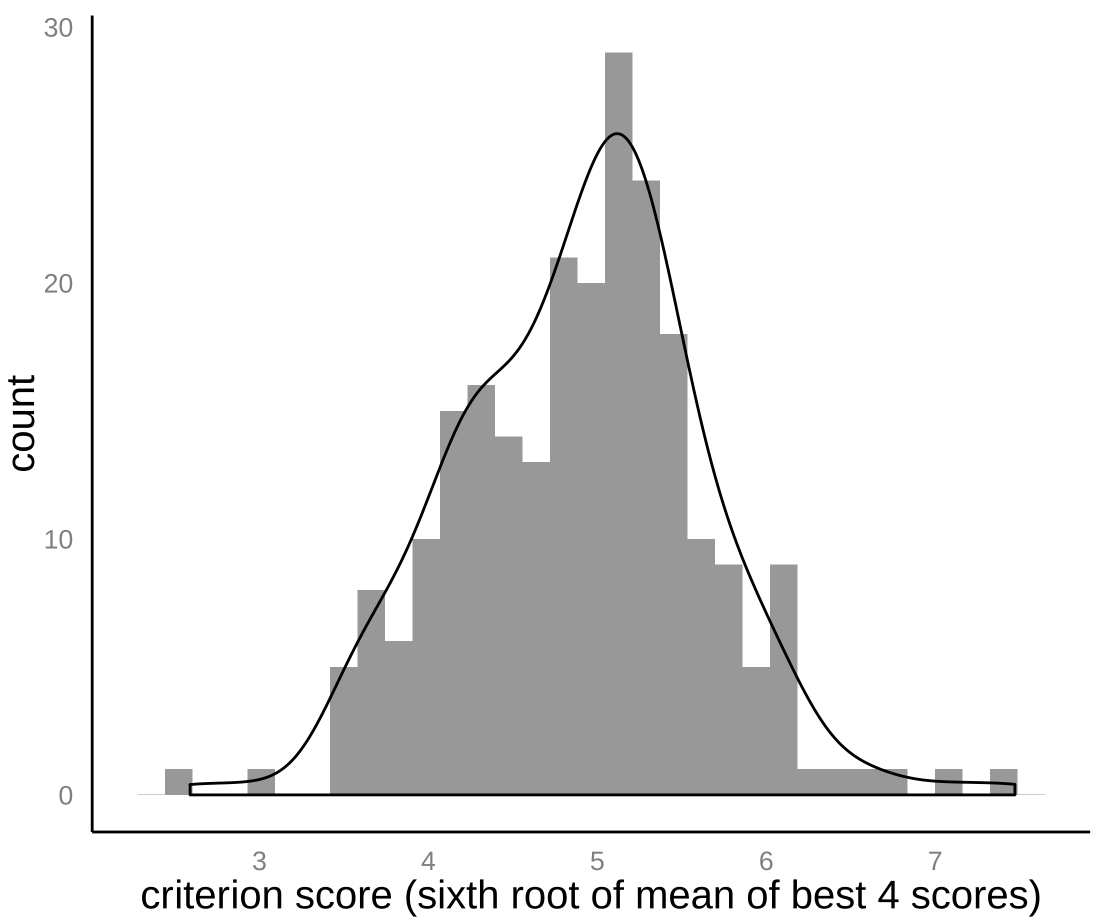


Figure 3.2: Distribution of player skill, normalized by taking the sixth root of each player's "criterion score" (mean of four highest game scores).

Table 3.1: Results of a Shapiro-Wilks test on a variety of different criterion score transformations to measure player skill. Lower W scores and lower p-values indicate stronger deviations from normality.

transformation	W	p
canonical	0.688	< .0001
log	0.979	< .01
square root	0.933	< .0001
cubic root	0.976	< .001
quartic root	0.988	< .05
fifth root	0.991	0.165
sixth root	**0.992**	**0.240**
seventh root	0.992	0.235
eighth root	0.992	0.199
ninth root	0.991	0.158

3.2.2 Task feature decomposition

For each episode of gameplay, I decomposed the Tetris task into an array of measurable features each of which reflects some aspect(s) of player behavior (a complete list of these features is provided in Appendix A). Many of these features are measures of the game state that a player can visually observe and reason over. These features include information about the shape of the piles a player builds and features of the locations in which a player places the falling zoid. Most of these "game state" features are derived from the machine learning literature (C. P. Fahey, 2015; Gabillon et al., 2013; Şimşek et al., 2016; Szita & Lorincz, 2006; Thiery & Scherrer, 2009a), where the focus has been on developing metrics of the "goodness" of each zoid placement.

I interpret many of these features as measuring (a) errors of commission and recovery (e.g., the number of 'pits' present in the game state), (b) risk and reward (e.g., how high the player may allow a board to grow in favor of a high payoff maneuvers), and (c) more abstract elements a player may be sensitive to, such as entropy (e.g., the general disarray of the board).

Machine learning models do not worry about motor times or the effort required to move a zoid to a location. Hence, my features also include measures of how a player maneuvers a zoid to its ultimate destination. These features correspond with more fundamental elements of behavior in most any interactive task, including counts and kinds of keypresses, initial and average response latencies, and measures of efficiency of move execution.

Figure 3.4 illustrates some of the features measured. Appendix A defines the full list of 39 features measured in the current analysis.

This list of features suffers from one important drawback: there is substantial overlap among the definitions of many of these features such that they are highly correlated. Figure 3.5 shows the correlation matrix for this full set of features based on all of the available episode data (minus that which was filtered out). Although it is tempting to point out correlated clusters with the naked eye (e.g., latency features appear to cluster together and correlate), I turn instead to principal component analysis to help group and separate the features most important for capturing variation in player behavior.

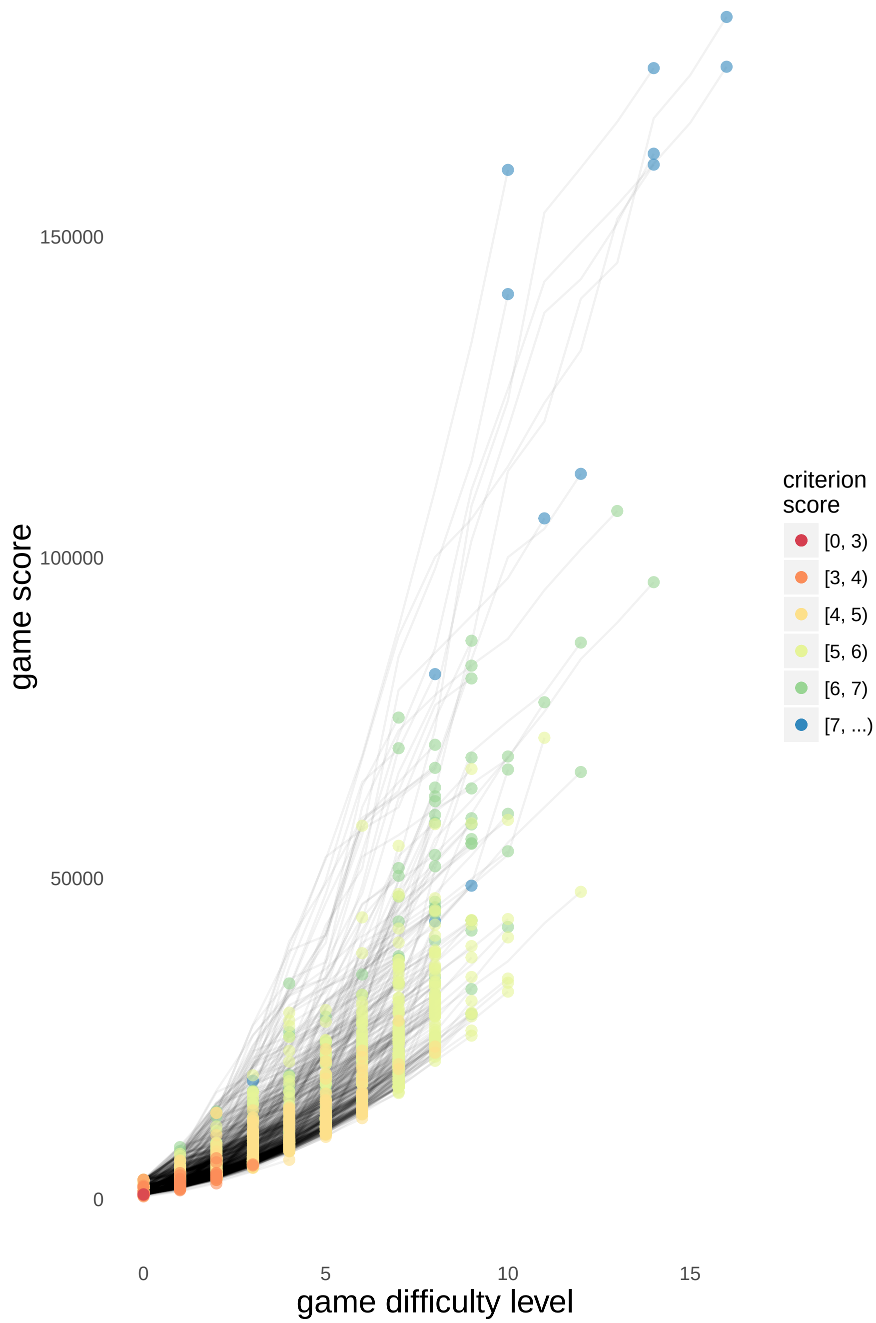
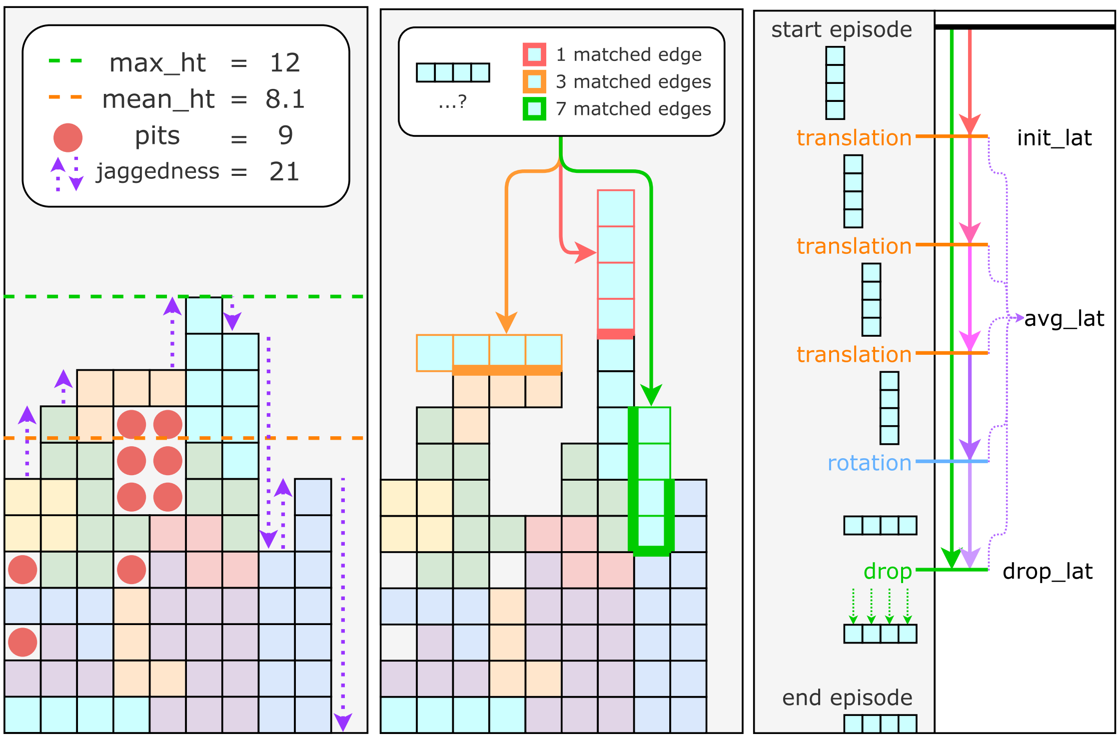


Figure 3.3: Plot of the trajectory of game score over the best four games played by all players, giving an impression of the "shape" of the data available for analysis. Lines represent score progression over level, whereas points represent players' ultimate game scores. Color represents each player's criterion score (calculated later), showing some of the breakdown of score across game levels for players of different skill levels.



(a) (b) (c)

Figure 3.4: Some features of gameplay in Tetris: (a) features of the board state, including the maximum height of the pile (max_ht), the number of unworkable holes buried in the pile (pits – indicated by the red circles), the mean height of all columns in the pile (mean_ht), the "jaggedness" of the pile's perimeter (jaggedness); (b) a feature describing how well a particular zoid-placement fits with the surrounding pile (matches); and (c) the three types of keypress actions in the game (translations, rotations, and drops) and three temporal features of the episode including the initial latency of the first keypress of the episode (init_lat), the mean latency of keypresses throughout an episode (avg_lat), and the time until the player finally (if at all) drops the zoid into the pile (drop_lat).

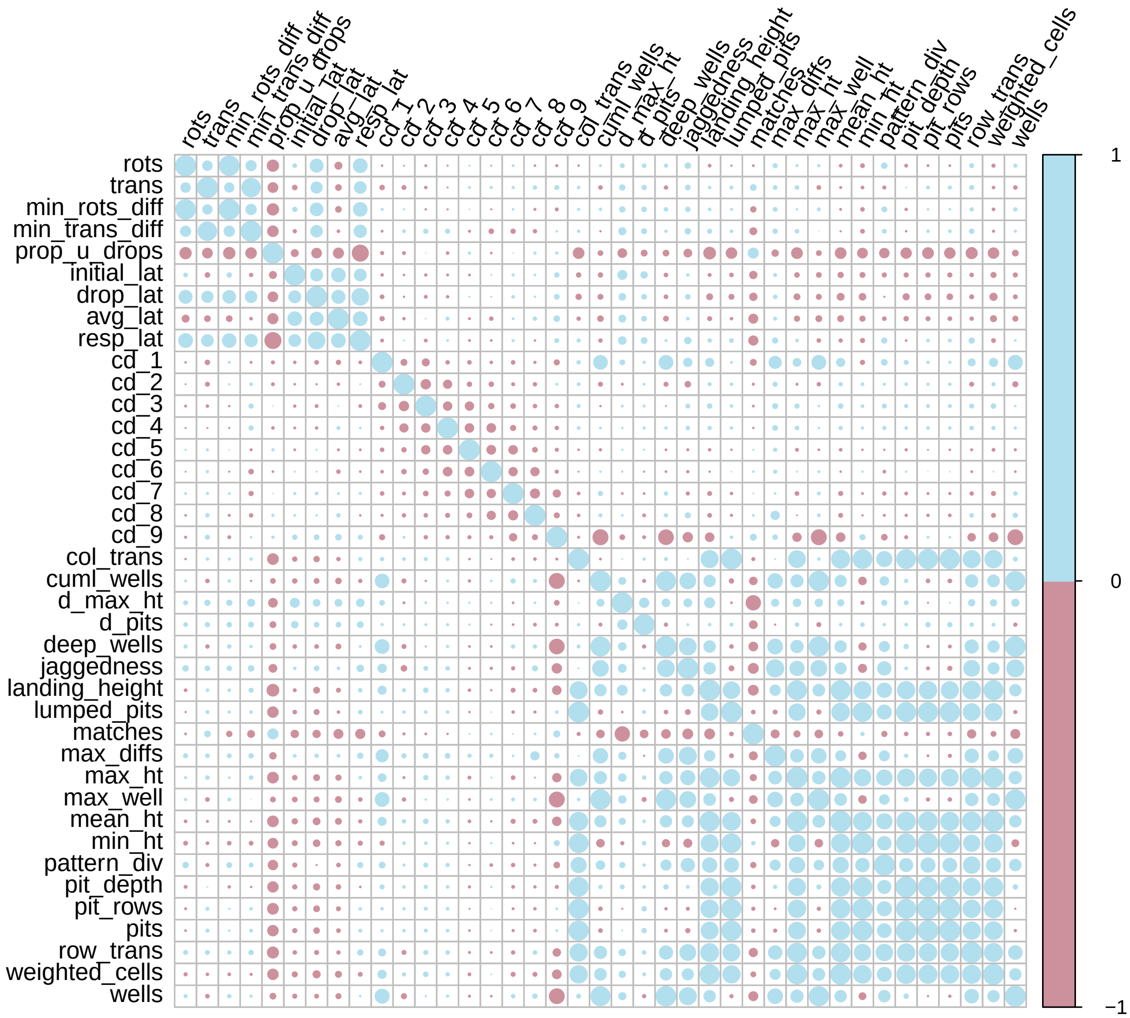


Figure 3.5: Correlation matrix of all features measured across all episodes analyzed. The circle in each cell represents the correlation between two features, with the color indicating the correlation's direction (light blue = positive, reddish = negative) and the size indicating the correlation's magnitude. Notably, all keystroke-level interaction features are on the top-left, while measurements of the structure of the pile are on the bottom right. Note the high degree of correlated clusters of features throughout.

3.2.3 Principal component analysis

Principal component analysis finds a reduced set of orthogonal, latent features – or, components – that weigh the 39 features in various combinations to produce a set of components that capture variations in player behavior. Using this process, I reduce my initial 39 features to 4 components and examine them for more intuitive descriptions of how behavior varies in this task.

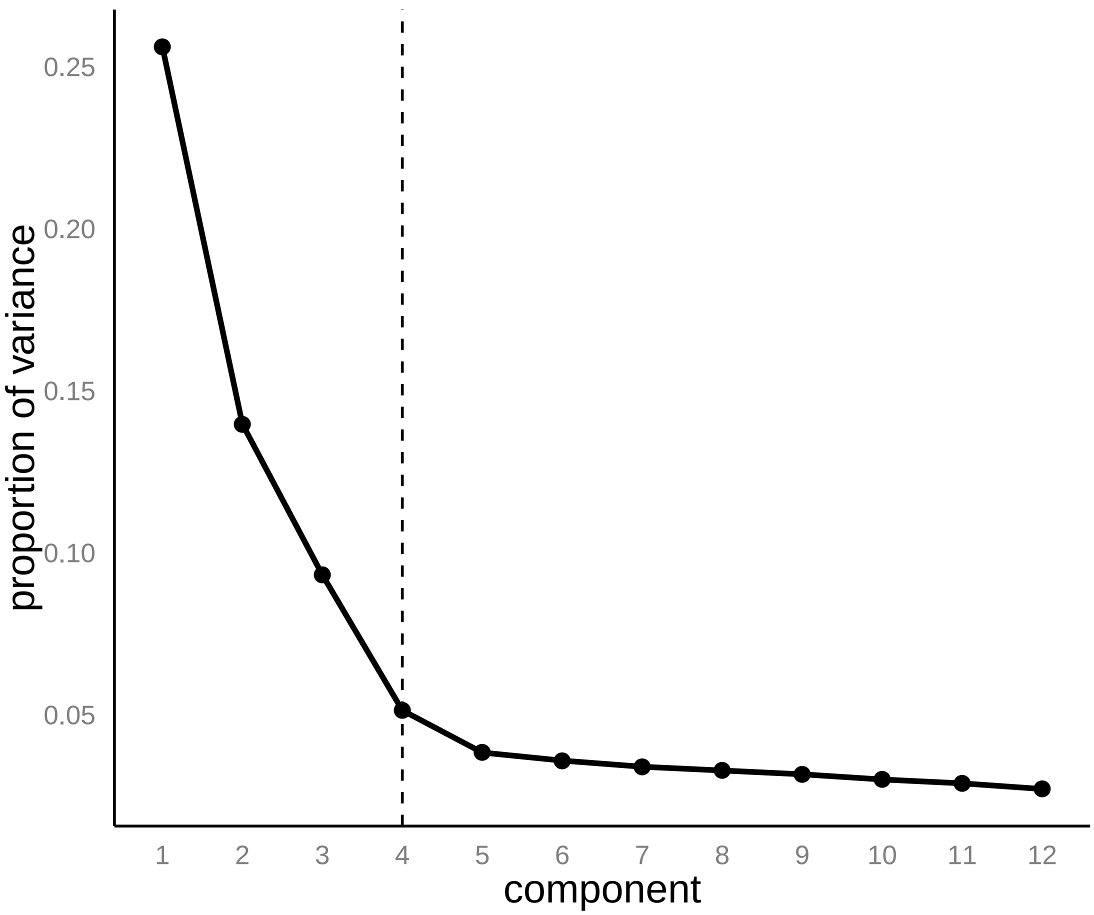


Figure 3.6: Scree plot of the proportion of variance explained by each component. The dashed line represents the "elbow" after which components offer highly diminished returns in terms of variance explained. The fourth component was included with the first three in the analysis.

I ran the principal component analysis on the pre-filtered data set of 213,322 episode-level observations, including all 39 features of the game state described in appendix A. (As discussed earlier and illustrated in Figure 3.1, I did this for each player for each game played, after the pre-filtering removed all data from any level not completed during a game.) Figure 3.6 is a scree plot[2] showing the variance in the data explained by each component. Typically, one searches for the inflection point, or "elbow", in a scree plot to determine how many components to include in further analysis, as the explanatory power gained tapers off with each additional component considered. Here the elbow is at component 4, which suggests I could use only the first 3 components in the analysis. However, because I suspect subtler aspects of player behavior may be important in explaining expert performance, I err on the side of caution and also include component 4.

Table 3.2 shows that the four principal components included in the analysis explained 25.6%, 14.0%, 9.3%, and 5.1% of the variance in the observed data, for a cumulative total of 54.0%. I give a summary of each component below. Table A.1 in Appendix A shows the exact feature loadings for each component.

Table 3.2: Results of the principal component analysis.

	component label	highest loaded feature	% variance explained
1	entropy	+mean_ht	25.6
2	structure	+wells	14.0
3	decide-move-placed	-resp_lat	9.3
4	pivoting/recovery	-avg_lat	5.1
		Total	54.0

3.2.3.1 Component 1, "entropy"

Entropy is associated with larger piles, more holes and disarray, and pile structures that are generally unfavorable to scoring points and surviving longer. A positive value here means the pile is larger and more difficult to work with, while a negative value relates more to empty piles.

3.2.3.2 Component 2, "structure"

Structure is associated with large but orderly piles that leave a deep empty well on the far left or right of the game board– an ideal setup for performing the valuable but risky 4-line clear maneuver (a "Tetris"). Higher values imply more progress toward preparing for a 4-line payoff, while lower values are associated with shorter piles, and perhaps more short-sighted strategies.

[2]"A scree plot displays the proportion of variance explained (synonymous with the eigenvalues of the new basis) associated with a component or factor in descending order versus the number of the component or factor", see (Minitab, 2017).

3.2.3.3 Component 3, "decide-move-placed"

This is the "human execution" component whose features I group into three subcategories. The "decide" subcategory measures choice-response-time on four features and, for a fifth feature, includes the proportion of time in the episode during which the player intentionally "dropped"[3] a zoid. Its "move" subcategory compares the minimum number of four movement features with the actual number used. The "placed" subcategory measures the goodness of the zoid's placement in terms of minimizing the height of the pile and maximizing the number of edges matched (as per Figure 3.4-B).

3.2.3.4 Component 4, "pivoting/recovery"

Pivoting/recovery is associated with very inefficient paths for the zoid, but executed quickly and in service of explicitly clearing lines off the board and uncovering pits. As such, this component appears to be related to quick changes mid-episode and/or rapidly searching the space for a good zoid position when time pressure is high, all for the purpose of improving the state of the pile.[4]

3.2.3.5 Principal component analysis discussion

The 4 components selected in the PCA explain more than 50% of the variance among all observed episodes of the Tetris data. The component explaining the most variance, "entropy", is a general assessment of the disarray in the pile, followed closely by the component related to building orderly piles, "structure". This is unsurprising, as orderliness is a defining feature of the Tetris game wherein one must neatly fill rows to perform well. The third component, "decide-move-placed", captures *quick decisions and efficient movements that result in good zoid placements*. Finally is a component that appears to be related to emergent conditions in the game – "pivoting/recovery". This feature relates to episodes wherein the player has either struggled or failed to find a desirable destination for the zoid, and ultimately altered course mid-stream to place the zoid elsewhere. Together, these are the principle components, the *elements of expertise*, in Tetris.

[3] Zoids will drop by themselves. However, players can hasten drop speed by pressing and holding the drop button. Doing this indicates that the player has decided exactly where they want the zoid to land.

[4] I suspect a similar component would emerge prominently in a future analysis concerning the "failing game levels" that have been trimmed off in the present analysis.

3.3 STUDY 1: RESULTS

With both (1) a measure of each player's level of expertise and (2) a set of meaningful behavioral measures for the Tetris task, I constructed several multiple linear regression models using the 4 principal components to predict player expertise. These models predicted expertise based strictly on each player's execution of the four components, aimed at answering the question "can we predict overall player expertise from small slices of their performance?" Specifically, I asked whether these four components– measures of behavioral process collected at one level of Tetris game difficulty and not full game outcome scores– can predict each player's score-based level of expertise (their "criterion score").

3.3.1 Analyzing by difficulty level

The goal of this analysis is to uncover what aspects of performance distinguish players of different levels of expertise. Ideally I would simply compare subjects of all skill levels under the same game conditions, but doing so is impossible for two reasons: first, the game conditions change as a natural course of the game – the speed of the falling zoid increases, restricting time for the player to decide on a destination and maneuver the zoid to it. It is possible that player behavior will be quite different under the lax time pressure of level 0 (wherein the zoid falls in 16 seconds) as compared to the extreme time pressure of level 9 (the zoid falls in just 2 seconds). Thus I cannot simply average behavior across all difficulty levels.

A second issue is that, due to the progressive nature of the game, novice players lack gameplay data for higher difficulty levels, as they fail in the game before reaching the higher game speeds (indeed, only 27 of the players reached level 9 or higher). Hence, in Tetris, it is impossible to simply compare performance across game levels.

I take a contrastive approach to examining what differentiates players at different levels of game difficulty. I begin by constructing two separate multiple linear regression models: one for level 1 and one for level 9. As Tetris starts at level 0, for level 1 only three of the very least experienced players are absent in the data (see Table 3.3), marking this model as a sort of "model of the masses." An open question is whether my four components capture any differences in level 1 play between players of different skill levels; that is, do experts display their expertise when there is no imminent pressure to do so?

Table 3.3: Game speed and proportion of data at different levels. The time taken (in seconds) for a zoid to fall from the top of the screen to the bottom for a given level of game difficulty. Decisions about where to place the zoid and all maneuvering of the zoid to the proper position must take place in this span of time. Asterisks indicate game levels chosen for constructing the multiple regression models.

Difficulty	Time to	Players with data	
Level	Impact	Count	Percentage
0	16.00 s	239	99.5%
*1	14.33 s	236	98.3%
2	12.67 s	225	93.8%
3	11.00 s	211	87.9%
4	9.33 s	189	78.8%
5	7.67 s	158	65.8%
6	6.00 s	139	57.9%
7	4.33 s	106	44.2%
8	2.67 s	67	27.9%
*9	2.00 s	27	11.3%
10-12	1.67 s	8	5.8%
13-15	1.33 s	3	2.5%
16-18	1.00 s	1	0.4%
19-28	0.67 s	0	0.0%
29+	0.33 s	0	0.0%

The level 9 model shifts the focus to only those 27 players who are skilled enough to survive at higher game speeds, highlighting what distinguishes the very best players from one another – the "best of the best" model.

I conclude my analyses by constructing a level 1 model that includes only the 27 players who made it to level 9 or beyond. As the level 9 model will show, these players do appear to differ from each other. The question I ask in constructing a level 1 model of these players is whether those differences are manifested during what is, for these players, an easy and relaxed level of play.

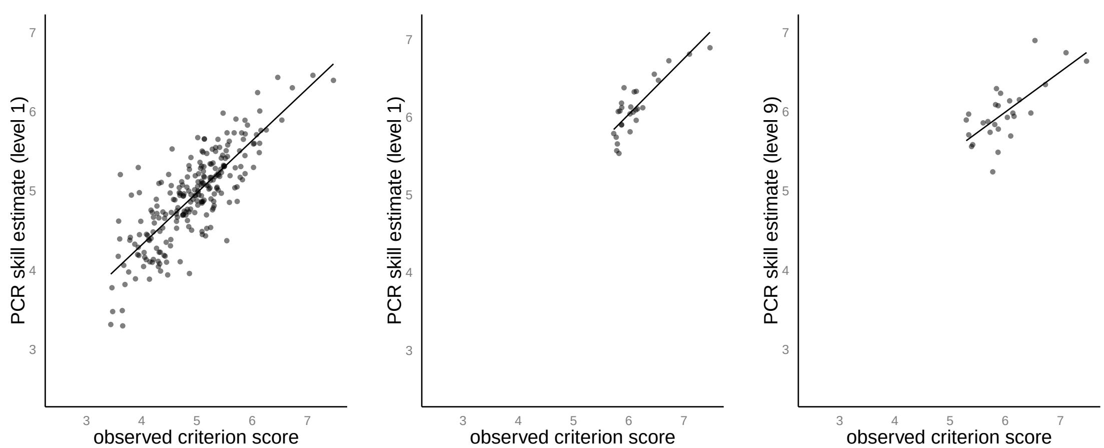
3.3.2 Integrating components with player data

Each player who participated in the study provided one hour's worth of play during which they completed as many games of Tetris as possible. To standardize the number of games per player and focus the analysis on each player's "best performance", I begin by selecting each player's best four games for analysis. This also presumably minimizes the number of games that were aborted early.

Next, I took the mean of a player's component scores at each difficulty level completed (across games). By reference to Figure 3.1, for player 3117, I took the mean of games 2, 4, 5, and 6 for level 1, but only games 2, 4, and 6 for level 9. (This follows as player 3117 lost the game at level 9 for game 5 and I am only using data for levels that were successfully completed). Hence, the data used for player 3117 is based on fewer games than that of a player who completed level 9 on each of his/her four criterion games.

3.3.3 Principal component regression results

For both game difficulty levels 1 and 9 I calculated a principal component regression (PCR) to predict player expertise (criterion score) based on the array of 4 principal components of behavior calculated earlier (entropy, structure, decide-move-placed, pivoting/recovery). Each model was then run through a bidirectional stepwise model selection process based on minimum Akaike Information Criterion (AIC) to eliminate superfluous components in the model. The left two columns of Table 3.4 show the results of the multiple linear regression models for both difficulty levels.



(a) (b) (c)

Figure 3.7: Fitted values of principal component multiple regression models for the two game levels examined.

For the Level 1 model, a significant regression equation was found (F(4,231) = 110.9, $p <$ .0001) with an adjusted R^2 of 0.652. Players' predicted expertise level (criterion score) is equal to 5.01 - 0.12 (entropy score) + 0.22 (structure score) + 0.49 (decide-move-placed score) - 0.10 (pivoting/recovery score). The "entropy", "structure", and "decide-move-placed" components were significant predictors of player expertise at level 1, while the "pivoting/recovery" component showed near-alpha significance ($p = .054$). None of the components were eliminated during model selection.

Table 3.4: Results of each of two multiple linear regression models constructed using data from difficulty level 1 and 9 using 4 principal component scores for each of 236 participants to predict criterion score (player expertise).

	Level 1			Level 9			Level 1 (best 27)		
Adj. R^2	0.652			0.470			0.681		
df	(4 , 231)			(2 , 24)			(3 , 23)		
F	110.9			12.9			19.5		
p	<0.0001			<0.001			<0.0001		
factors	coeff.	sign.		coeff.	sign.		coeff.	sign.	
(Intercept)	5.01	<0.0001	***	5.65	<0.0001	***	5.57	<0.0001	***
entropy	-0.12	<0.0001	***	0.13	0.058	.	–	–	
structure	0.22	<0.0001	***	–	–		0.11	<0.05	*
decide-move-placed	0.49	<0.0001	***	1.11	<0.001	***	0.67	<0.0001	***
pivoting/recovery	-0.10	0.054	.	–	–		-0.30	0.063	.

For the Level 9 model, a significant regression equation was found (F(2,24) = 12.9, p < .001) with an adjusted R^2 of 0.470. Players' predicted expertise level (criterion score) is equal to 5.65 + 0.13 ("entropy" score) + 1.11 ("decide-move-placed" score). Only the "decide-move-placed" component was a significant predictor of player expertise at level 9, while the "entropy" component was near-alpha ($p = .058$). The "structure" and "pivoting/recovery" components were eliminated from this model.

To investigate whether the "best of the best" players behavior was differentiable under low time pressure conditions, I also constructed a multiple regression using level 1 data for only those 27 best performing players (those who successfully completed level 9 at least once). A significant regression equation was found (F(3,23) = 19.5, p < .0001) with an adjusted R^2 of 0.681. Players' predicted expertise level (criterion score) is equal to 5.57 + 0.11 ("structure" score) + 0.67 ("decide-move-placed" score) - 0.30 ("pivoting/recovery" score). Both the "structure" and "decide-move-placed" components were significant predictors of player expertise at level 1 for the 27 best players, while the "pivoting/recovery" component was near-alpha ($p = .063$). The "entropy" component was eliminated from this model. Figure 3.7 shows the fit of each model's predictions to its training data.

3.4 STUDY 1: DISCUSSION

3.4.1 All players, level 1 model

For the level 1 model with 236 players (those who completed level 1 at least once), the component scores successfully predict differences in player expertise. Even though level 1 is not especially challenging, the component analysis successfully differentiates between low and high skilled players.

3.4.2 27 best players, level 9 model

The 27 players, Level 9 model (all who completed level 9 at least once) differentiates among the relatively small number of players with data at this high difficulty level. Importantly, the only significant component score in the model is the decide-move-placed score. This component is the human-execution factor whose 11 features fall into three categories: decision efficiency (choice-reaction-time and drop probability), efficient movements (minimizing the number of rotations and transpositions used to move the zoid to the targeted location), and goodness of the zoid's placement (i.e., placement of the zoid so as to maximize the number of edges matched to the piece and to lower the overall height of the board).

3.4.3 27 best players, level 1 model

I ran this model because I wanted to know whether I could differentiate among the 27 best players based on their performance at level 1. After all, a naive perspective would be that although this group could not help being better on my metrics at level one than the lesser skilled players, the skill differences among this top group might not show through at this level of play. The fact that there are individual differences among this group at level 1 play suggests two conflicting hypotheses: (a) the best players are intentionally rushed or hasty, versus (b) the best players are quick but not rushed, or at a deliberate speed but not hasty. The former suggests that players are executing movements as fast as possible perhaps to warm up or practice. The latter suggests that from the players' perspective, they are not feeling rushed or hasty but are simply making decisions with their acquired efficiency and precision that results in faster decisions, more efficient movements, and more effective placements.

Although I list the factors of the decide-move-placed component in the serial order of (a) decide, (b) move, and (c) placed, I believe this component is more complex than this ordering suggests. For example, with increasing expertise I believe that the ability to spot the best placement comes first. Hence, the end of the "decide" phase (which is marked by the first keypress for a transposition, rotation, or drop) signals that the player has decided not simply to "move" the zoid but fully "where best to place the zoid". This location decision also implies that the player has evaluated the board to determine an open path exists to that location. I will have more to say about this hypothesis in the Study 1 and 2 combined discussion section later on.

3.5 STUDY 1: SUMMARY AND CONCLUSION

These analyses show how the most general components, those that emerged from the dataset of 236 players, are differentially important at discriminating among players at different levels of expertise. Hence, only the decide-move-placed factor remains discriminative among the 27 players represented by the level 9 analyses. Isolating these 27 players to look at their level 1 performance shows that this decide-move-placed factor discriminates among these 27 even when they are playing under low pressure conditions.

CHAPTER 4
STUDY 2: EXPERTISE IN TETRIS TOURNAMENTS

4.1 STUDY 2: APPROACH

This study asks whether findings from the previous study can be used to predict success in a more naturalistic setting – in this case, real-world Tetris tournaments.

4.2 STUDY 2: METHOD

Typically, to validate statistical models of player expertise one would employ methods such as leave one out (LOO) cross-validation or splitting the data into training and test sets. While such paths of analysis are well traveled and more than adequate, I am in a unique position to perform two somewhat unusual methods of model validation. The first increases the *external validity* (Gray & Salzman, 1998, p. 217) [see also, (Cook & Campbell, 1979)] of my PCA model by generalizing my conclusions across persons, settings, and times; namely, (a) to a new dataset of players (persons), (b) who were playing in the "qualifying rounds" of a tournament and not "laboratory" conditions (places), and (c) in some cases, were playing years before (and even sometimes after) the laboratory data collection (times). The second validation uses the same components as the first (the PCR factors for each individual for each of the 4 components) and, hence, for the same reasons, also increases the external validity of my findings. However, the second validation goes beyond that to increase the *effect construct validity* (Gray & Salzman, 1998, pp. 213-215); that is, the effect construct for the first method discussed in this section is the same construct used in developing it in Study 1; namely, the PCR adjusted R^2 of player criterion score with the components. Here I use a different effect construct; namely, can my PCA components be used to predict tournament winners?

4.2.1 Procedure

In conjunction with the CogWorks laborators, I hosted Tetris tournaments using my Meta-T software (Lindstedt & Gray, 2015) at Rensselaer Polytechnic Institute's annual Genericon conventions in 2014, 2015, and 2016. Genericon's attendees consist of a mix of RPI students, local area residents, and some fans flown in from longer distances. Entrance into each tournament was free, such that even low-skill players could could be encouraged to participate and contribute (after all, “it's for science!”). Cash prizes were awarded in the amount of $300, $200, and $150 for 1st, 2nd, and 3rd place tournament winners respectively, to attract skilled players to come to compete. Data collection was approved by the Rensselaer IRB in an IRB proposal written specifically for the tournaments.

Each tournament consisted of a qualifying round and a single-elimination tournament bracket. In the qualifying round, players completed two games to the best of their ability. To promote both control of the random nature of the task environment and fairness of competition, the order in which the sequence of zoids appeared in each qualifying game was fixed such that all players would experience the same two lists of zoids, though the order in which those sequences were presented to each player was randomly selected. In total, 33 players entered the 2014 tournament, 43 entered the 2015 tournament, and 15 entered the 2016 tournament, for a total of 91 entrants across all three years. Some entrants were repeat visitors, and as such I used only their most recent data.

Each player's highest qualifying score (HQS) was used to determine who would take part in the tournament playoffs. The top scoring 8 players in the qualifying round were invited to participate in a single-elimination Swiss bracket style tournament. In this style of tournament, each player is paired with an opponent based on their qualifying round ranking, but arranged to avoid forcing highly ranked players to eliminate one another early; i.e., the highest ranking player starts by playing the lowest. This arrangement ensures the match-ups are fair, but also carries the potential for exciting upsets; that is, the 2nd ranked player is likely to face lesser skilled opponents and thus more likely to proceed to the finals, but can still potentially lose to the 7th ranked player, which would be considered a major upset.

The tournament consisted of 3 rounds: quarterfinals, in which all 8 players competed; semifinals, in which the 4 winners of the quarterfinals competed; and the finals, where the two best players compete for the grand prize. To determine the 3rd place prize, the two players who lost in the semifinals played a runoff game.

During a tournament match, both players' game screens were projected onto a blank wall for audience members to watch. The winner of a match was the player who achieved the highest score – thus, even if a player lost their own game early, they could still win if their opponent did not surpass the score they achieved. Similar to the qualifying round, the sequential order of zoids presented in each round of the tournament was fixed, such that all quarter-finalists faced the same sequence of zoids. The same was true of all players in the semifinals, the finals, and the runoffs, with a different zoid sequence selected for each round of players.

4.2.2 Data selection

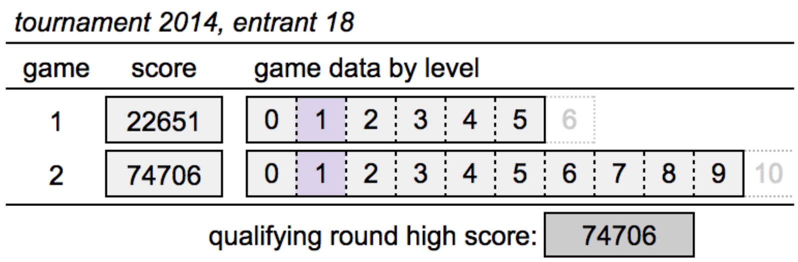


Figure 4.1: Illustration of data collected for tournament player 18 in the 2014 tournament, similar to Laboratory player 3117 presented earlier. The highlighted game score represents the player's qualifying round high score (QHS). Dimmed game levels were incomplete and omitted from analysis. The shaded purple region indicates the small slice of behavioral data used by the Level 1 Principal Component Regression model to estimate player skill.

In common with the laboratory analyses reported above, for both tournament analyses, I excluded the final incomplete level of each player's game and then extracted level 1 data from each tournament entrant's two qualifying round games. Then all four component scores were computed and all values were averaged per player such that each player was represented by one four-dimensional behavioral data point for the purposes of estimating their skill using the level 1 PCR model. Figure 4.1 illustrates the available data for a sample tournament entrant.

4.3 STUDY 2: RESULTS

I first conducted a traditional validation of the PCA models but extended the external validity of these models by using a new dataset of players (persons) who were playing under "tournament" not "laboratory" conditions (places) and in some cases, were playing years before (or after) the laboratory data collection (times). I then use the models to predict the winners from three years of Tetris Tournaments. This increases the *effect construct validity* of my findings.

4.3.1 Increasing external validity: PCR versus qualifying rounds

I predicted each player's individual expertise level based only on their level 1 behavioral components, using the exact same components and PCR model developed in the laboratory in Study 1. I then ran a simple linear regression examining whether the PCR skill estimates correlated with player performance in the qualifying round. The resulting model had an adjusted R^2 of 0.544 ($F(1,88) = 107.3$, $p < .000001$), showing a relationship between the level 1 model's skill estimates and players' game performance during the qualifying round. Figure 4.2 illustrates this relationship. This finding compares favorably to the in-laboratory test reported above which found an adjusted R^2 of 0.652.

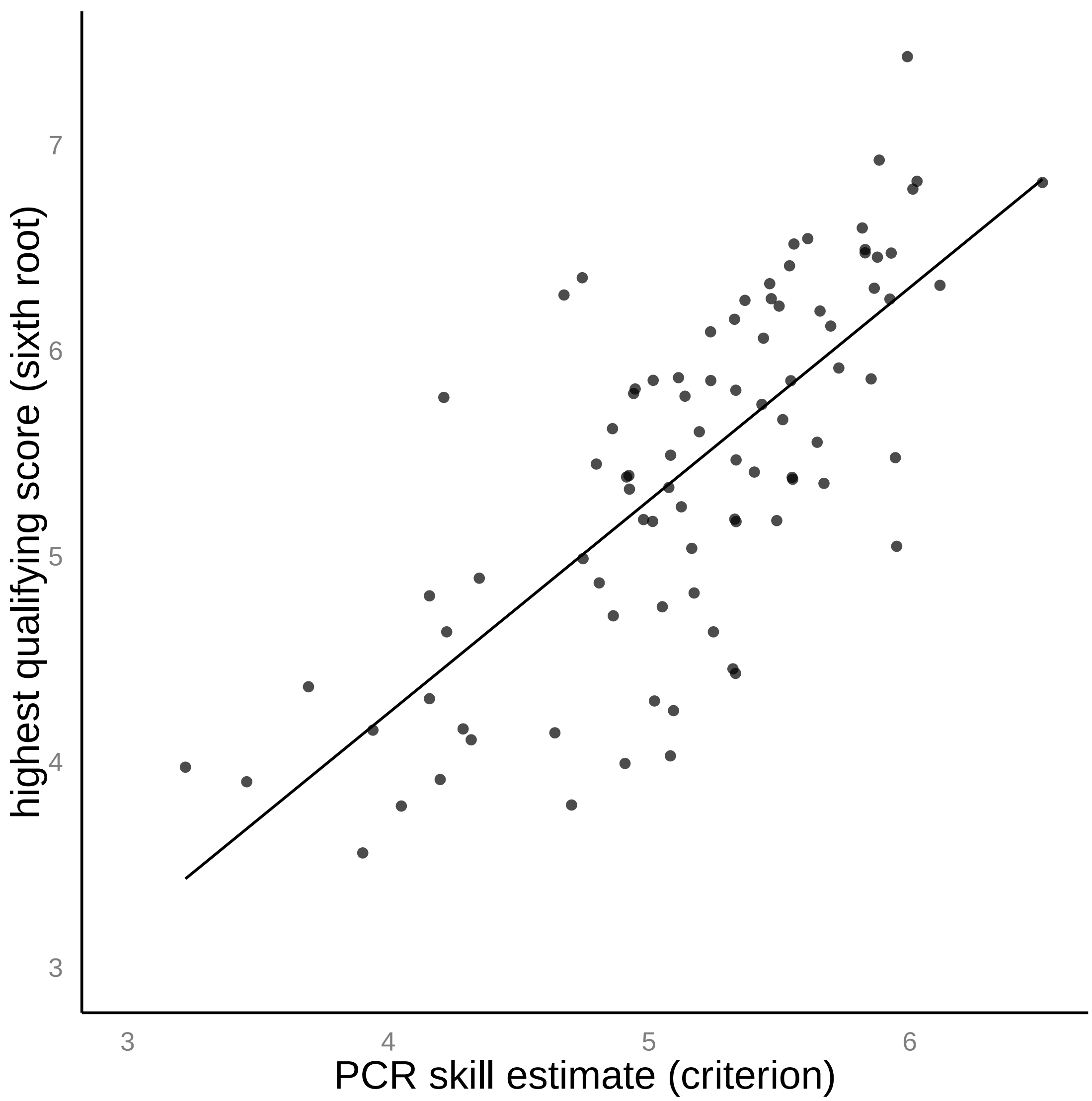


Figure 4.2: PCR predictions. R^2 = .544. Correlation between qualifying round high score and principal component regression (PCR) skill estimates based on level 1 player data during those games of the qualifying round.

4.3.2 Increasing external and effect construct validity: PCR versus tournament matches

To extend the construct validity of the PCA model, I compare each player's PCR skill estimate with their playoff opponent. Across all rounds of all three tournaments, the PCR skill estimates based on level 1 performance in the qualifying round were accurate in predicting 83.3% of all match winners during the tournament, which was significantly more than chance (OR = 5.0, 95% confidence interval (CI) = [1.311, 19.075], $p < .05$).[1] Hence, this finding supports the construct validity of the PCR skill estimate.

I then go beyond validation to compare how well betting using the PCR estimate would compare with a better using each player's qualifying round high score (as illustrated in Figure 4.1). Across all rounds of all three tournaments, the qualifying round score was accurate for predicting the winner of 58.3% of all matches, which did not differ significantly from chance (OR = 1.4, 95% confidence interval (CI) = [0.448, 4.376], $p = .563$). Thus, it appears the PCR skill estimate (based on only level 1 data) is better at predicting tournament winners than the players' actual qualifying round scores used to determine the tournament bracket.

Table 4.1 compares the qualifying score's winner predictions to the PCR skill estimates' predictions for each tournament round. Notably, the PCR skill estimates correctly predicted every winner of the runoffs and finals (100%), whereas the qualifying round score only predicted one third of those winners (33%), implying the PCR estimate performed quite well at distinguishing the "best from the best".

Table 4.1: Predictions of the outcomes of tournament elimination matches made using QHS and PCR scores to assess player skill. The table shows the accuracy of each type of skill assessment for each round tournament elimination matches, and highlights the PCR model's sensitivity when predicting whether matches will result in upsets.

tournament round	N games	qualifier accuracy	PCR model accuracy	upsets N	upsets predicted
quarterfinals	12	58.3%	83.3%	5	60.0%
semifinals	6	83.3%	66.7%	1	0.0%
runoffs	3	33.3%	100.0%	2	100.0%
finals	3	33.3%	100.0%	2	100.0%
total	**24**	**58.3%**	**83.3%**	**10**	**70.0%**

[1] Odds ratio was computed using MedCalc's simple online calculator. https://www.medcalc.org/calc/odds_ratio.php

4.4 STUDY 2: DISCUSSION

I validated the level-1 PCR model's ability to assess Tetris players' expertise twice over. The model did well in predicting both: (1) performance during the qualifying rounds, and (2) the winners of individual tournament matches. Both of these predictions extend the external validity of my findings by using a new dataset of players (persons), who were playing under "tournament" not "laboratory" conditions (places), and in some cases, were playing years before (or after) the laboratory data collection (times). I then used my models to predict the winners from three years of Tetris Tournaments. This increased the *effect construct validity* of my findings.

The model's success is especially exciting as these predictions are based only on data collected during level 1 of game play, long before it is clear what the final scores for those games will be, implying that something more fundamental to player ability is being measured. The PCR model does well at predicting qualifying round performance (adjusted R^2 of 0.544), somewhat less than on the laboratory data on which it was trained (adjusted R^2 of 0.652). Where the PCR model shines is in predicting the tournament outcomes. Indeed, as is typical in Tetris and other tournaments, I assigned players to tournament matches based on their highest qualifying round score with the general expectation that the player with the highest qualifying round score would win. As column four of Table 4.1 shows, the model predicted more than two thirds of all upsets, and *all* of the upsets during the tournament finals. Compared to the qualifying round score rankings, the PCR model's skill estimates are superior at picking out the winners of individual matches.

4.5 STUDIES 1 AND 2: COMBINED DISCUSSION

My primary finding is that various elements of moment-to-moment behavior in the complex real-time, decision making task of Tetris are able to predict player expertise. The common sense interpretation of these components ranges from measures of how different players build stacks of zoids to how deftly players execute keypress maneuvers. I validated the model twice.

First, rather than dividing the lab data into a development data set and a test data set, I used all of the lab data from 236 players (see Table 3.3 for the count of players by level) as the development set and used data collected during the qualifying rounds of Tetris tournaments as the test data set. Hence, the players and the settings of these two venues were very different.[2] The model did

[2]The lab players were individually run in sound isolation pods using volunteers from the Rensselaer subject pool. The tournaments attracted players who were not part of the development set. The tournament data were collected in a large and noisy room, with up to 3 other people also attempting to qualify at the same time. These "others" sat at adjacent

well both in extrapolating to new people, places and times (the classic validation method), and in predicting tournament match winners (the novel validation method). This increased the *effect construct validity* of my findings.

The model's success is especially exciting as these predictions are based only on data collected during level 1 of game play, long before it is clear what the final scores for those games will be. Interestingly and essential to my continuing investigation, my models are capturing something about the nature of Tetris expertise that is more predictive of performance than the scores themselves.

4.5.1 Components of excellence

The PCA yielded four components which I named entropy, structure, decide-move-placed, and pivoting/recovery. *Entropy* measures the orderliness or disorderliness of the pile and is negatively associated with expertise, with the better players scoring lower on this component.

Structure measures orderliness and especially captures deep, empty wells on the extreme left or extreme right columns. These wells are, of course, areas where an I-beam zoid could be placed, thereby clearing multiple lines at once and gaining the points bonus that comes from that maneuver. For the 27 best players in Study 1, structure does not appear in the level 9 model but is significant in the level 1 model. The number of tetrises (clearing 4 lines at once) declines sharply for even the best players at higher levels. Presumably the disappearance of structure for the 27 level 9 players reflects the disappearance of this maneuver.

Pivoting/recovery is associated with inefficient paths, executed quickly, in service of clearing lines off the board and uncovering pits. In terms of evidence for its efficacy, this is the weakest component. It does not appear at all for the level 9 analysis and has non-significant p-values for all level 1 data as well as for level 1 play of the 27 best players. Although it seems tempting to make up a story to explain it, I leave it here as a mystery requiring further analysis.

The component that I believe is most associated with expertise is *decide-move-placed*. This component differentiates the best players from everyone else when the level one performance of the 236 players is considered. It continues to differentiate the best 27 players from each other at level 9. Indeed, somewhat to the surprise, when I looked at just the level 1 data for these 27 players, decide-move-placed even differentiated these players from each other under the very slow speeds of level 1.

computers, while a small crowd of 20-30 onlookers came in and out to watch, at times both encouraging and distracting the players with their comments and advice.

These faster rates for expert players at early levels of play are seen in other video games such Starcraft 2™ (Huang, Yan, Cheung, Nagapan, & Zimmermann, 2017; Thompson, McColeman, Stepanova, & Blair, 2017). For the Tetris players, I suggest that the experts are “quick but not rushed” or “working at a deliberate speed but not hasty”. They are not feeling rushed or hasty but are simply making decisions with an efficiency and precision that results in faster decisions, more efficient movements, and more effective placements. This perspective also suggests a decision sequence that is not necessarily implied by the temporal ordering of decide, move, and placed. I hypothesize that the end of the “decide” phase (which is marked by a keypress for a transposition, rotation, or drop) signals that the player has decided not simply to “move” the zoid but “where best to place the zoid”. This location decision also implies that the player has evaluated the board to determine an open path exists to that location. Hence, by this hypothesis the “decide” is not simply a decision to drop, rotate, or transpose the zoid, but a decision on where to place it.

4.5.2 Unraveling expertise

I see the elements of expertise for Tetris as composed of both perceptual and cognitive elements with the distinction between these two being more notional than real. My past work (Sibert et al., 2017) established that what expert Tetris players “see” is based on the nature of the rewards they are seeking. That is, unlike machine learning players who tend to optimize the number of lines cleared, human players tend to optimize score so that the set of feature weights that best account for their zoid placements are those that emphasize clearing multiple lines at once, as opposed to the slow but steady single line clears of the machine learners. For humans, this implies a large role for perceptual learning (Goldstone, 1998; Kellman & Massey, 2013) or perceptual expertise (Chase & Simon, 1973, 1973; Reingold & Sheridan, 2011) which, as with de Groot’s chess players (de Groot, 1965) or radiologists looking for tumors (Reingold & Sheridan, 2011), enables experts to immediately fixate on locations most likely to advance the game or most likely to contain evidence of a tumor. In the current study, perceptual expertise is marked in the decide-move-placed component by the experts’ seemingly instant ability to locate places where the zoid placement will minimize the height of the pile while maximizing the number of edges matched.

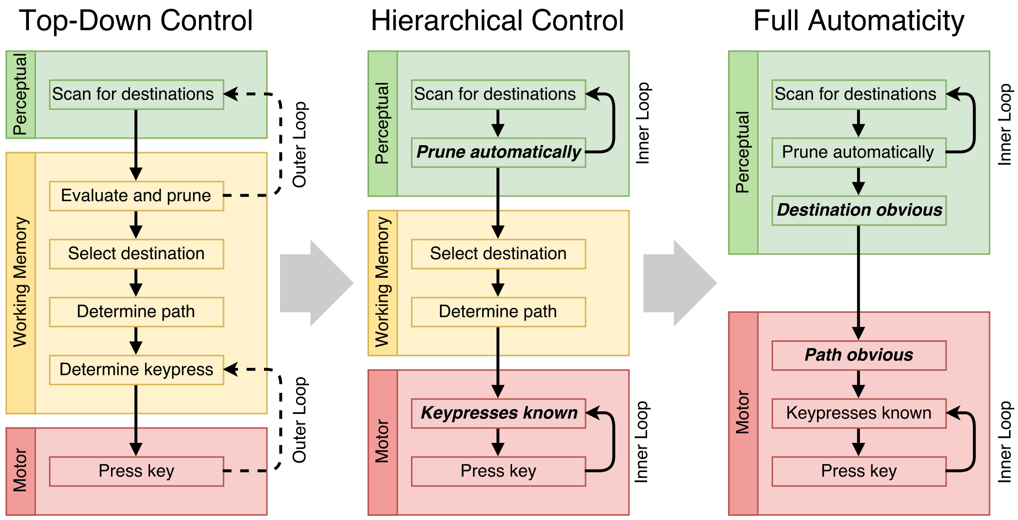


Figure 4.3: Notional figure capturing the transition from top-down control to hierarchical control and ultimately full automaticity for skilled Tetris players. I see the transition from novice to expert as entailing a transfer of control from top-down processes which require repeatedly consulting and updating working memory (outer loops) to more hierarchical control, wherein repetitive task elements are shifted to the more relevant systems (inner loops) allowing working memory to serve primarily as a coordinator between systems. With sufficient practice, some subtasks may be nearly completely automatized, freeing working memory up nearly entirely to focus on other subtasks. (Created using draw.io.)

I also see choice reaction times (Card, Moran, & Newell, 1983, pp. 71-76) minimizing as expertise increases. This conclusion is supported by recent findings from Logan et al. (2016) in skilled typing showing that the Hick's Law coefficient for choice (Hick, 1952) goes effectively to zero in skilled typists. For Tetris, the best players minimize their response times to drop, rotate, and transpose a zoid. They also minimize the number of keypresses required rotate and transpose zoids to their final resting locations. These factors suggest some combination of rapid decision making for each action in the sequence of rotations, transpositions, and drops as well as for identifying the optimal resting location for the current zoid. In terms of a notional, non-mathematical, non-computational model, that captures the changes in cognition, perception, and action across levels of Tetris expertise I present Figure 4.3.[3] This figure walks through the process of transitioning between stages of skill acquisition, from a more deliberate, cognition-heavy approach, to a more automated, perceptual-motor approach.

[3]The attentive reader will note that I have borrowed the style, while modifying some of the substance, of the flow chart provided by Logan et al. (2016) for skilled typing. My elaborations are meant to keep my formalism simple while better fitting the demands of Tetris for perceptual learning and placement of the current zoid into an optimal board location.

4.5.3 Still better skill exists

I have found the "decide-move-placed" component to be a critical factor in determining player skill for the current dataset which spans those players who cannot escape difficulty level 1 all the way through those who capably exceed difficulty level 9 (and some much beyond). But Figure 4.4 shows that there is likely to be more to this picture: even the very best players struggle to execute the high-scoring, but risky, 4-line clear maneuver when the game's time pressure increases. If one compares these players' line-clearing patterns to video of the players in the final round of the 2016 *Classic Tetris World Championship* (Classic Tetris, 2016), one sees players who are not only capable of continuing to perform these 4-line clears late into the game, they are able to do so while reaching the so-called "kill screen" at level 29 where the game's speed begins to exceed the very limits of human capability. There is clearly much more to discover about the space of expertise charted by these extreme expert Tetris players.

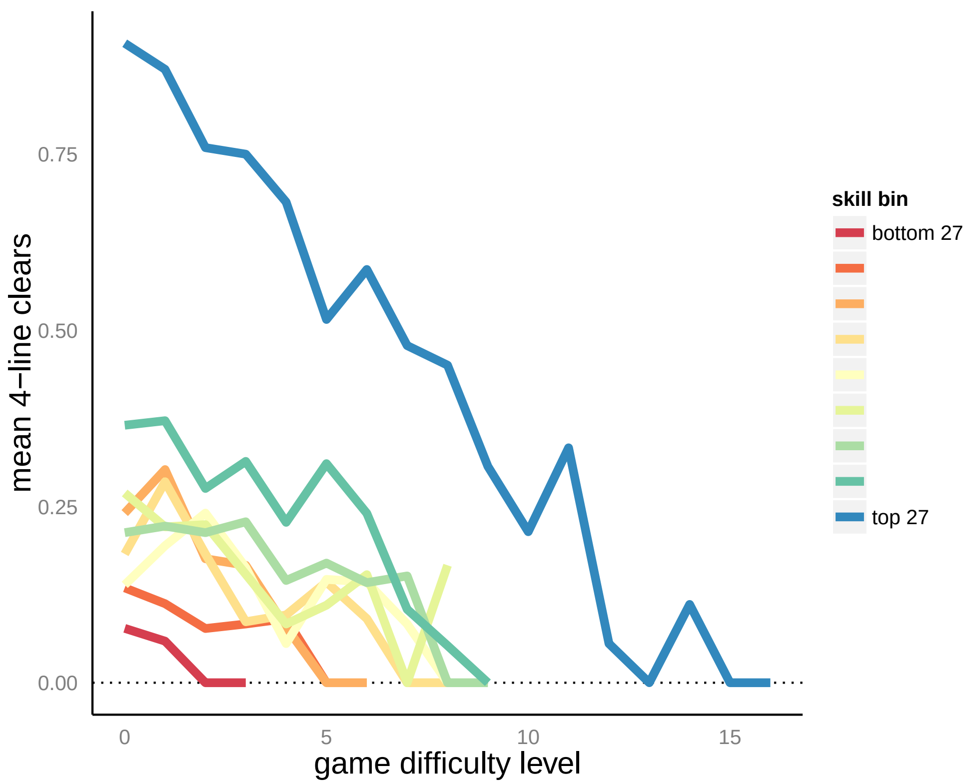


Figure 4.4: Shows the drop-off in number of 4-line clears (y-axis) as time pressure increases (game level, x-axis), even for the highest performing players.

4.6 STUDIES 1 AND 2: SUMMARY AND CONCLUSIONS

I used Tetris as an experimental paradigm (or Game-XP) to study dynamic, sequential decision making in a manageable but complex task environment, sampling Tetris expertise by collecting 1 hour of play data from each of 240 college students. After filtering, the resulting data set contained 239 players, 1814 games (mean 7.6 per player) and produced 213,322 episodes in which one player, placed one zoid on the Tetris board. These data were subjected to a principle component analysis (PCA) that yielded 4 interpretable components. The external validity of the PCA analyze was enhanced by applying the four components to a new set of data collected during tournament play across three Tetris tournaments and its construct validity was enhanced by using the PCA measures to predict the tournament winners.

These analyses provide clear evidence for integrating complex cognitive, perceptual, and motor processes that result in real-time, sequential-decision-making in Tetris. Up to the limits of their expertise, experts are quick but not rushed and move at a deliberate speed but are not hasty. I see this finding as an important step in bringing the tools and theories of cognitive psychology to bear in the analyses of a complex, cognitive task performed by millions of human players of widely different levels of expertise. This work has benefited from prior work on Tetris especially the feature modeling work of Sibert et al. (2017) and the sequential decision-making work of Şimşek et al. (2016).

In the next study, I gather data from players more expert than those in the college student population– the top contenders in the annual Classic Tetris World Championship.

CHAPTER 5

STUDY 3: COGNITIVE STRATEGIES IN WORLD CHAMPIONSHIP TETRIS PLAY

As alluded to in the previous studies, extreme levels of skill do indeed exist for Tetris. In Study 3 I investigate the cognitive underpinnings of extreme expertise in Tetris to better understand the natural process by which such skill is acquired. To achieve this, I compiled the existing local tournament data with world champion Tetris player data collected on-site at the annual Classic Tetris World Championship (CTWC) in Portland, Oregon, USA. Tetris world champions play the game at the knife edge of human cognitive ability, and, as such, adopt specific strategies overlooked by those earlier in the skill-acquisition process. The analysis pulls apart the individual features of the expertise-essential "decide-move-placed" component from Study 1 to illuminate the cognitive strategies employed by extreme Tetris experts.

5.1 STUDY 3: GUIDING RESEARCH AND RATIONALE

A primary function of expertise is to increase efficiency in a domain using specialized retrieval structures. To achieve this, the brain makes use of chunking processes to enhance pattern recognition, thereby reducing decision-making times (Ericsson & Kintsch, 1995). In terms of Hick's law for decision processing time, this chunking process can be thought of as simply reducing the task search space by way of grouping together the most task-relevant states and actions. Logan (2016) suggests that the process of acquiring these chunks, and thus expertise, can be seen as a balancing act between the invariant elements of the task environment– in this case, Fitts' Law for motor actions– and Hick's law for decision-making. Fitts' law is considered to be relatively stable with practice; i.e., practice will only negligibly reduce the time taken to maneuver fingers to known targets over some distance. Thus, Logan argues that the majority of the work done by the cognitive system striving for typing expertise is in working to reduce the time taken to make decisions, or rather, reduce the impact of Hicks' law.

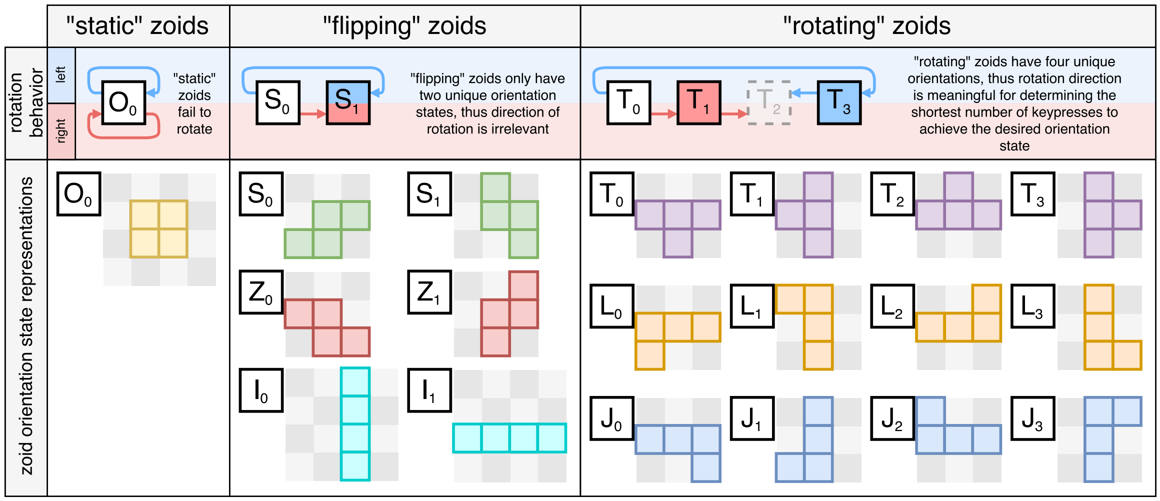
The process for acquiring any skill, however, is iterative; players can only innovate on what is obvious and available in their current interpretation of the task at hand. The "soft constraints hypothesis" (Gray, Sims, Fu, & Schoelles, 2006) suggests that the human cognitive system tends to select the fastest strategy, in-the-head or in-the-world, as it currently appears. Thus at certain times in the learning process players may be misled to believe they have found a good strategy when, in fact, they have arrived at a "stable suboptimal solution" (Fu & Gray, 2004) – i.e., by minimizing their search space early on, they may head down a strategic path doomed to plateau. An example of these Hick's law trade-offs in Tetris can be found in the zoid-rotation subtask.

5.1.1 Rotational strategies in Tetris

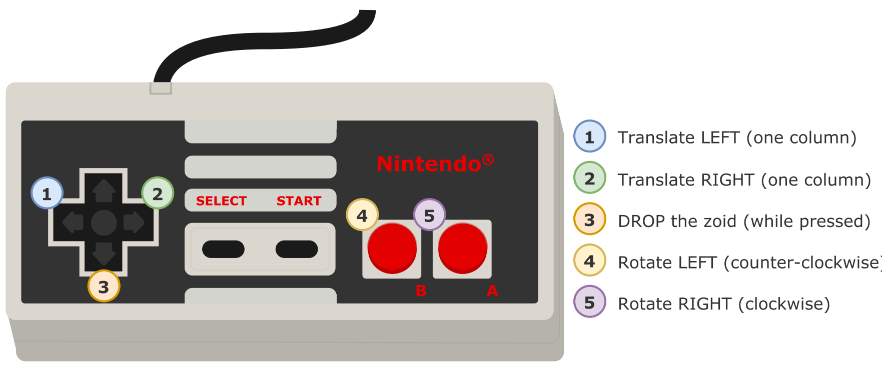
Tetris, like any complex task, can be broken down into several subtasks: visual search, evaluation of options, aligning the zoid vertically, recovering from (or avoiding) errors, etc. Here, I examine a fundamental element of even novice Tetris play, the zoid-rotation subtask. The Tetris champion's strategic reference book "Tricks of the Classic NES Tetris Masters" (Smith, 2014) highlights a simple, but important, element of the game: to get the zoid into its desired ultimate destination, players can press a button to rotate the zoid either left (counterclockwise) or right (clockwise). This distinction is deceptively simple, as employing both buttons in a strategy carries some significant cognitive baggage.

Figure 5.1 shows the rotation behavior of zoids in Tetris. For many zoids, rotation or direction of rotation is a non-issue, as all rotation actions amount to the same result. For example, "static" zoids (the "O"-zoid) do not rotate at all and thus button presses are not applicable to rotation, and "flipping" zoids (e.g., the "I"-zoid) do rotate, but because they only have two possible orientation states the direction of rotation is irrelevant, as both directions result in the same resultant orientation. The remaining zoids, "rotating" zoids, have four distinct orientations, and thus the direction of rotation is meaningful for determining how many button presses are required to achieve a desired orientation.

Two strategies emerge for dealing with the third category of "rotating" zoids: the mono-rotational strategy, wherein one dominant rotation direction is used for all rotation keypresses; and the bi-rotational strategy, wherein both rotation directions are employed to minimize keypresses.



(a) Zoid behavior.



(b) NES controller.

Figure 5.1: (A) Internal representations of Tetris zoids and their rotation behavior. "Static" zoids do not rotate and require no rotation button presses; "flipping" zoids only have two orientations, so rotation is possible, but direction is irrelevant, as both direction result in the same final orientation; "rotating" zoids have four unique orientations, and thus require either multiple button presses, or use of both rotation directions to reach all possible states. (B) The NES controller, its buttons, and their functions in the game.

5.1.2 Mono-rotational strategy

The mono-rotational strategy is simple: to get a zoid from its current orientation to the desired orientation, all a player needs to know is how many times to press the rotate button. This strategy is simple in that it does not even consider that there are two directions to rotate the zoid– as far as this strategy is concerned, there is only one rotate button.

The mono-rotational strategy does have a critical inefficiency: for a number of common scenarios in the game this strategy requires pressing the rotate key three times in succession, which is not ideal for the break-neck pace of high-speed expert Tetris play.

5.1.3 Bi-rotational strategy

The bi-rotational strategy is more complex: the player utilizes both rotation buttons to rotate the zoid only in the direction that requires the fewest keypresses. As a result, the player must track more information, which may result in some cognitive slowdown, but as a reward for that additional effort the number of keypresses required to rotate the zoid is kept to a minimum.

5.1.4 Cognitive implications of rotation strategies

Each strategy optimizes different criteria, and is adaptive in its own way. The mono-rotational strategy, though simple, is adaptive in nature because it minimizes decision-making by pruning the search space of all possible solutions involving the player's non-dominant rotation direction. Early in the game, when time is plentiful, the difference between using one or both rotation buttons is essentially null– players may feel they have all the time in the world to rotate the zoid, and thus the cognitive savings appear to outweigh the motor costs. The bi-rotational strategy, on the other hand, is adapted to the later stages of the game, when shaving off the milliseconds wasted on additional rotations is critical to success. In that scenario, the player overcomes the cognitive difficulty of rotating the zoid in both directions to minimize their motor activity and achieve higher levels of success.

Adaptive rationality aside, there are additional considerations for understanding which strategy a player is likely to adopt. Consider the memory structures required to support making decisions in each strategy. The mono-rotational strategy requires a relatively simple memory structure: for a given episode's initial-orientation and destination-orientation pair, all that is required is a simple memory look-up operation to identify how many times to press the player's dominant rotation key. The bi-rotational strategy, on the other hand, requires the player to identify both (a) the direction, and (b) the number of rotations required to achieve the minimum keypresses. Moreover, this bi-rotational memory structure must also be sensitive to the current zoid's type ("static", "flipping", or "rotating"), or risk wasting time making meaningless decisions for zoid types that do not require information about rotations or rotation direction.

Because of the relative complexity of each strategy's memory structures, their maintenance and development should also differ. Though I ultimately disagree (Destefano et al., 2011) with Kirsh and Maglio's (1994) conclusion that complementary actions are indicative of high degrees of skill, I find their suggestion that such actions may facilitate learning compelling. They suggest that during the early stages of skill acquisition in Tetris, players will over-rotate zoids while searching for solutions, doing direct visual comparison of the zoid and pile shapes in-the-world rather than in-the-head. This operation may in fact lighten the cognitive load while learning to navigate the other elements of the task, but more importantly performing these over-rotations more frequently also lends to the reinforcement of basic button-press-to-zoid-rotation action chunks, helping players gain an intuitive understanding of how their button pressing actions affect the progression of zoids through their respective sets of orientations.

Considering each strategy in this light, the mono-rotational strategy memory structures would emerge as a natural consequence of minimizing the initial task complexity: if a novice is rotating the zoid in only one direction to reduce the decision-space, and rotating it more often to reduce computation in-the-head, then the memory chunks being constructed and strengthened early on would be those involved in single-direction rotations, i.e. the mono-rotational strategy. Moreover, the natural development of these single-direction chunks produces a feedback loop: any player who initially avoids rotating zoids in both directions will be more likely to employ the mono-rotational strategy, which will, in turn, only strengthen the single-direction memory structures. Thus, because of this "rich get richer" effect (Lebiere, 1999), the bi-rotational strategy is unlikely to develop in the first place, as players experimenting with rotating both directions may face competition between chunks when attempting to recall which rotation button will achieve which destination, ultimately falling back on their tried-and-true, naturally strengthened mono-rotational strategy.

Figure 5.2 illustrates a possible arrangement of the cognitive "pipeline" of the two rotational strategies. Both the mono-rotational and bi-rotational processes assume players have accurately harvested the perceptual board information and already selected a destination for their current zoid. The paths diverge in terms of interaction with memory: the mono-rotational strategy uses the current zoid and the destination orientation to key into a simple look-up table to determine the number of rotations necessary; whereas the bi-rotational strategy requires finding two solutions, comparing them, and finding the shortest number of keypresses. Once the number (and direction) of keypresses is computed, motor actions can be programmed and executed.

The two strategies for rotating zoids do well to illustrate the balance between reductions to the motor cost and reductions to the decision space of a task: the mono-rotational strategy minimizes the decision-space (Hick's law) while failing to optimize for the additional keypresses required in some scenarios, whereas the bi-rotational strategy trims down the overall keypresses, but with some expected cognitive cost. Moreover, these two strategies also represent an excellent example of a “stable suboptimal solution”, wherein the attractiveness and self-reinforcing nature of one strategy makes it unlikely that a player would even encounter, let alone develop, a superior solution.

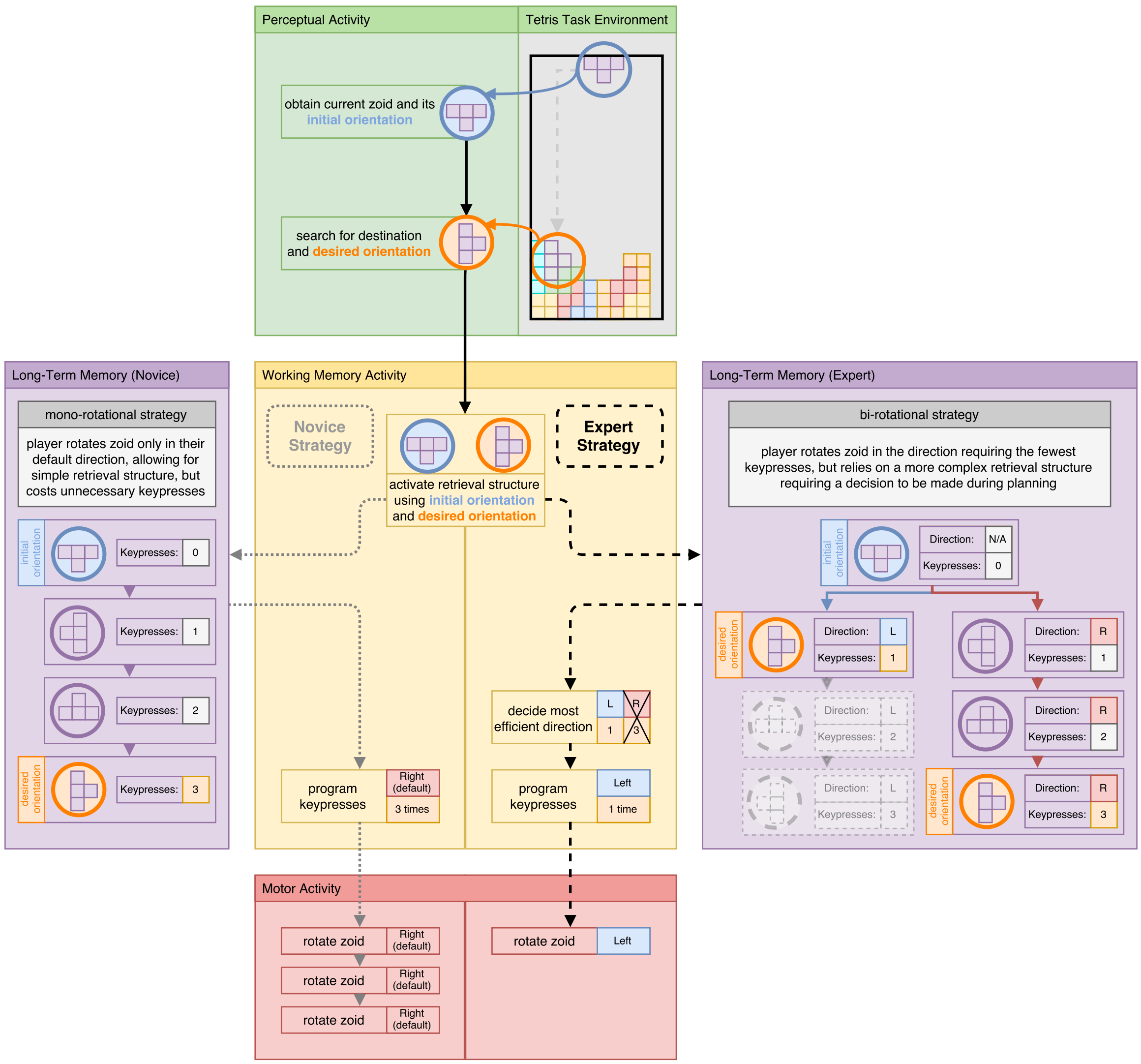


Figure 5.2: Diagram of the possible flow of information for each of two cognitive strategies involved in the zoid-rotation subtask. Players using the mono-rotational strategy rely only on simple memory retrieval structures to perform their task (minimizing the impact of Hick's law), but at the cost of sometimes executing two additional keypresses. Conversely, players using the bi-rotational strategy minimize number of rotations by employing both rotation directions, but at the cognitive cost of relying on a more complex retrieval structure. The retrieval structures presented here are intended to illustrate the relative complexity of the information involved in each strategy, not make strong predictions about the precise structure of memory.

5.1.5 Hick's law and extreme expertise

One question remains: are extremely skilled players who adopt this strategy still impacted by Hick's Law? The literature suggests that as expertise increases, the impact of Hick's Law approaches zero (A. R. Jensen, 1987; Logan et al., 2016). If expertise can be thought of as a reduction in the decision space of a task, does the same apply to expert memory structures, or is there some inherent limit to this theorem?

To investigate this question, I turn to extreme examples of skill in the Tetris task, those who compete in the Classic Tetris World Championship. [1] These world champions manage to perform under such extreme conditions of time pressure that the fact that they can navigate the task at all is impressive. As such, one might reasonably assume that their memory structures are similarly finely tuned. Because Hick's law is associated with reaction times, I ask the following: do world champion players who have adopted the bi-rotational strategy exhibit slower reaction times for "rotating" and/or "flipping" zoids? If these world champions' reaction times for each type of zoid do not differ, then these world champions have managed to reduce the complexity of the bi-rotational strategy to a "flat" memory structure, similar in its complexity to that of the mono-rotational strategy. If these reaction times do differ, then even the best Tetris players in the world would appear to be unable to reduce the complexity of their memory structures below some lower limit, thus questioning the generality of the conclusion reached by Logan et al. (2016).

5.1.6 Hypotheses

My primary hypothesis in this investigation is that skilled bi-rotational strategy users will exhibit increased initial reaction time for "rotating" zoids over "flipping" and "static" zoids, whereas skilled mono-rotational strategy users will exhibit no such reaction time penalties due to the relative simplicity of the strategy. This hypothesis assumes that: (1) initial reaction time (i.e., the time from the start of an episode to the first keypress) represents purely cognitive processing time, and (2) the experts have reached asymptote on their relative reaction times. If supported, the findings would indicate that the expert memory strategy for each zoid type appears to have some element of irreducible complexity, and that Hick's law cannot bend to reduce decision time beyond that point. If not, then experts have managed to solve the task to such a degree that their memory

[1] Most of the CTWC players reported having practiced for thousands of hours attempting to improve their skills to win championships, some of whom have won multiple consecutive championships in various incarnations of the game. They also have all reached the difficulty level of the game known as the "kill screen", which is so fast that it is considered humanly impossible to survive in it for more than a few moments. These players are motivated, skilled, and operating at the edge of human ability.

structure is flat, requiring no more initial time to react than they would using the mono-rotational strategy, suggesting that Hick's law continues to bend late into skill acquisition.

Before asking this question, however, I must verify some basic premises about the dataset:

Basic premise 1: world champion ("global"), local champion ("regional"), and low skill ("novice") Tetris players will all differ significantly from one another in terms of both overall game score (sixth root, see Study 1) and execution time, measures related to performance outcomes and performance processes, respectively.

Basic premise 2: Global champion players will exhibit significantly more non-dominant rotations for "rotating" zoids than will novice players, and these players will show enhanced rotation efficiency and overall latency. That is to say, global champions will show evidence that they employ the bi-rotational strategy, as it is more adaptive for high-speed play by eliminating costly unnecessary rotation button presses (and also as illuminated by the subjective accounts in "Ecstasy of Order" documentary). Novices, on the other hand, will employ the mono-rotational strategy, as they will have opted to minimize their decision space, and thus the impact of Hick's law on their fledgling skill in the task. Because I am not sure precisely when along the progression of expertise to expect players to adopt the bi-rotationl strategy, it is unclear which strategy regional champion players would adopt; however, because of the initial attractiveness and self-reinforcing nature of the mono-rotational strategy described in the previous section, I expect it is more likely that these regional champion players employ the mono-rotational than the bi-rotational strategy.

5.2 STUDY 3: METHOD

5.2.1 Participants

I selected three groups of participants from both the local and global Tetris tournaments: **novices**, **regional** champions, and **global** champions. The local tournaments analyzed were the same as those used in Study 2, except in this case I selected two sets of players: 10 **novices**, the ten lowest-scoring qualifying round entrants across all three years; and 10 **regional** champions, the ten *unique* champions or runners-up across all three years (two of these players either placed or were runners-up in two different annual tournaments).

The third group consists of 10 **global** champion players from the Classic Tetris World Championship. I arranged with the tournament organizers to meet with, interview, and collect data from these ten players (the most I could get with the time allotted). Among the players I met were the five-time 1st place champion, as well as the only other player to win 1st place apart from him. The rest of the players all routinely ranked among the top 8 across the various championships. Suffice to say, these players represent sampling from the extreme high end of expertise.[2] As in Study 2, data collection was approved by the Rensselaer IRB in an IRB proposal written specifically for the tournaments.

5.2.2 Procedure

The conditions of the local Tetris tournaments were as described in Study 2. The CTWC players participated under similarly experimentally challenging conditions: the venue was somewhat loud, open, and distracting, conditions that come with the territory in field-work.

All the data used in this investigation were collected during the qualifying rounds, or during my individual sessions with the CTWC players; i.e., I did not include data from the tournament brackets themselves. During the qualifying round played by the **novice** and **regional** players, each player played one game that used a predetermined random number generator seed to produce identical sequences of zoids in the game. The same seed was used during one of the games in my sessions with the **global** players. Thus, all players played one complete game in which the zoid sequence was identical. Only data from these games are used in the present investigation, resulting in one game's worth of data from each participant.

Though Meta-T attempts to perfectly replicate the 1985 NES version of Tetris, there are some subtle differences that the CTWC players were able to pick up on. [3] However, because all of the data were collected using the same program, I consider them internally consistent and thus a valid lens through which to examine the differences between all of the players of different skill levels.

5.2.3 Data partitioning

Data collection produced an abundance of data for each player at the episode level. However, my analyses target higher levels of aggregation, and as such, all data are averaged at the level relevant to the analysis presented. For example, when examining only differences for a single

[2]For a video example of CTWC players at peak performance during the tournament finals (with announcer commentary), refer to https://www.youtube.com/watch?v=DdfRQjb5o9k

[3]Due to some bugs in Meta-T's code, it does not capture the behavior of the 1985 NES Tetris as closely as I had initially thought. Some subtle inconsistencies in how Meta-T tracks timers internally caused some scenarios wherein the CTWC players immediately noticed that they could not "tuck" the zoid the way they expected to. This difference— which was imperceptible for the code's designer (me) and even the regional tournament champions— made it difficult for these Tetris world champions to succeed at the game much past level 19 (a feat which is effortless for them on the NES version). This interesting finding might warrant its own investigation, but as I have insufficient observations to be able to ask such questions, I relegate them here to a footnote.

outcome measure across the between-subject skill categories (novice, regional, and global), each player's performance data will be reduced to a single data point. Whereas if I am examining that same outcome across those same skill categories, but also the within-subject zoid rotation type (static, flipping, and rotating), each player's data will instead be reduced to one data point for each of the three zoid rotation types (or thrice as many observations).

5.3 STUDY 3: RESULTS

I present evidence to address three main questions. First, do the skill groups truly differ in performance? Second, to what extent do different players adopt different rotational strategies? And third, are extremely skilled players who adopt the bi-rotational still subject to Hick's law, or has their expertise diminished its impact to near-zero as suggested by the findings of Logan, et al., (2016)?

5.3.1 Differences in performance between skill groups

First, to verify that the players vary in their overall game performance, I conducted a one-way between subjects ANOVA to compare the effect of player skill on total game score (sixth root) for players in the novice, regional, and global skill conditions. There was a significant effect of player skill on game score (sixth root) for the three conditions [$F(2, 27) = 161.8$, $p < .00001$, $\eta^2=0.92$]. Figure 5.3 shows the trace of player scores over game difficulty levels, and the distribution of final game scores (sixth root). Post-hoc comparisons using a pairwise t-test with Bonferroni correction revealed that game score (sixth root) was significantly higher for global (M = 7.97, SD = .29) vs regional (M = 6.41, SD = .33) players, both of which differed significantly from novices (M = 4.11, SD = .71). Regional players's scores also significantly differed from novices. These results indicate that the players do, in fact, differ significantly in their overall game performance.

I also examined whether players differed on more immediate measures of skill by comparing total execution time (i.e., the time to either dropping a zoid, or the game automatically placing it due to gravity) for each category of player skill. For this analysis, I examined only data from game difficulty level 0, as this is the only game level that all players completed, and this is the level at which time pressure is lowest. I conducted a one-way between subjects ANOVA to compare the effect of player skill on mean execution time for players in the novice, regional, and global skill conditions. There was a significant effect of player skill on execution time for the three conditions [$F(2, 27) = 51.14$, $p < .00001$, $\eta^2=0.79$]. Post-hoc comparisons using a pairwise t-test

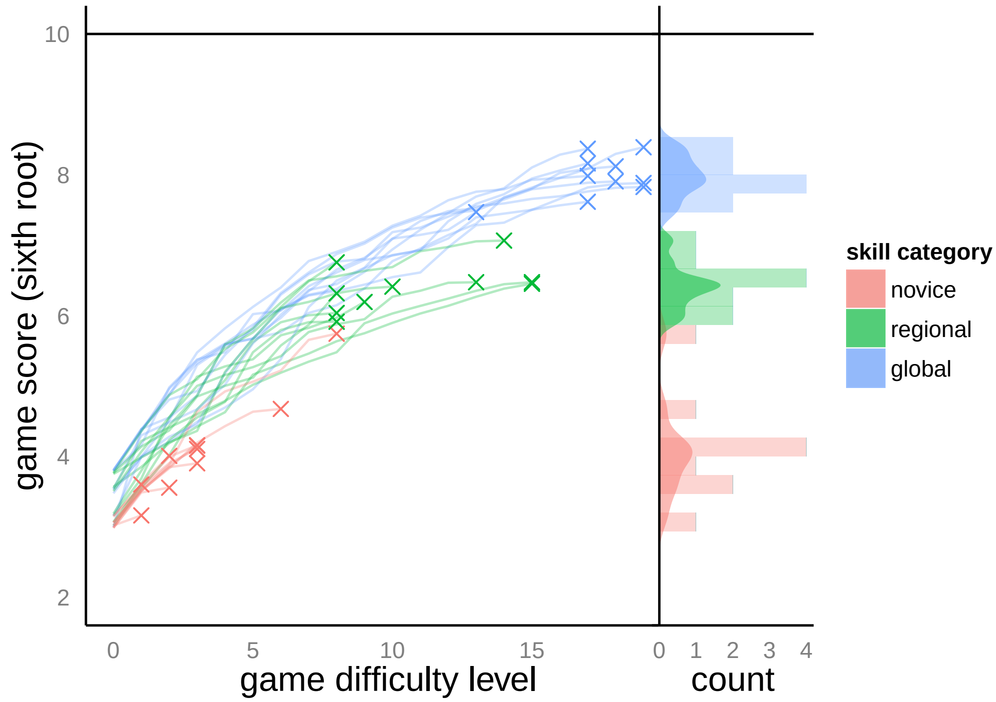


Figure 5.3: The trace of game scores across all game levels for all 30 participants. The ’x’ marks the point at which the game terminated. The right panel shows the distribution of final game scores for players in each skill category.

with Bonferroni correction revealed that execution time was significantly lower for global (M = 629.7ms, SD = 38.8) vs regional (M = 1423, SD = 562.3) players, both of which differed significantly from novices (M = 2933ms, SD = 701.2). Regional and novice players also differed significantly. Figure 5.4 shows the comparison of execution times for players of each skill category. Figure 5.5 shows how execution time changes for players of each skill category as game difficulty level increases.

Together, these findings satisfy a basic premise of this investigation: that players sampled from low skill tournament entrants (novices), local tournament winners (regional), and world champions (global) all differ in terms of both their performance outcomes (game scores) and immediate interactive performance throughout the game (execution time).

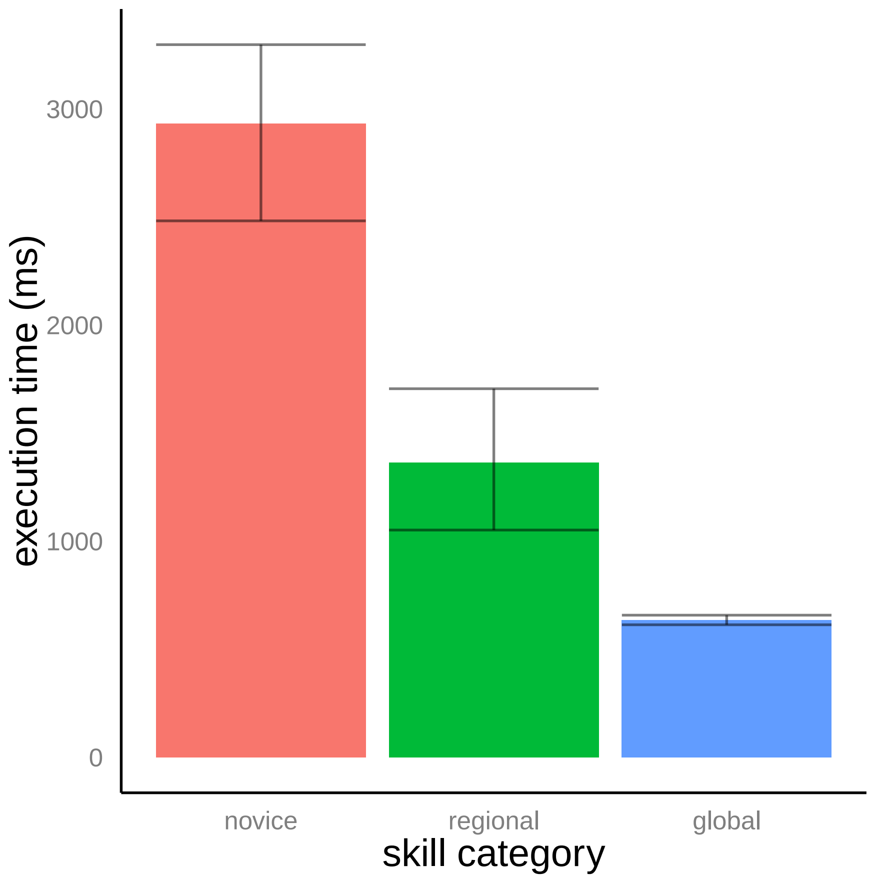
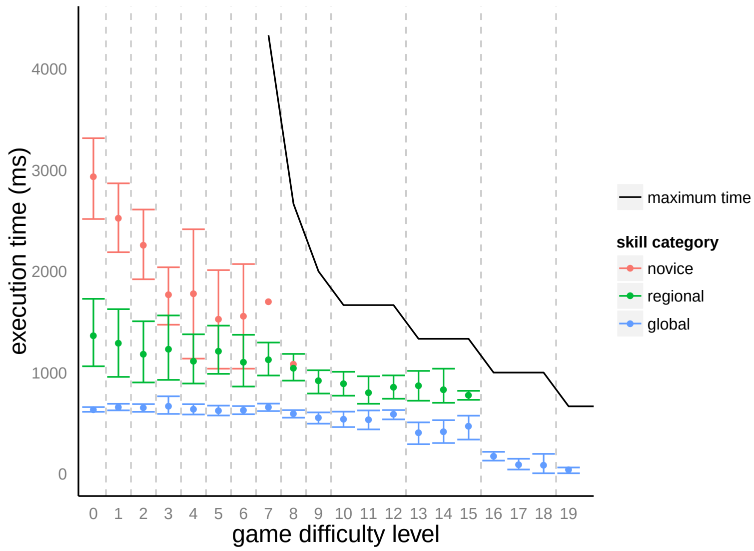
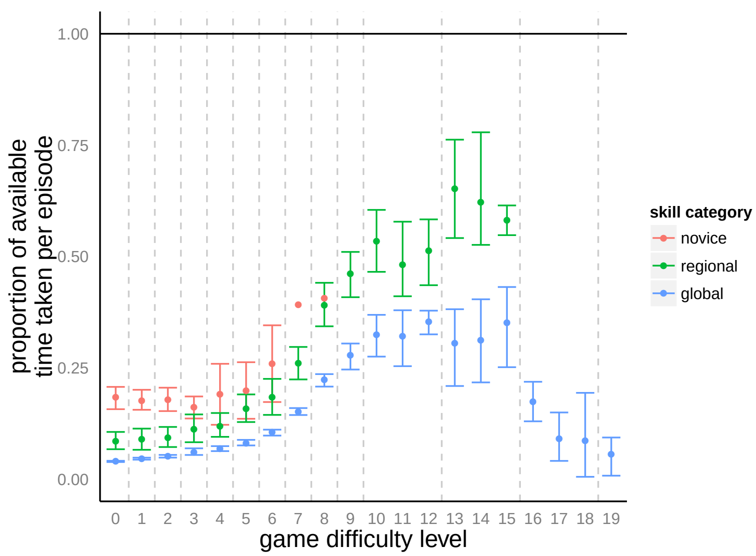


Figure 5.4: Total execution time per episode for players of each skill category at game difficult level 0 (the beginning of the game). Execution time is defined here as the time until either: (1) the moment the zoid is manually dropped, or (2) the moment the zoid stops at rest in the pile. Error bars represent 95% confidence intervals.



(a)



(b)

Figure 5.5: Total execution time per episode for players of each skill category across all game levels. Execution time is defined here as the time until either: (1) the moment the zoid is manually dropped, or (2) the moment the zoid stops at rest in the pile. Figure (a) shows the execution time in milliseconds across all levels, while figure (b) shows the proportion of time used by the player compared to the maximum available in in the episode, as determined by the current difficulty level. The black line in both figures indicates the maximum time available in an episode for a given difficulty level. Error bars represent 95% confidence intervals.

5.3.2 Evidence of rotational strategy use and efficacy

Figure 5.6 shows the balance of rotation directions employed by player, skill level, and zoid type. For both novices and regional players, there is very little variation from a player's "dominant" rotation direction, implying these players use the mono-rotational strategy. In contrast, global players appear to use a dominant rotation direction for "flipping" zoids, but are more selective for the "rotating" zoids, implying they use the bi-rotational strategy.

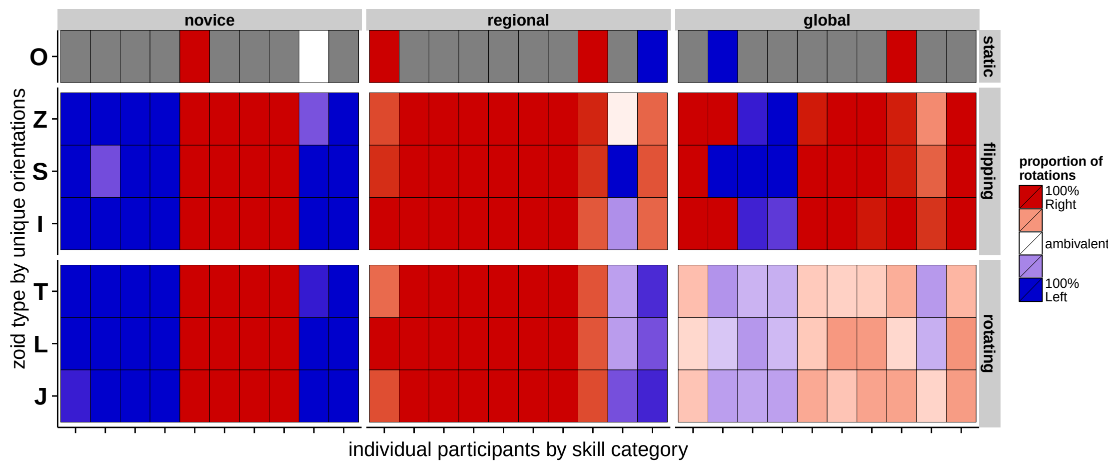


Figure 5.6: Balance of rotation direction usage (color) for each individual player (individual columns) grouped by player skill category (column groups) for each zoid type (individual rows) grouped by zoid type (row groups).

To more rigorously compare players' strategy use, I measured the proportion of each player's rotation button-presses that were in that player's non-dominant rotation direction. First, I conducted a one-way between subjects ANOVA to compare the effect of player skill on log-proportion of non-dominant rotations for players in the novice, regional, and global skill conditions. Because the bi-rotation strategy is only meaningful for "rotating" zoids, I restricted the analysis to only include those zoids. There was a significant effect of player skill on non-dominant rotations for the three conditions [F(2, 27) = 98.86, $p < .00001$, η^2=0.88]. Post-hoc comparisons using a pairwise t-test with Bonferroni correction revealed that the proportion of non-dominant rotations was significantly higher for global (M = .206, SD = .105) than for regional (M = .086, SD = .153) and novice (M = .005, SD = .011) players. Regional and novice players' proportions of non-dominant rotations did not differ. Figure 5.7-A shows this effect.

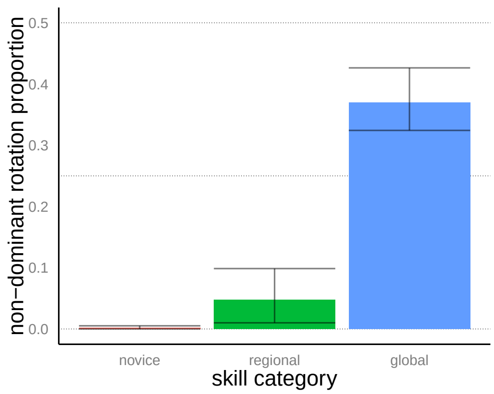
Second, I examined the impact of rotational strategy adoption on motor efficiency by comparing the number of extra rotations for "rotating" zoids exhibited by players in each skill category. I conducted a one-way between subjects ANOVA to compare the effect of player skill on extraneous rotations for players in the novice, regional, and global skill conditions (collapsed across game levels and only for "rotating" zoids). There was a significant effect of player skill on non-dominant rotations for the three conditions [F(2, 27) = 37.18, $p < .00001$, η^2=0.73]. Post-hoc comparisons using a pairwise t-test with Bonferroni correction revealed that extra rotations were significantly lower for global (M = .022, SD = .018) than for regional (M = .756, SD = .262) players, and both were significantly lower than novices (M = 1.132, SD = .434) players. This result suggests that global players (who all appear to have adopted the bi-rotational strategy) are more efficient at rotating zoids than regional and novice players.

Together, these findings suggest that global champion players do, in fact, employ the bi-rotational strategy, while regional and novice players rely most heavily on the mono-rotational strategy. Moreover, the bi-rotational strategy appears to achieve its intended effect of reducing the number of unnecessary rotations required to get the zoid to the desired orientation.

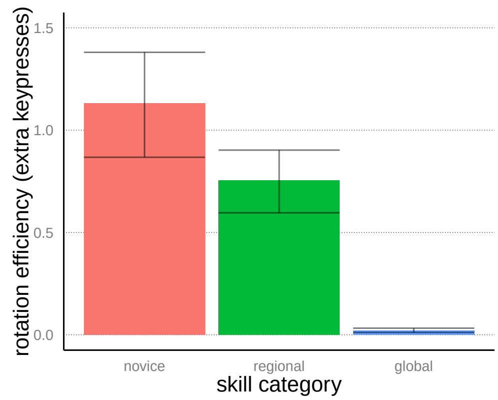
5.3.3 Impact of Hick's law on expert decision-making

To this point, the analysis has focused on sorting out (1) whether the groups differ in demonstrated expertise, (2) which strategy is adopted by each group, and (3) whether the strategies result in gains in efficiency as expected. I have presented evidence that global champion players both (1) exhibit reduced execution times, and (2) employ the bi-rotational strategy, so my final analysis asks whether or not the added complexity of the bi-rotational strategy impacts these players' initial reaction times (i.e., their initial cognitive processing times) compared to skilled players employing the mono-rotational strategy. For this analysis I am interested in examining fine-grained differences in initial reaction times for players who have achieved some level of mastery over the strategies they employ, thus novices are omitted. Initial reaction time for a given episode is considered to be the time between the zoid appearing on the screen and the player making their first button-press, time during which I assume any delays are due to additional cognitive processing. I consider regional players to by and large exhibit skilled usage of the mono-rotational strategy, while global players unanimously exhibit skilled usage of the bi-rotational strategy. In addition, I want to examine those game episodes where players are close to the limit of their capability, when time pressure is high and a high degree of performance is critical, so I limit the analysis to only performance data from the highest level of game difficulty fully completed by each player.

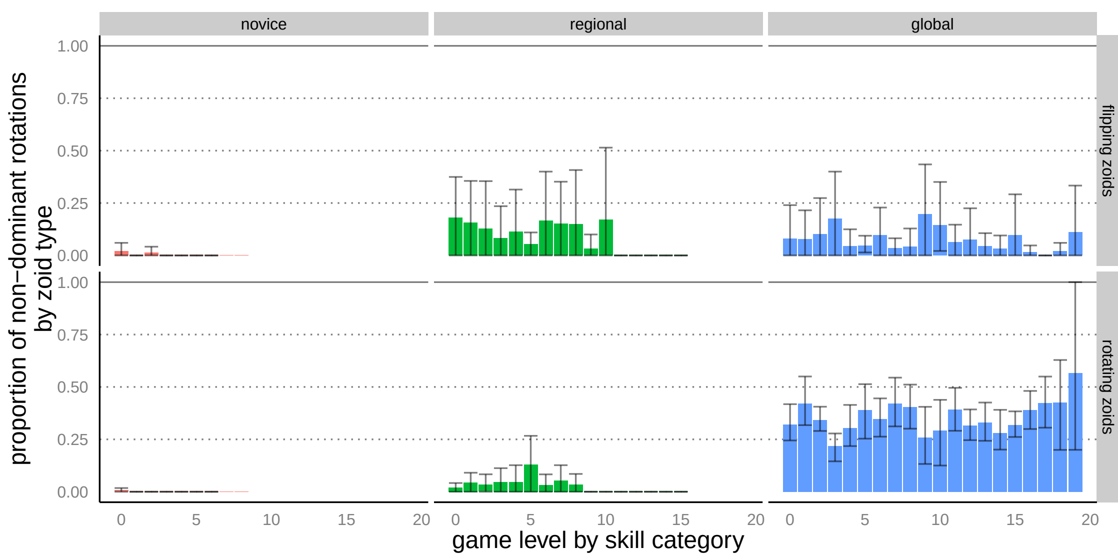
I conducted two one-way ANOVAs, one for each player skill level (regional or global), to compare the effect of each zoid rotation type (static, flipping, or rotating) on players' initial reaction times. For the regional champion model (i.e., the skilled mono-rotational players), there was no significant main effect of zoid rotation type on initial reaction time [F(2, 27) = .20, $p = .821$, $\eta^2 = .01$]. For the global champion model (i.e., the skilled bi-rotational players), there was a significant effect of zoid rotation type on initial reaction time [F(2,27) = 82.55, $p < .00001$, $\eta^2 = .86$]. Post-hoc comparisons using a pairwise t-test with Bonferroni correction revealed that global players' initial reaction times for rotating zoids (M = 126.3ms, SD = 17.89) were significantly higher than for flipping zoids (M = 80.29ms, SD = 9.22ms) and static zoids (M = 57.33ms, SD = 6.55ms), and reaction times for flipping zoids were higher than for static zoids. Table 5.1 shows the means and standard deviations for each zoid rotation type for both regional and global champion players. This result suggests that global players (those employing the bi-rotational strategy) are differentially affected by the zoids' number of possible rotations. Figure 5.8 shows this effect, and also expands to see how it plays out over the course of a game as difficulty increases. For a more detailed exploration of individual players' initial latencies and tables of group means for each game level, see Appendix C.



(a) Nondominant rotations ("rotating").

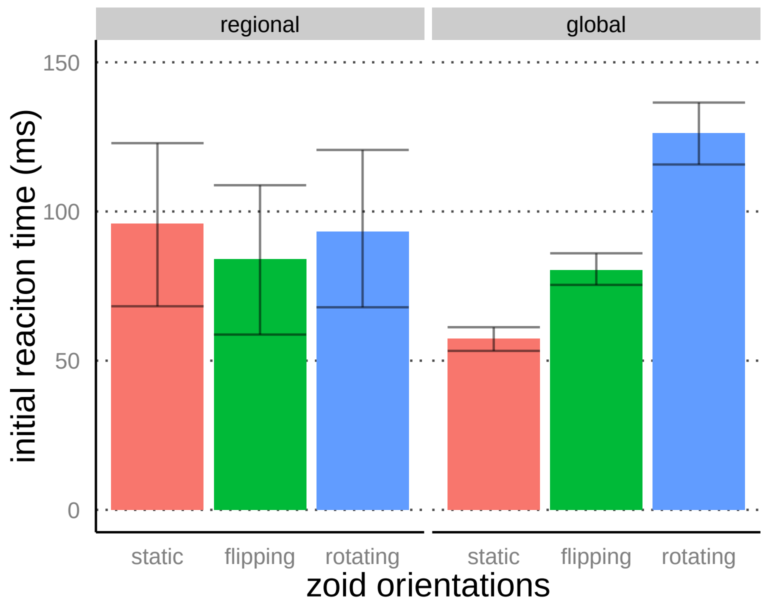


(b) Efficiency ("rotating").

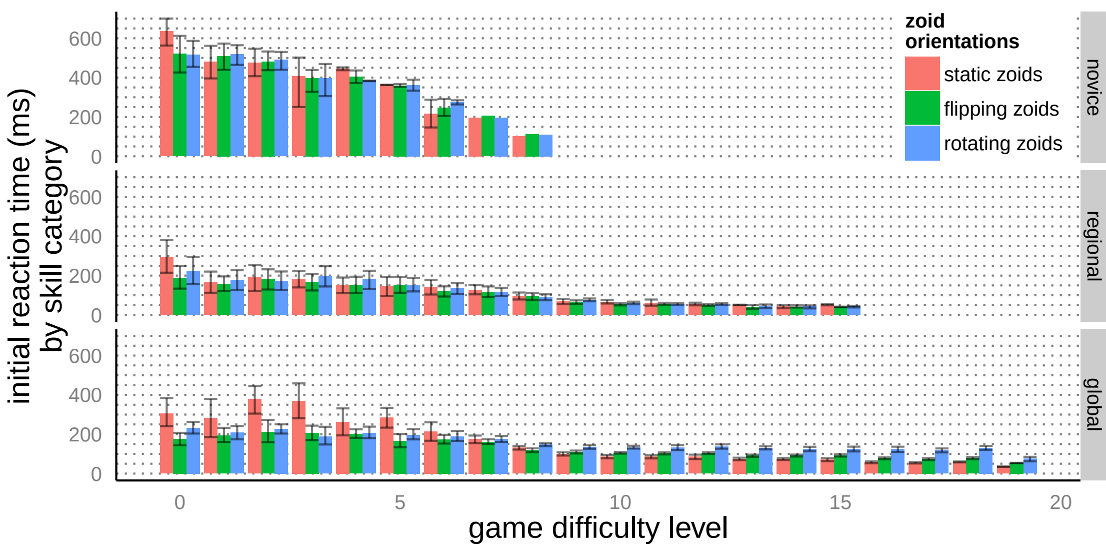


(c) Nondominant rotations (by level and zoid type)).

Figure 5.7: Evidence and effect of different rotation strategies. Figure (a) shows the proportion of rotation keypresses in a player's non-dominant direction. Figure (b) shows the rotation efficiency in terms of how many extraneous button presses are made to get the zoid into its ultimate orientation (a zoid could, at maximum, require two button presses to achieve a desired orientation). Figure (c) shows in further detail how non-dominant rotations progress over the course of the game (levels on x-axis), for players of each skill category (columns) and for each style of zoid rotation behavior (rows). Error bars represent 95% confidence intervals.



(a) Initial reaction time at highest level completed.



(b) Initial reaction time by zoid type across level.

Figure 5.8: Figure (a) shows the mean initial reaction times per episode for champion players (regional and global) for the player's highest completed game level. Regional players adopt the mono-rotational strategy, while global players adopt the bi-rotational strategy. Figure (b) shows the initial reaction times for each zoid type for players in all three skill categories across the full breadth of game levels. Error bars represent 95% confidence intervals.

Table 5.1: Means (and standard deviations) of reaction times for regional and global players for each zoid rotation type.

	static zoids		flipping zoids		rotating zoids	
regional champions	95.94	(47.16)	84.01	(40.74)	93.27	(45.22)
global champions	57.33	(6.55)	80.29	(9.22)	126.3	(17.90)

5.4 STUDY 3: DISCUSSION

The preceding analysis produced the following results. First, each skill category (novice, regional, and global) produced game scores and execution times that significantly differed from one another. Second, I demonstrated that novice and regional players adopted the mono-rotational strategy, and global champions adopted the bi-rotational strategy. Third, I showed that global champions (using the bi-rotational strategy) reduced the number of extraneous zoid rotations per episode to near-zero. Finally, fourth, I found expert players using the bi-rotational strategy had initial reaction times that were significantly impacted by zoid orientation type (static, flipping, and rotating, in increasing order of complexity), while there was no systematic difference for skilled users of the mono-rotational strategy. Next I discuss the implications of these findings.

5.4.1 The "cognitive speed-bump" of the bi-rotational strategy

My primary finding in this study is that even extremely skilled players experience a cognitive "speed-bump" when adopting a strategy that involves more complex decision-making, and that this is true even at extreme levels of expertise. The global champions do, in fact, save time by adopting the bi-rotational strategy to avoid unnecessary zoid rotations, as evidenced by their decreased extraneous button presses (Figure 5.7b) and their overall faster episode execution times (Figure 5.5). However, these global champion players appear to achieve their overall more efficient performance at the cost of considering more information, evidenced by their increased initial reaction times for more complicated zoid types (Figure 5.8a). By contrast, the regional champions, who have adopted the simpler mono-rotational strategy, show no systematic slowdown when determining how much to rotate each type of zoid, regardless of how many orientation states that zoid may possess.

The "speed bump" becomes even more evident when considering Figure 5.8b. While the statistical analysis focused on the "highest game level achieved" for each player to attempt to best characterize each player's performance near their own cognitive limits, this figure compares gameplay data from each game level. Though the number of players begins to decrease, it is apparent that the variability of reaction times reduces, and the means seem to converge: regional players using the mono-rotational strategy under the same time pressure as global players using the bi-rotational strategy show flat reaction times near 50ms, while global players show a minimum reaction time near 50ms that increases with zoid orientation type, and thus decision complexity. This finding implies that the strategy adopted by the highly skilled global champions appears to introduce temporal costs above and beyond the use of the mono-rotational strategy.

These findings imply some interesting things about the decision structures involved in the two rotation strategies. While on the surface the bi-rotational strategy initially seemed to require a more complex decision process, it was also reasonable to expect that this complexity could be reduced with continued practice and chunking processes. This was evidently not the case. Even the best Tetris players in the world are still susceptible to slowdowns associated with Hick's law.

However, also consider regional champions. Though these players cannot boast being the best in the world, they are skilled enough to outshine dozens of others in local tournaments, even when employing the mono-rotational strategy. When considering these regional champions' reaction times across each zoid type, they appear flat, especially in the later stages of the game; regional players' reaction times for all zoid types all appear on par with global players' *fastest* reaction times, those for "static" zoids where there is no rotation decision to be made (see Table 5.1). This implies that skilled mono-rotational strategy users are not sensitive to the same decision-based slowdown as the bi-rotational strategy users, and that even if there ever was any such slowdown, it must have been eliminated with practice (i.e., it is possible that Hick's law *did* bend for players using the simpler strategy).

The zoid rotation subtask is ultimately very simple: transform an object from state A to state D using 1 to 3 button-presses. Nevertheless, there exist two strategies for approaching it, a simple strategy which can be practiced to such an extent that it appears to require no decisions, and a more complex strategy which appears to possess some irreducible complexity. This implies that, despite the wonders of the human cognitive system to adapt to any task, there may be limits on just how cognitively streamlined a given strategy can be, even for simple tasks. Players at this level of skill may not, in fact, be merely plateaued, but have reached a "cognitive asymptote" for the adopted strategy in the task (Gray & Lindstedt, 2016). This leads me to a friendly amendment to Logan et al.'s (2016) suggestion that choice reaction times reduce with expertise: "Hick's law bends with practice, *except when it cannot*."

5.4.2 Implications for strategy discovery

A secondary finding was that my prediction about strategy adoption was accurate for novices and global champions– novices used the mono-rotational strategy, while global champions used the bi-rotational strategy. But more interesting is the case of the regional champions: though they clearly relied primarily on the mono-rotational strategy, it is apparent that their non-dominant rotation usage was at least non-zero, implying that they may have been toying with the idea of using the bi-rotational strategy. Figure 5.7c shows the incidence of non-dominant rotations across game levels and for both "flipping" and "rotating" zoid types. This illustrates that early on in a game, when time pressure is low and there is time to spare, some regional champions made use of both rotation directions (which is unlikely to happen purely by chance). Later in the game, when time pressure is increased and cognitive resources are spread thin, this usage falls back off to zero and they resume using their primary mono-rotational strategy. This suggests that some of the regional players may be just on the cusp of developing the bi-rotational strategy by practicing it whenever possible, but it is too costly to do so when time pressure rises.

Figure 5.7c supports my earlier discussion that it is a difficult process to "unlearn" one strategy in favor of a different one, particularly when the task tends to put extreme cognitive demands on the player. Although the mono-rotational strategy is ultimately sub-optimal in terms of the overall task (as it involves more button presses than are strictly necessary), it appears to be an obvious choice early on as even novices pick a single rotation button and stick with it. Moreover, when players achieve enough skill to reach game levels high enough to notice difficulties with the mono-rotational strategy and consider implementing the bi-rotational strategy– as some of the regional champion players seem to do–, when the time pressure is high they retreat back to the sub-optimal but better-practiced mono-rotational strategy.

These findings do well to illustrate how stable sub-optimal solutions may come to be employed, and how the cognitive system may resist certain types of change, despite the obvious advantages of alternative solutions. In the end, if a sub-optimal strategy is not only initially easier to comprehend, but easier to develop, then it may become ever more difficult to "fight one's own history".

5.5 STUDY 3: SUMMARY AND CONCLUSION

In this study, I examined how novice, regional champion, and global champion Tetris players dealt with the zoid rotation subtask, adopting one of two rationally adaptive strategies: the mono-rotational strategy, which reduces cognitive complexity at the cost of motor inefficiency; and the bi-rotational strategy, which pays the cost of cognitive complexity for reduced motor activity, saving precious milliseconds required for high-level task performance.

My findings showed that even the best players in the world could not reduce the complexity of the decisions required for their preferred strategy, leading me to my somewhat commonsense conclusion that "Hick's law bends with practice, *except when it cannot*." At times, it appears the cognitive machinery simply cannot reduce the complexity of the task any further, and thus expertise in that task may hit a *cognitive asymptote*.

CHAPTER 6
GENERAL DISCUSSION AND CONCLUSIONS

6.1 SUMMARY OF FINDINGS

The three studies presented comprise an investigation of the nature of expert performance in the complex, real-time Tetris task. In the first study, I used feature analysis (PCA) of laboratory gameplay data to extract performance features related to expertise. The most stable predictor of expertise was the "decide-moved-placed" feature, which relates to reaction times, movement efficiency, and "goodness of fit". Together with more task-specific features, I produced a multiple linear regression model which captured 65% of the variance when predicting player expertise after examining just a single level of early gameplay data (game difficulty level 1 only).

In the second study, I validated my laboratory model by using it to make skill estimates for novel data collected in-the-field at locally hosted Tetris tournaments. The model performed well at predicting qualifying round game scores (accounting for 54% of variance). Moreover, the model's skill estimates also predicted 83% of all tournament winners, compared to qualifying round score (58%, not significantly different from chance). This model exceeded expectations, in that it was validated twice over in its ability to accurately estimate player skill, and it did so by only looking at an array of immediate gameplay measures (with no knowledge of ultimate outcomes).

In the third study, I extended my field research to include players from the Classic Tetris World Championship, the best Tetris players in the known world. I drilled deeper into the available performance data and compared the strategies employed by novice, regional champions, and global champions for the zoid rotation subtask of Tetris, one simple strategy (the mono-rotational strategy) and one complex strategy (the bi-rotational strategy). This study produced two substantial findings. First, global champion players exhibited a "cognitive speed bump" (related to Hick's law) which I attributed to their use of the more complex, but ultimately more successful, bi-rotational strategy, whereas the regional champions experienced no such slowdown while using a simpler but less efficient strategy. This finding suggests that these players may have reached a cognitive asymptote on this subtask that cannot be further reduced, implying that the "bending" of Hick's law with expertise (A. R. Jensen, 1987; Logan et al., 2016) has reached its limit for this particular strategy. Second, I saw some evidence of regional players beginning to experiment with the bi-rotational strategy, but abandoning it when the game's time pressure grew too great, suggesting that adapting to the new, better strategy is difficult, and it is easier to revert to the established path of least resistance.

6.2 GENERAL DISCUSSION

The common theme throughout this investigation is the importance of reaction times and efficiency for understanding expert performance in complex, real-time domains– at every stage of expertise, and for every difficulty level of Tetris, measures related to decision speed significantly predict player expertise. This finding is reflected in all of the analyses and visualizations presented here, in that players with more skill were always faster and more efficient than players with lower skill, with one exception: expert players experience a "cognitive speed bump" associated with the adoption of a more complex strategy, while less-skilled players adopting a simpler strategy experienced no such slowdown. This finding indicates that for some strategies in complex tasks Hick's law cannot bend. Even world-class competitive Tetris players with years of practice playing at the edge of their own capabilities needed to take the time to evaluate a simple, but important, choice about rotation direction before they could take action.

My findings are compatible with Simon's "gelatin mold" metaphor for adaptive cognition: "The shape of a gelatin dessert cannot be predicted from the properties of gelatin, but only from the shape of the mold into which it was poured. [...] Behavior cannot be predicted from optimality criteria alone without information about the strategies and knowledge agents possess and their capabilities for augmenting strategies and knowledge by discovery or instruction." (Simon, 1992) In particular, I emphasize his latter statement in the quote, that we must understand an individual's strategies as well. My findings in this work seem to indicate that, indeed, we must look at the task to understand the shape a player's cognitive skill takes in the "gelatin mold", but we also see that the shape of the mold itself changes depending on the player's chosen strategies. In the case of the Fosbury flop, he reshaped his mold by eliminating the physical considerations of dangling legs from the high jump task, whereas the Tetris champions have reshaped their mold with additional decision-making slowdowns in favor of overall performance speed. In either case, though the shape of the task remains objectively intact, the players' chosen strategies have quite literally "changed the game."

6.3 FUTURE INVESTIGATIONS

6.3.1 ACT-R modeling

The "cognitive speed bump" for expert users of the bi-rotational strategy remains a curiosity, as the decision process involved, though more complex than the mono-rotational strategy, is relatively simple: from the current orientation, how far and in what direction should one rotate to achieve the desired orientation? In terms of expert ACT-R models– a computational cognitive modeling system (Anderson, 1993)– the memory structure required for the bi-rotational strategy would not necessitate more complexity in the decision-making process than for the mono-rotational strategy despite requiring additional information. Due to the nature of chunks as specified in the architecture, it is possible to simply store additional information in a given chunk, meaning that both the bi-rotational strategy and the mono-rotational strategy could follow the same simple routine: (1) retrieve the memory chunk to progress from a given initial orientation to a destination orientation, and (2) access the either 1 or 2 pieces of information stored therein to begin programming motor button presses. However, my empirical findings indicate that expert players' reaction times are indeed sensitive to the complexity of each zoid's rotation structure, suggesting that there is some additional cognitive cost not reflected by a simple memory retrieval strategy. Thus, a future ACT-R modeling study exploring this conflict of basic architectural assumptions and my empirical findings would likely prove illuminating.

6.3.2 Investigation with the Deliberate Practice approach

I have stated multiple times that the bi-rotational strategy appears to have a kind of "irreducible complexity". However, it may be the case that these world champion Tetris players simply are not practicing in a way that *could* result in these gains. Ericsson's (1993) Deliberate Practice Framework suggests that to continue advancing, one must both (1) bring oneself to the limit of one's ability such that one makes performance errors, and (2) get the appropriate feedback needed to encourage adjustments toward more favorable cognitive strategies. It is possible that, because of how fine-grained these reaction times are, it is difficult for a cognitive system– even a finely-tuned one– to detect these planning costs, which appear to take fewer than 50 ms. Using an experimental platform such as Meta-T, it would be a simple modification to incorporate millisecond-level reaction time feedback to motivated, extreme expert players (i.e., the CTWC population) to study whether it is possible to shave off this "cognitive speed bump" through continued practice with enhanced feedback.

6.3.3 Longitudinal studies

The studies in this investigation took a "sampling expertise" approach, in which I collected data from various cross-sections of the Tetris-playing population. Several of my findings relating to the acquisition of skill, strategy discovery, plateaus and stable suboptimal solutions, would be best served by a longitudinal approach, wherein I could monitor progress and changes in player performance, as well as conduct interviews and other qualitative approaches to produce a robust profile of player skill acquisition.

6.4 CONCLUSION

This work has focused on examining the features and theoretical structures associated with extreme expertise in the complex, real-time cognitive task of Tetris. To facilitate this research, I developed a platform for experimental research in Tetris, Meta-T, which allows for careful experimenter manipulation and offers high-fidelity behavioral performance logging.

My first study of undergraduate participants in the laboratory identified a set behavioral features– the "decide-move-placed" component– of interactive behavior in the Tetris task associated with player skill to be primarily those of speed and efficiency. These features are reliable for differentiating player skill across a wide array of expertise, and also across different levels of time pressure within the game, to the extent that experts are identifiable by my model "at a glance". This finding leads to my first conclusion: player skill can be reliably predicted by their efficiency of execution.

My second study validated my model of expertise on novel players in three real-world Tetris tournaments. By examining only a small subset of early-game performance, I was able to predict both player performance in the qualifying round of the tournament, and tournament match outcomes better than chance (83%), even predicting many major upsets, and the winner of every match of the tournament finals and run-offs. This two-fold validation leads to a second conclusion: when it comes to estimating skill by either single outcomes or behavioral process, it pays to bet on process.

Finally, in my third study I found that by examining the strategies employed by extreme experts in Tetris with fine detail, we begin to see the definition of the human cognitive system in greater relief. I find that even world champion Tetris players remain impacted by Hick's law for decisions about how to rotate zoids in the game– a "cognitive speed bump". This finding indicates that there may be some lower limit to the ability of human expertise to adapt to and reduce a task's decision space, leading me to two conclusions: (1) "Hick's law bends with practice, *except when it cannot*", and (2) "both task and strategy together define the mold into which human expertise may shape itself."

My approach has been to take a body of complex human behavioral data, and then iteratively "zoom in and unfold." The work presented here details this approach to exploring and validating questions of performance expertise in a given complex task domain. I find that the behavior itself is more fundamental to understanding expertise than its associated outcomes, particularly when attempting to define the limits of human cognition within complex tasks. A great deal of further work can be done to test the limits of these findings, but as they stand presently, I believe they serve to embellish the cognitive portrait of human expertise in the complex, real-time task domain of Tetris.

REFERENCES

Anderson, A. F. & Bavelier, D. (2011). Action game play as a tool to enhance perception, attention, and cognition. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 307–329). Charlotte, North Carolina: Information Age Publishing.

Anderson, J. R. [John R.]. (1987). Skill acquisition - compilation of weak-method problem solutions. *Psychological Review*, *94*(2), 192–210. doi:10.1037//0033-295X.94.2.192

Anderson, J. R. [John R]. (1993). *Rules of the mind*. Hillside, NJ: Psychology Press.

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

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, *129*(3), 387–398. doi:{10.1016/j.actpsy.2008.09.005}

Bryan, W. L. & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. *Psychological Review*, *4*(1), 27–53. doi:10.1037/h0073806

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chase, W. G. & Ericsson, K. A. (1982). Skill and working memory. *The Psychology of Learning and Motivation*, *16*, 1–58. doi:10.1016/S0079-7421(08)60546-0

Chase, W. G. & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. doi:{10.1016/0010-0285(73)90004-2}

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

Classic Tetris. (2016). Classic Tetris World Championship – Final Round. Retrieved March 12, 2017, from https://youtu.be/DdfRQjb5o9k

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally.

de Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton.

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 An-*

nual Conference of the Cognitive Science Society (pp. 2709–2014). Austin, TX: Cognitive Science Society.

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science*, *18*(6), 321–326.

Ehret, B. D., Gray, W. D., & Kirschenbaum, S. S. (2000). Contending with complexity: developing and using a scaled world in applied cognitive research. *Human Factors*, *42*(1), 8–23.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*(4448), 1181–1182.

Ericsson, K. A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.

Fahey, C. (2007). Tetris.

Fahey, C. P. (2015). Tetris AI. [Online; accessed 2015-Jan-30].

Fitts, P. M. & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks Cole.

Fu, W. T. & Gray, W. D. (2004). Resolving the paradox of the active user: Stable subuptional performance in interactive tasks. *Cognitive Science*, *28*(6), 901–935. doi:10.1016/j.cogsci.2004.03.005

Gabillon, V., Ghavamzadeh, M., & Scherrer, B. (2013). Approximate dynamic programming finally performs well in the game of Tetris. In *Advances in neural information processing systems* (Vol. 26, pp. 1754–1762).

Gigerenzer, G. & Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585–612. doi:10.1146/annurev.psych.49.1.585. eprint: http://dx.doi.org/10.1146/annurev.psych.49.1.585

Gray, W. D. (2017a). Games-XP: Action Games as Cognitive Science Paradigm. *Topics in Cognitive Science*, *9*(2). doi:10.1111/tops.12260

Gray, W. D. (2017b). Plateaus and asymptotes: Spurious and real limits in human performance. *Current Directions in Psychological Science*, *26*(1), 59–67. doi:10.1177/0963721416672904

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world performance. *Human-Computer Interaction*, *8*(3), 237–309.

Gray, W. D. & Lindstedt, J. K. (2016). Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance. *Cognitive Science*, *[Early View]*, 1–33. doi:10.1111/cogs.12412

Gray, W. D. & Salzman, M. C. (1998). Damaged merchandise? a review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, *13*(3), 203–261.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: a rational analysis approach to resource allocation for interactive behavior. *Psychological review*, *113*(3), 461–82. doi:10.1037/0033-295X.113.3.461

Green, C. S. & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537.

Green, C. S. & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception And Performance*, *32*(6), 1465–1478.

Haier, R. J., Karama, S., Leyba, L., & Jung, R. E. (2009). Mri assessment of cortical thickness and functional activity changes in adolescent girls following three months practice on a visual-spatial task. *BMC Research Notes*, *8*(174).

Haier, R. J., Siegel, B. V., MacLachlan, A., Soderling, E., Lottenberg, S., & Buchsbaum, M. S. (1992). Regional glucose metabolic changes after learning a complex visuospatial/motor task: a positron emission tomographic study. *Brain Research*, *570*, 134–143.

Hays, R. T., Jacobs, J. W., Prince, C., & Salas, E. (1992). Flight simulator training effectiveness: a meta-analysis. *Military Psychology*, *4*(2), 63–74. doi:10.1207/s15327876mp0402_1. eprint: http://www.tandfonline.com/doi/pdf/10.1207/s15327876mp0402_1

Hick, W. E. (1952). On the rate of gain of information. *The Quarterly Journal of Experimental Psychology*, *4*, 11–26.

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

Huang, J., Yan, E., Cheung, G., Nagapan, N., & Zimmermann, T. (2017). Master maker: understanding gaming skill through practice and habit from gameplay behavior. *Topics in Cognitive Science*. doi:10.1111/tops.12251

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833.

Jensen, A. R. [A. R.]. (1987). Individual differences in the Hick paradigm. In P. A. Vernon (Ed.), *Speed of information processing and intelligence* (pp. 101–175). Norwood, NJ: Ablex.

Jensen, A. R. [Arthur R.]. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93–132). New York: Springer.

Jensen, A. R. [Arthur R.]. (1998). *The g factor: the science of mental ability.* Human evolution, behavior, and intelligence; ISSN: 1063-2158 (Print). Santa Barbara, CA: Praeger Publishers/Greenwood Publishing Group.

Johnson, R. C., McClearn, G. E., Yuen, S., Nagoshi, C. T., Ahern, F. M., & Cole, R. E. (1985). Galton's data a century later. *American Psychologist*, *40*(8), 875–892. doi:{10.1037/0003-066X.40.8.875}

Keehner, M. (2011). Spatial Cognition Through the Keyhole: How Studying a Real-World Domain Can Inform Basic Science-and Vice Versa. *Topics in Cognitive Science*, *3*(4), 632–647. doi:{10.1111/j.1756-8765.2011.01154.x}

Keehner, M., Tendick, F., Meng, M., Anwar, H., Hegarty, M., Stoller, M., & Duh, Q. (2004). Spatial ability, experience, and skill in laparoscopic surgery. *American Journal of Surgery*, *188*(1), 71–75. doi:{10.1016/j.amjsurg.2003.12.059}

Kellman, P. J. & Massey, C. M. (2013). Perceptual learning, cognition, and expertise. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 58, 117–165). doi:{10.1016/B978-0-12-407237-4.00004-9}

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

Kirsh, D. & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, *18*(4), 513–549. doi:10.1207/s15516709cog1804{_}1

Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft*, *8*(1), 5–19.

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

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

Logan, G. D., Ulrich, J. E., & Lindsey, D. R. B. (2016). Different (key)strokes for different folks: How standard and nonstandard typists balance Fitts' law and Hick's law. *Journal of Experimental Psychology-Human Perception and Performance*, *42*(12), 2084–2102.

Maglio, P. & Kirsh, D. (1996). Epistemic action increases with skill. In *18th annual meeting of the cognitive science society* (pp. 391–396). Cognitive Science Society.

Martin-Gutierrez, J., Luis Saorin, J., Martin-Dorta, N., & Contero, M. (2009). Do video games improve spatial abilities of engineering students? *International Journal of Engineering Education*, *25*(6, SI), 1194–1204.

Minitab. (2017). What is a scree plot? Retrieved February 6, 2017, from http://support.minitab.com/en-us/minitab/17/topic-library/modeling-statistics/multivariate/principal-components-and-factor-analysis/what-is-a-scree-plot/

Newell, A. & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, N. J.: Lawrence Erlbaum Associates.

Okagaki, L. & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: gender effects in late adolescence. *Journal of Applied Developmental Psychology*, *15*(1), 33–58.

Proctor, M. D., Bauer, M., & Lucario, T. (2007). Helicopter flight training through serious aviation gaming. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, *4*(3), 277–294. doi:10.1177/154851290700400305. eprint: http://dms.sagepub.com/content/4/3/277.full.pdf+html

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*(2), 359–379. doi:10.1037/a0029082

Reingold, E. M. & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook of eye movements* (pp. 528–550). Oxford University Press.

Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory using EPAM-IV. *Psychological Review*, *102*(2), 305–330. doi:{10.1037//0033-295X.102.2.305}

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information-processing .1. detection, search, and attention. *Psychological Review*, *84*(1), 1–66.

Shepard, R. N. & Metzler, J. (1971). Mental rotation of 3-dimensional objects. *Science*, *171*(3972), 701–703. doi:{10.1126/science.171.3972.701}

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information-processing .2. perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*(2), 127–190.

Sibert, C., 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), 1–21. doi:10.1111/tops.12225

Simon, H. A. (1992). What is an "explanation" of behavior? *Psychological Science*, *3*(3), 150–161.

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

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

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

Şimşek, Ö., Algorta, S., & Kothiyal, A. (2016). Why most decisions are easy in tetris—and perhaps in other sequential decision problems, as well. In *Proceedings of the 33rd international conference on machine learning* (pp. 1757–1765). JMLR.

Smith, B. (2014). *Tricks of the classic nes tetris masters*. San Bernardino, CA: CreateSpace Independent Publishing Platform.

Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. J. F. Michelene T.H. Chi Robert Glaser (Ed.), *The nature of expertise* (pp. 71–128). Hillside, NJ: Psychology Press.

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

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

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

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

Thompson, J. J., McColeman, C. M., Stepanova, E. R., & Blair, M. R. (2017). Using video game telemetry data to research motor chunking, action latencies, and complex cognitive-motor skill learning. *Topics in Cognitive Science*, *9*(2), 1–18. doi:10.1111/tops.12254

Wikipedia. (2016). Classic Tetris World Championship — Wikipedia, The Free Encyclopedia. Retrieved February 6, 2017, from https://en.wikipedia.org/w/index.php?title=Classic_Tetris_World_Championship&oldid=746037457

APPENDIX A

BEHAVIORAL FEATURES OF TETRIS

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

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

Actions: keypresses made by the player.

rots - Rotations. The number of zoid-rotations performed in an episode.

trans - Translations. The number of times the zoid was moved left or right– translated– in an episode.

prop_u_drops - Proportion of user drops. The proportion of top-to-bottom movement that was intentionally dropped by the player in an episode. 1 indicates the player dropped the zoid the full length of the screen, and 0 indicates they never dropped it at all.

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

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

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

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

initial_lat - Initial latency. Time elapsed in milliseconds from the start of the episode until the first keypress.

drop_lat - Drop latency. Time elapsed in milliseconds from the start of the episode until the player first drops the zoid.

avg_lat - Average latency. The mean time between all keypresses in an episode.

resp_lat - Response latency. The time from the start of the episode in milliseconds until *either* the zoid is first dropped *or* the zoid is locked into its final position. This is a more clear measure of actual response time than drop_lat, as it accounts for episodes in which the zoid was never dropped.

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

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

mean_ht, max_ht, min_ht - Mean, Maximum, and Minimum height. The mean, maximum, and minimum height among all 10 columns in the pile. Note that only maximum height needs to exceed 20 to result in a game over.

cd_1 - cd_9 - Column difference (0,1)-(8,9). 9 features representing the difference in height between each successive pair of columns. Positive values imply a raise in height from left to right, while negative implies the opposite.

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

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

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

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

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

lumped_pits - Lumped pits. A measure of pits considering all adjacent groups of pits to be identical. Thus four isolated pits in the pile would have more weight than one 2x2 cluster of pits.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table A.1: Feature loadings from the principal component analysis. Five features are omitted (cd_2 - cd_6) as they did not contribute to any of the selected components.

	Comp.1	Comp.2	Comp.3	Comp.4
Label	entropy	structure	decide-move-placed	pivoting/recovery
% Variance	25.6	14.0	9.6	5.1
rots			-0.348	0.325
trans			-0.275	0.308
prop_u_drops	-0.105		0.314	
min_rots_diff			-0.350	0.299
min_trans_diff			-0.306	0.259
initial_lat			-0.180	-0.407
drop_lat			-0.390	
avg_lat			-0.193	-0.483
resp_lat			-0.452	-0.111
mean_ht	0.305			
min_ht	0.255	-0.218		
max_ht	0.290			
cd_1		0.192		
cd_7				
cd_8				
cd_9		-0.201		
max_diffs		0.260		
pits	0.272	-0.177		
pit_depth	0.273	-0.150		
pit_rows	0.278	-0.173		
lumped_pits	0.265	-0.180		
wells	0.119	0.373		
deep_wells	0.115	0.372		
cuml_wells	0.107	0.365		
max_well	0.112	0.367		
jaggedness	0.113	0.309		
col_trans	0.273	-0.161		
row_trans	0.291			
pattern_div	0.199			
weighted_cells	0.295			
landing_height	0.285			
matches			0.137	0.324
d_max_ht			-0.133	-0.278
d_pits				-0.133

APPENDIX B

EXAMINING TOURNAMENT MATCH-UPS AND RANKINGS

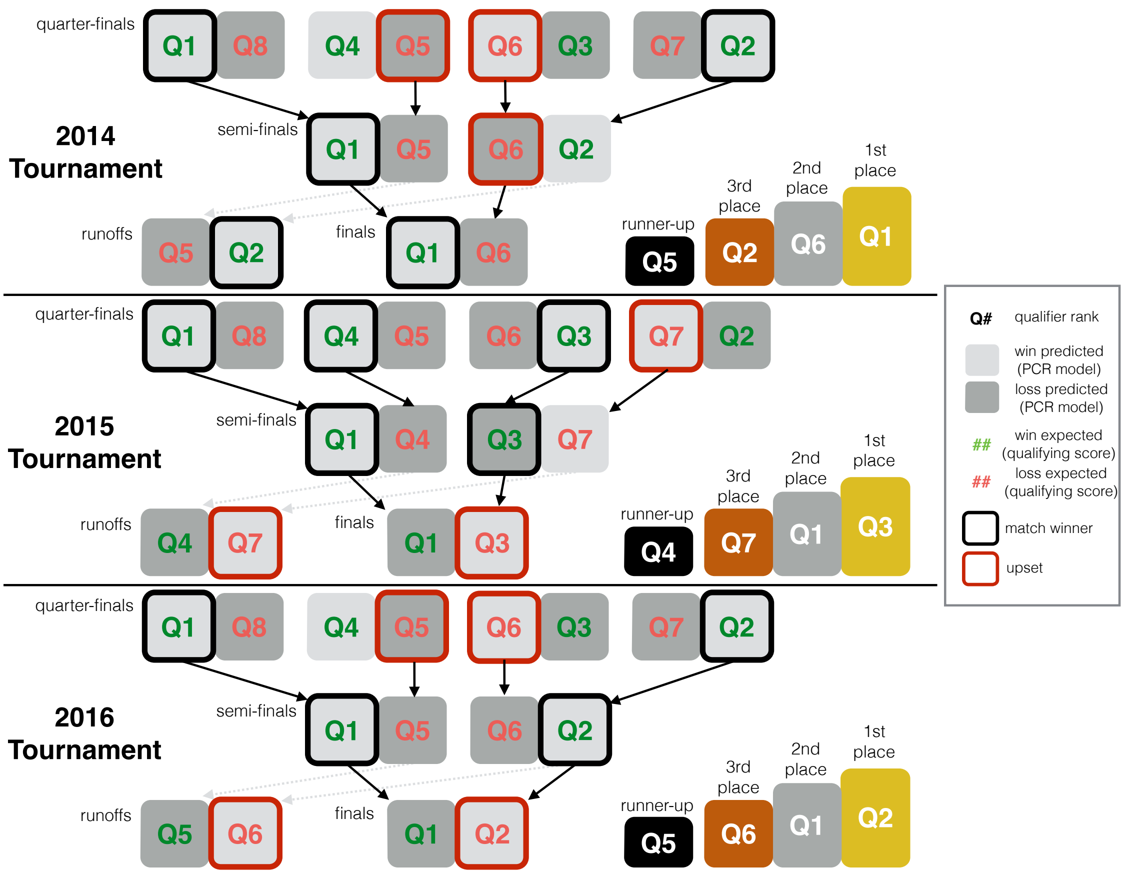
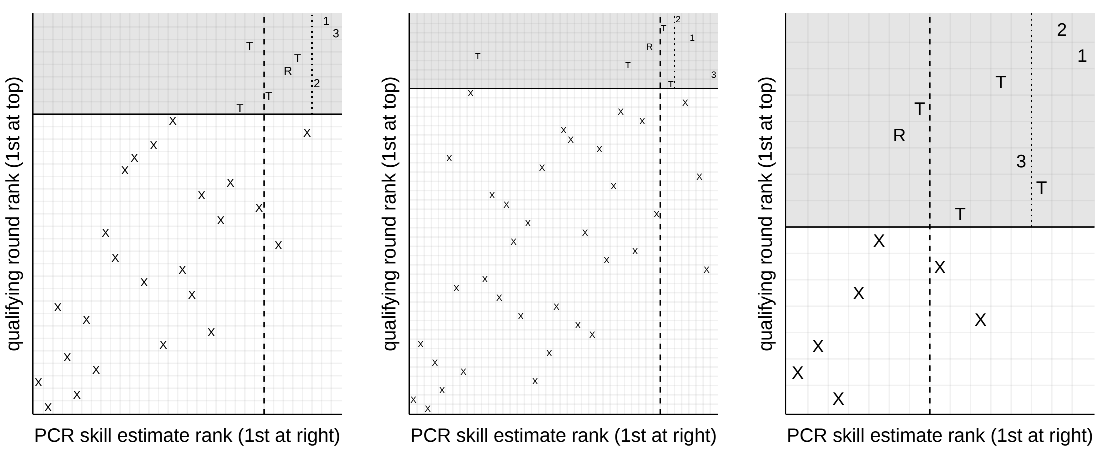


Figure B.1: Table of tournament match-up progression for each year, illuminating when upsets occurred (red border) and whether qualifying round score (green text for a win, red text for a loss) and/or PCR skill estimates (light fill for win, dark fill for a loss) predicted the winner of the match. Notably, low-ranking qualifying round players often place in the top three.

Table B.1: Tournament rankings comparing champion players' tournament outcomes to (1) rankings based on qualifying round high scores, and (2) rankings based on the PCA model's skill estimates based on level 1 gameplay data. Asterisks indicate an upset; i.e., a player who ranked lower initially in their respective tournament, but ultimately claimed a championship prize. Bold values indicate perfect predictions of tournament champion rankings.

		Rank Estimates		Distance (in ranks)	
Year	Final Standing	Qualifiers	PCR Model	Qualifiers	PCR Model
2014	Gold (1st)	1st	2nd	0	1
	Silver (2nd)	6th*	3rd	4	1
	Bronze (3rd)	2nd	1st	1	2
2015	Gold (1st)	3rd	2nd	2	1
	Silver (2nd)	1st	3rd	1	2
	Bronze (3rd)	7th*	1st	4	2
2016	Gold (1st)	2nd	1st	1	0
	Silver (2nd)	1st	2nd	1	0
	Bronze (3rd)	6th*	4th*	3	1
			Total distance:	17	10



(a) (b) (c)

Figure B.2: Tournament standings as determined by both qualifying round high score (QHS) and Principle Component Regression (PCR) skill predictions. Subfigures (a), (b), and (c) highlight where QHS and PCR disagree about a player's skill level. Letter indicates player type: 1, 2, and 3 are the champions who won a cash prize, R indicates the runner-up, T indicates tournament player that did not win, and X indicates players excluded from the tournament. The shaded region highights players with the top 8 qualifying round scores (i.e., those who participated in the tournament), the dashed vertical line shows the top 8 players according to their PCR skill estimate, and the dotted vertical line segment in the shaded region indicates the PCR model's estimate of the best 3 players in the tournament. The top-left and bottom-right quadrants show the sometimes substantial disagreement on which players should be included in the tournament between the qualifying round score ranking and the PCR skill estimate ranking.

APPENDIX C

EXAMINING INITIAL REACTION TIMES

Table C.1: Player count (n), mean (m), and standard deviation (sd) for initial reaction times (in milliseconds) by game level and zoid type for global champion players. Time to impact (tti) represents the maximum amount of time (seconds) available per episode at the given game level before the zoid naturally collides with the bottom of the game space.

global champions								
			static zoids		flipping zoids		rotating zoids	
level	tti	n	m	sd	m	sd	m	sd
0	16.00	10	305	123	176	55	233	51
1	14.33	10	282	183	193	65	209	62
2	12.67	10	378	121	211	94	227	42
3	11.00	10	370	153	206	61	188	79
4	9.33	10	263	112	200	37	205	52
5	7.67	10	285	84	167	55	197	47
6	6.00	10	214	79	174	38	189	46
7	4.33	10	176	33	161	20	176	25
8	2.67	10	132	16	119	15	147	13
9	2.00	10	100	14	109	12	136	16
10	1.67	10	87	15	105	7	133	12
11	1.67	10	86	13	103	10	133	22
12	1.67	10	85	19	104	8	138	19
13	1.33	10	75	10	91	10	131	13
14	1.33	9	74	8	92	8	124	18
15	1.33	9	72	13	93	10	125	20
16	1.00	9	58	8	78	9	124	21
17	1.00	9	55	6	74	9	119	16
18	1.00	5	59	3	79	7	130	13
19	0.67	3	35	1	54	1	75	12

Table C.2: Player count (n), mean (m), and standard deviation (sd) for initial reaction times (in milliseconds) by game level and zoid type for regional champion players. Time to impact (tti) represents the maximum amount of time (seconds) available per episode at the given game level before the zoid naturally collides with the bottom of the game space.

regional champions

			static zoids		flipping zoids		rotating zoids	
level	tti	n	m	sd	m	sd	m	sd
0	16.00	10	296	142	185	101	222	119
1	14.33	10	165	93	157	65	176	91
2	12.67	10	190	111	181	86	172	77
3	11.00	10	181	72	167	73	196	89
4	9.33	10	152	63	153	68	180	85
5	7.67	10	144	87	154	68	151	64
6	6.00	10	142	62	119	44	135	45
7	4.33	10	128	40	116	45	118	35
8	2.67	10	97	29	96	31	90	25
9	2.00	6	68	17	66	14	77	11
10	1.67	5	65	12	54	7	60	8
11	1.67	4	61	19	57	6	55	6
12	1.67	4	55	8	51	5	56	5
13	1.33	4	51	1	41	11	45	11
14	1.33	3	44	9	42	6	43	9
15	1.33	2	52	5	41	2	42	6

Table C.3: Player count (n), mean (m), and standard deviation (sd) for initial reaction times (in milliseconds) by game level and zoid type for novice players. Time to impact (tti) represents the maximum amount of time (seconds) available per episode at the given game level before the zoid naturally collides with the bottom of the game space.

novice players

			static zoids		flipping zoids		rotating zoids	
level	tti	n	m	sd	m	sd	m	sd
0	16.00	10	635	119	521	159	515	120
1	14.33	10	480	144	509	119	519	82
2	12.67	8	477	103	482	75	491	70
3	11.00	6	407	181	397	84	398	114
4	9.33	2	445	10	404	45	384	2
5	7.67	2	363	0	360	10	361	39
6	6.00	2	217	100	248	61	274	15
7	4.33	1	196	–	206	–	196	–
8	2.67	1	102	–	113	–	111	–

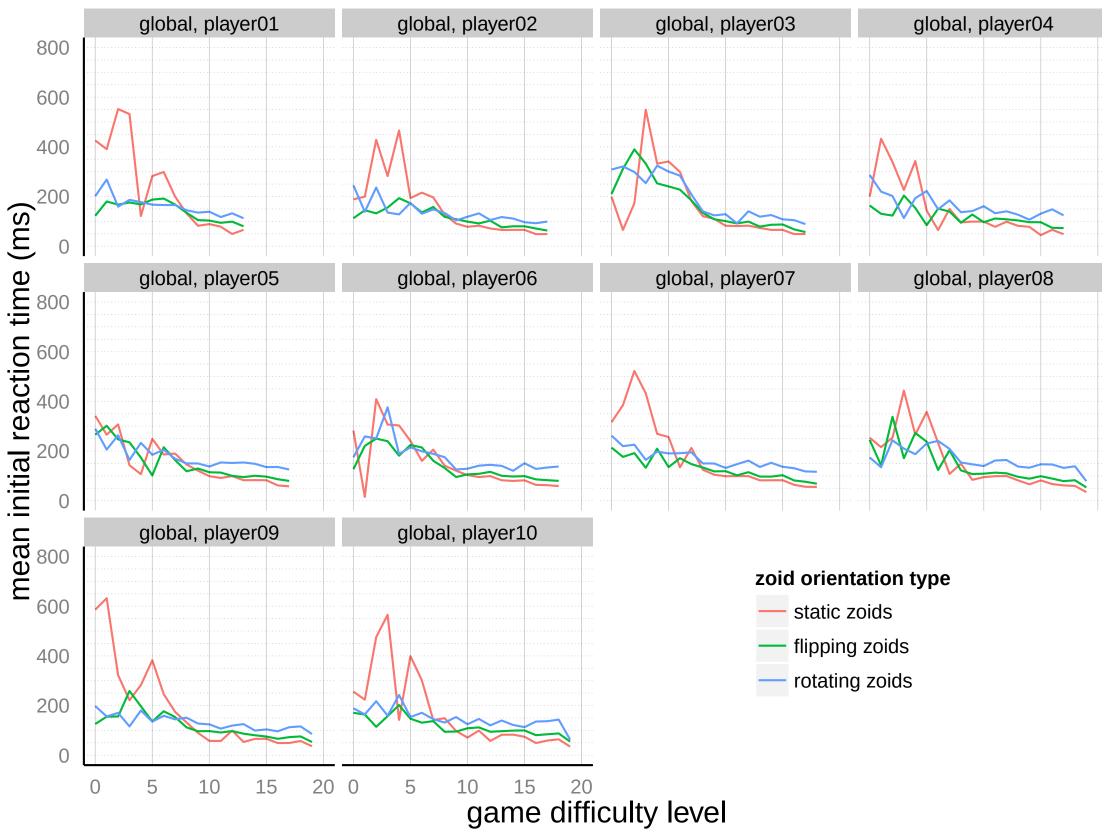


Figure C.1: Initial reaction times for global champions. Shows the mean initial reaction times (y-axis) for each zoid rotation type (color) by game difficulty level (x-axis) for each global champion player. Players are sorted within each skill category in ascending order of the highest game difficulty level reached (left to right, row-wise).

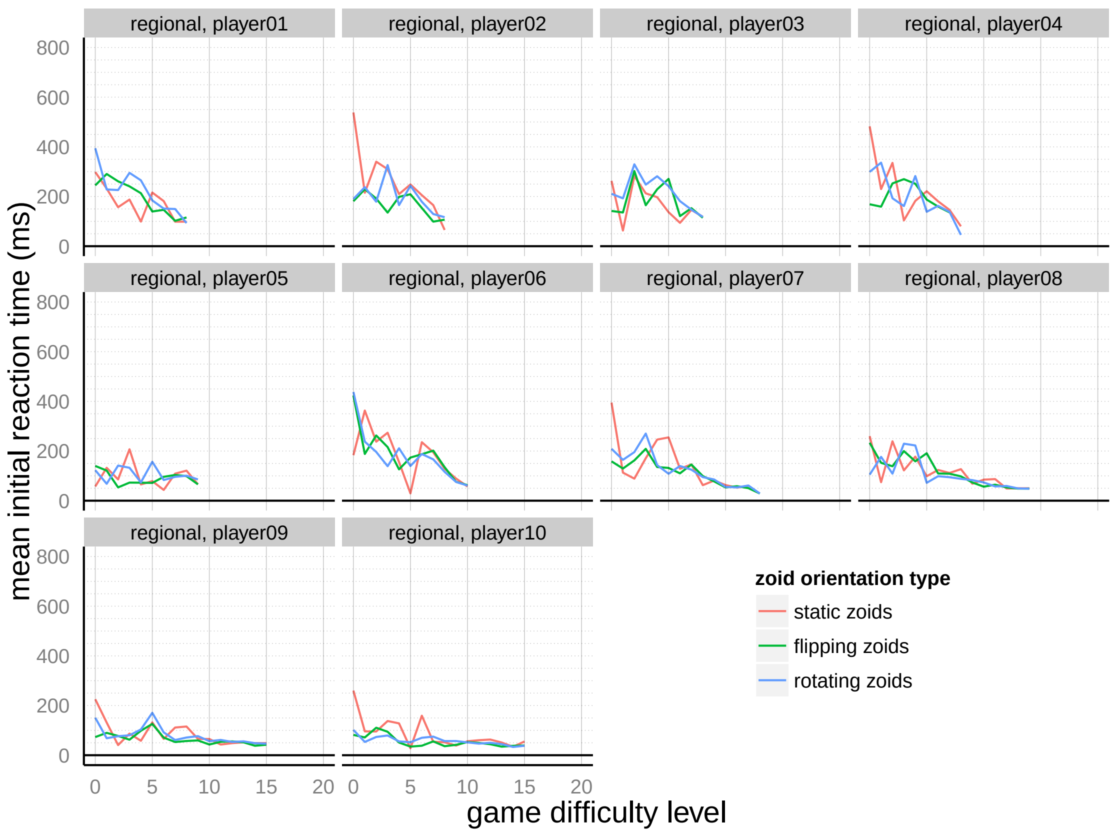


Figure C.2: Initial reaction times for regional champions. Shows the mean initial reaction times (y-axis) for each zoid rotation type (color) by game difficulty level (x-axis) for each regional champion player. Players are sorted within each skill category in ascending order of the highest game difficulty level reached (left to right, row-wise).

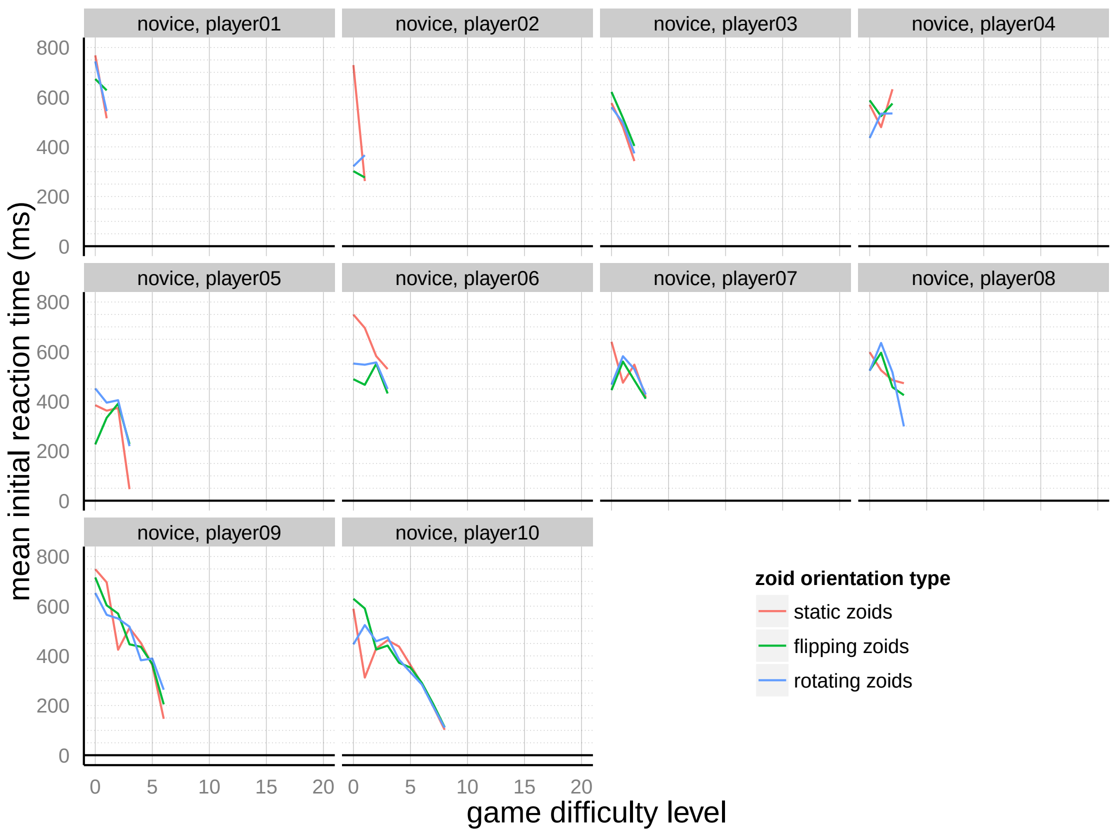


Figure C.3: Initial reaction times for novices. Shows the mean initial reaction times (y-axis) for each zoid rotation type (color) by game difficulty level (x-axis) for each novice player. Players are sorted within each skill category in ascending order of the highest game difficulty level reached (left to right, row-wise).

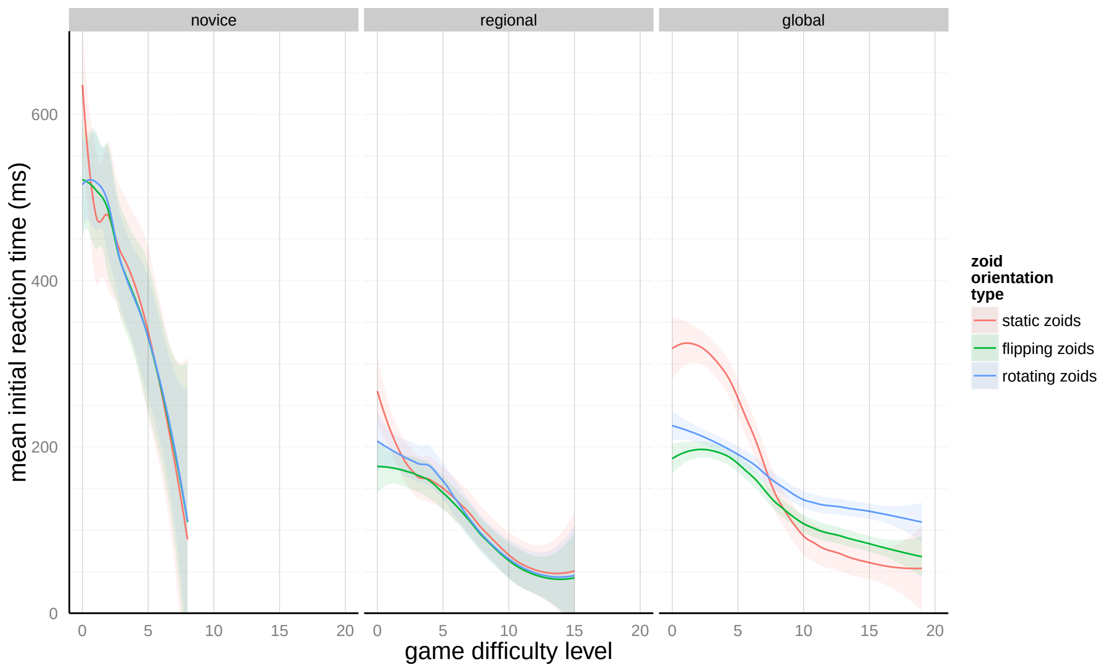


Figure C.4: Mean initial reaction times (y-axis) for each category of zoid (color) by game difficulty level (x-axis), averaged for each group of player skill (columns). The curve is the result of a local polynomial regression (LOESS curve fitting), and the colored ribbons for each group represents the 95% confidence interval of the estimate. Of particular interest is the difference between the relationship of reaction times for each zoid category for regional champion players (middle column) and global champion players (right column): as time pressure increases, regional players appear to converge, whereas global players consistently exhibit different initial reaction times for each zoid rotation type.