Master Maker: Understanding Gaming Skill through Practice and Habit from Gameplay Behavior

Jeff Huang, Eddie Yan, Gifford Cheung, Nachiappan Nagappan, Thomas Zimmermann
Brown University, University of Washington, University of Washington, Microsoft Research, Microsoft Research

Abstract

The study of expertise is difficult to do in a lab environment due to the challenge of finding people at different skill levels and the lack of time for participants to acquire mastery. In this paper, we report on two studies that analyze naturalistic gameplay data using cohort analysis to better understand how skill relates to practice and habit. Two cohorts are analyzed, each from two different games (*Halo Reach* and *StarCraft 2*). Our work follows skill progression through 7 months of *Halo* matches for a holistic perspective, but also explores low-level in-game habits when controlling game units in *StarCraft 2*. Players who played moderately frequently without long breaks were able to gain skill the most efficiently. What set the highest performers apart was their ability to gain skill more rapidly and without dips compared to other players. At the beginning of matches, top players habitually warmed up by selecting and re-selecting groups of units repeatedly in a meaningless cycle. They exhibited unique routines during their play that aided them when under pressure.

Introduction

Bruce Lee, the famed martial artist, once said *"I fear not the man who has practiced 10,000 kicks once, but I fear the man who has practiced one kick 10,000 times."* He understood that mastery requires dedication to one routine, and implied that the skill one gains comes from persistent practice. What is it about the habits one develops when practicing a kick that makes them someone to fear? Unfortunately, this is not simple to measure or analyze, as physical practice is hard to capture and encode.

In scientific methodologies, tasks are assigned to participants in lab studies—controlled environments where a single factor can be manipulated and the outcomes observed. However, the artificiality of lab tasks along with the limited time for participants to perform can restrict what we can discover about long-term skill progression and habit

[0] This paper extends work published in Huang, Zimmermann, Nagappan, Harrison, and Phillips (2013); Yan, Huang, and Cheung (2015)

development. This paper adopts competitive games as an environment for large-scale analysis of skill, which do not have the same restrictions of typical lab settings. The underlying investigation is about how skill develops under the lens of individual practice and habits.

Careful analysis of data gathered from competitive gameplay offer a chance to look at behavioral patterns of the players, including the most skilled players of the game. This data is both naturalistic, as it is captured broadly by the gaming server or players themselves, and is available at a large scale, covering thousands of players. This paper emphasizes skillful players in *Halo Reach* and *StarCraft 2*, two games from different genres. *Halo Reach* is a first-person shooter where players engage each other in weapon-based combat. *StarCraft 2* is a fast-paced strategy game where a player controls up to two hundred space soldiers, vehicles, and alien creatures in order to defeat the opposing team's army. Both of these games are played competitively and assign players a quantified skill level. There are numerous factors affecting what makes people play at the highest skill levels, including tactics, reaction time, and game knowledge, but we focus more abstractly on higher-level behaviors. Cohort analysis allows us to group players by their start date to learn about practice and progression, as well as form cohorts of players in the same skill level so that we can compare how higher skilled players distinguish themselves. These behaviors illustrate how practice and habit relate to skill, particularly for the top players of these games.

Halo Reach is used for a study of practice and progression. Being able to understand practice over time for thousands of players can provide insight about how players improve their skill. Also, we can review early matches from top players' careers to see how they differ in their evolution. Some literature concerning practice emphasizes "deliberate practice" (Ericsson, Krampe, and Tesch-Römer (1993)), when people conduct intense, focused activities in private to improve skill. These practice sessions are highly focused and require rest and recuperation between sessions. However, in online video games like *Halo Reach*, players who improve are not usually "private" in the same way that a violin player practices alone. Instead, the online player is matched against other players who play for their own reasons. Thus, competitive video game sessions are similar in form to repeated sessions of competitive chess matches. This kind of tournament play is not covered by the original definition of "deliberate practice," but has been shown to still be effective. Gobet and Campitelli (2007) examines Chess, where group practice (including tournament play) was statistically more effective at improving a person's rating than deliberate practice (as defined by individual study). Thus, like Ericsson et al., we are interested in the relationship between intense activity and rest; and, like Gobet et al., we study an alternate kind of "practice" that is effective in increasing skill. The frequency and consistency of competitive matches in our *Halo Reach* dataset provides a particularly detailed perspective into this interest.

StarCraft 2 provides in-game insight into an even finer resolution analysis of habit and individualism in matches. Replays (automatically saved logs of game state) of *StarCraft 2* games include information about how players group and control their units under a feature we will call "unit groups." The mechanism is straight forward: if a player has five soldiers selected, she can assign those five into a group number (e.g. group 3). Later in the game, she will turn her attention to those five units and give them specific instructions. To do so, she presses the number 3 on her keyboard to re-select her original five soldiers. Using the entire row of number keys to assign, re-select, adjust, and re-assign unit groups, skilled players can simultaneously control hundreds of units in time-pressured situations. In essence, unit groups are a mechanism to efficiently manage units in the game, which makes them an appropriate manifestation for studying habits in gameplay. We identify a phenomenon where habitual game actions are practiced in relaxed situations to be relied upon later in high pressure situations. Players who have refined their unit group habits are higher skilled and multi-task better, particularly in time-pressured situations. However, these habits are truly individual; while there are fundamental differences between skill levels, there are even stronger differences between individuals.

Our main contributions are two case studies in gameplay behavior to show that mastery of a game takes place through sustained and intense practice that can result in bursts of improvement, and can manifest as deeply engrained, individualized habits that are available as second-nature, expert maneuvers when players are under pressure.

Related Work

This section situates the analyses presented later in this paper within the context of related work. It includes reviews of data analysis in games, studies in video game expertise, and literature on habit and practice.

Data Analysis in Games

Many modern multiplayer video game titles ship with the ability to record gameplay data and incorporate matchmaking features for players.

StarCraft 2 replay data has been studied previously to identify which characteristics explain a player's skill level (Thompson, Blair, Chen, and Henrey (2013)). Several features that were relevant to player skill were identified, including the "Perception-Action Cycle" (PAC), and actions per minute (APM). In this paper, we narrow down on the actions encompassed in the APM metric from an alternative perspective. In addition to the variation of replay data with player skill levels studied in Thompson et al., we also consider the variation between players at a similar skill level.

Weber and Mateas (2009) demonstrated a process for opponent modeling through data mining by analyzing *StarCraft: Brood War* replays. Their work transformed the replay logs into vectors representing the time each unit or building type was created. A model of the opponents' activity is generated from these vectors to predict what they will do next. In contrast, this paper demonstrates a process of identifying players from anonymous replays based on unit grouping habits which can enhance opponent modeling as it enables the focused study of a particular player's game to predict strategy.

Our study of *Halo Reach* adopts a similar methodology as other studies of multiplayer *Halo*, which use gameplay records and either player surveys or interviews. Mason and Clauset (2013) use the same data source as us—the *Halo Reach* multiplayer records, and supplement them with a survey. They find that players with more friends (both online and offline friends) on their team perform better individually, while also performing better as a team. Xu, Cao, Sellen, Herbrich, and Graepel (2011) take a different approach to studying social motivations, where they aim to understand the social relationships between *Halo 3* players through the gameplay records and interviews. They found that players were well aware of who they played with, and rather than only playing to win, also sought to enjoy the social experience of the game. While we employ similar methods and data as these two studies, our focus is on player skill and the change in expertise rather than social relationships.

Prior work on characterizing *Project Gotham Racing 4* (Hullett, Nagappan, Schuh, and Hopson (2012)) explained the diverse and extensive amount of data that is collected due to the constantly connected nature of the game consoles. The results of this analysis helped provide a better understanding of the differences between long-term and short-term players, the choices they make, their retention and the extent to which various options in the game are utilized (in this case for example, the type of track, vehicle class, or weather conditions). This led to recommendations for ways to reduce development costs by eliminating unused or unpopular options and to help keep new players engaged.

Studies of Video Game Expertise

Case studies situate the researcher inside the gaming experience, either as observers or as players themselves. Reeves, Brown, and Laurier (2009) take an ethnomethodological approach to analyzing expertise in the first-person shooter, *Counter-Strike*, by watching an expert *in situ*. They find that expert play involves an understanding of the terrain and a sense of where other players are in the environment. Reeves et al. also suggest regarding gameplay holistically, as it does not make sense when taken in pieces. Another researcher, Hock-koon (2012), becomes an expert himself in the game *Alien vs. Predator*. He rigorously kept a journal of his training and lessons learned, and developed a theory of elliptical learning. Hock-Koon argues that learning encompasses multiple levels of understanding for a single mechanism in the game. In contrast to these case studies that put the researcher into the game, we step back and look at aggregate data from thousands of players to seek generalizable patterns.

Using methods similar to ours, Stafford and Dewar (2014) track 854,064 players' scores in an online game called *Axon*. They notice that over time, scores generally increase but the top players start with a lead compared to other players and continue to grow this lead with each additional match. While we also examine skill trajectory in this paper, we look at other factors that impact skill in *Halo Reach*, and analyze within-game metrics of unit group usage in *StarCraft 2*.

Habit and Practice

People acquire skill from practice in a broad range of tasks. An early and well-known study is that of American telegraphers by Bryan and Harter (1897). The sending and receiving rates, measured in characters per minute, are plotted over time to produce figures showing different rates of acquiring expertise among the two tasks, and particularly a plateau in the middle of the receiving plot. This plateau has incited discussion in follow-up work, where the original authors believed that multiple practice curves existed (Bryan and Harter (1899)), characterized by the two separate skills of mapping Morse code into letters and predicting words from initial letters. However, Keller counters in a later study that there is no plateau effect (Keller (1958)), citing unpublished studies by Tulloss, where "there is no sign of a plateau in any of the Tulloss curves." In many cases, the plateaus may actually be instances of artificial asymptotes due to artifacts or poor system design (Gray and Lindstedt (in press)).

Other studies in software have looked at motivators for skill acquisition and differences between experts and non-experts in searching the Web. In a study of non-programmers playing a game that teaches programming, Lee and Ko (2011) found that participants completed more levels of the game, and thus acquired additional skill in programming if the goal was framed in terms of helping a personable robot rather than an inanimate terminal. In another study, White, Dumais, and Teevan (2009) examined experts and non-experts' behavior over a 3 month period of search logs. They found that expert searchers differed in terms of query vocabulary, sites they visited, and patterns of search behavior. The authors were also able to predict the expertise of a user with modest success; computer science experts were found to be easier to predict than medicine, finance, or legal experts. In our work, we focus less on predicting skill, and more on explaining factors that affect skill.

Mining Competitive Gameplay Data

Players of competitive games strive to excel in a structured environment, where those who find an edge are rewarded with wins and better ratings. The growing competitive landscape for video games has motivated game developers to implement features in support of professional and amateur players who play not just for entertainment, but to improve their skills and strategies. In competitive games, these features include sophisticated ranking systems that automate player matchmaking in online ranked matches and the ability to record sequences of commands performed by each player as replays. The existence and popularity of replays means that for many matches, a near-complete record of game state, user keystrokes, and mouse clicks are available.

The assumption is that players in these competitive games are constantly trying to improve themselves in order to win more and earn a higher rating. A side effect of ratings based on player-versus-player matches is that ratings gained or lost in matches become zero-sum among the players, and are essentially relative rather than absolute metrics of player skill. This effect precludes performance modeling such as those proposed by Anderson and Schunn (2000) in ACT-R Learning Theory. Nevertheless, these matches are a form of competitive practice as players learn while playing. In essence, the match results serve as precise records of the historical progression the players undergo, with the game replays offering insight into what actions the players issued in each match. These match results and replays are naturalistic and available in large numbers, but there are also limitations to taking the approach of mining gameplay data. To supplement the quantitative data, we also provide quotes from players gathered through sample surveys and retrieved from online comments to help explain some of the behavior observed in the gameplay data.

Study 1: Practice and Progression in *Halo Reach*

The first study is of a game in the popular *Halo* franchise on the Xbox console, *Halo Reach*. It is a first-person shooter, where players battle with rifles, grenades, plasma weapons, and swords. The matches start with the player spawning with initial weapons somewhere on a map; additional weapons, health, and other power-ups are available elsewhere. There are both singleplayer and multiplayer components, where the multiplayer games are played on an online gaming service called Xbox Live, on a local network, or on a single Xbox with split-screen.

In Team Slayer, by far the most popular multiplayer playlist (a set of game types with similar rules), teams earn a point whenever a member of their team kills an enemy player. When killed, players are resurrected at a random location to fight again. The team with the most points at 15 minutes or the first team to reach 50 kills wins the match. Thus, each match typically takes 12–15 minutes, with about 5 minutes following the match to view post-match statistics, assign the next teams and map, and load the next match. In this paper, we focus on studying skill in Team Slayer because of the simplicity of the game, its popularity, and its consistency of play from match to match. While half the players only play 40 or fewer matches of Team Slayer, the vast majority of the matches are from the minority of players who play hundreds of matches (Figure 1).

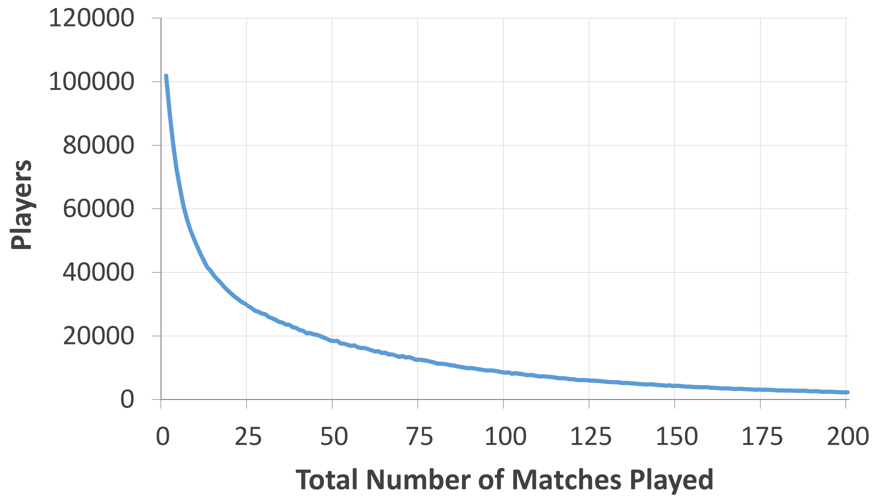


Figure 1. In Team Slayer, half the players played at least 40 matches, and a quarter played 95 matches or more. The chart extends beyond 200 matches as some people played over 1,000 matches during the 7 month period following the game's release.

TrueSkill

Halo Reach employs a skill rating system called TrueSkill (Herbrich, Minka, and Graepel (2006)), a generalization of the Elo chess rating (Elo (1978)). TrueSkill is currently used for matchmaking across numerous Xbox titles. The matchmaking system attempts to maximize the probability a match will end in a draw, which generally makes for an exciting match; of course, this is subject to practical constraints such as which players are currently looking for new matches. *Halo Reach* does not show players their current TrueSkill rating so there is little incentive for players to manipulate this rating.

TrueSkill represents a player's skill as a Gaussian distribution, parameterized with a mean μ and standard deviation σ; μ represents the best guess of that player's skill, and σ represents the variation in that guess. σ generally decreases over time as the player plays more matches since there is more information about their skill. μ starts with an initial value (a prior of 3), that adjusts to a player's "true" value for each multiplayer playlist. The matchmaking system attempts to pair up teams with equal skill (using a conservative estimate of skill computed by $\mu - C\sigma$, where C is a constant parameter), striving for balanced matches.

Player performance has been studied retroactively using TrueSkill for games of chess, showing that it can accurately predict the outcome of matches better than other rating systems (Dangauthier, Herbrich, Minka, and Graepel (2007)), and in *StarCraft*, where it agrees with public opinion about the top players in history (d__ijk__stra (2012)).

We use the TrueSkill "best guess" rating μ as the estimate of a user's skill. The ratings were retrieved from the official *Halo* servers that compute them for matchmaking. Our dataset consists of the complete first 7 months of matches from the 3.2 million *Halo Reach* players in its first week of release (September 13–20, 2012). We selected this cohort of players to control for the time when a person starts playing *Halo Reach*, and the remainder of this paper uses this cohort's historical game records from the 7 month period. Note that we are not sampling—this is the *complete population of players in this cohort*, and our dataset comprises every match played by that population. However, from this data we still know little about the mechanisms through which players improve their skills, which is examined in Study 2.

Our analysis can be reproduced by other researchers who download game histories from the *Halo Reach* API such as Mason and Clauset (2013), and the TrueSkill ratings can be recomputed using the published equations (Herbrich et al. (2006)). When plotting the players' skill in the charts, the median skill at every point along the x-axis was taken for each group. The median reduces the bias that occurs when plotting μ, a skewed variable that makes taking the mean exaggerate the effect of each factor.

Practice

Improving one's skill is tantamount to learning, and we wanted to look at specifically how *play intensity*, *breaks between matches*, and *initial skill progression* related to a player's skill in *Halo Reach* Team Slayer. These factors were chosen during discussion between the authors as potential determinants of skill, and relate to the phenomena of deliberate practice (Ericsson et al. (1993); Macnamara, Hambrick, and Oswald (2014)), the distribution of practice effect (Donovan and Radosevich (1999)), and the 'warm-up' decrement (Adams (1952)).

Play Intensity. We first investigated how skill is affected by a player's play intensity. Do players improve more if they play the same number of matches spread out over more weeks or played more compactly in fewer weeks? Do those who play more matches per week improve faster than those who play fewer, and is there a plateau of improvement? Prior research has shown that accumulated practice time predicts skill even when controlling for the current level of practice (Ericsson et al. (1993)). To explore these questions in our dataset, players are divided into cohorts of different play intensity measured by matches per week. Then each cohort is tracked in how their skill changes in each successive week of play, essentially weeks of practice.

From looking at these cohorts, two perspectives are needed. One perspective is at what rate of play intensity do players improve the quickest *per match*. Figure 2 presents in this information by plotting skill over matches for players grouped according to matches per week. The figure shows that those who play 4–8 matches per week seem to do best compared to other groups. However, from a different perspective of which players improve quickest *over time*, Figure 3 reveals that players who play more than 8 matches per week can surpass the less frequent players. Despite learning at a lower rate per match, the additional matches they played more than compensated for their slower skill gains. Essentially, cohorts with practice spaced out over longer periods of time progress in skill more efficiently. These results agree with studies that examine the effect of skill retention after practicing a task, which show that "individuals in spaced practice conditions outperformed those in massed practice conditions by almost one half of a standard deviation" (Donovan and Radosevich (1999)). Interestingly, those who play more frequently per week tend to start as less skilled players, but improve more rapidly, as shown by the 32–64 and >64 matches per week groups (i.e., players who logged over 8 hours a week of multiplayer Team Slayer).

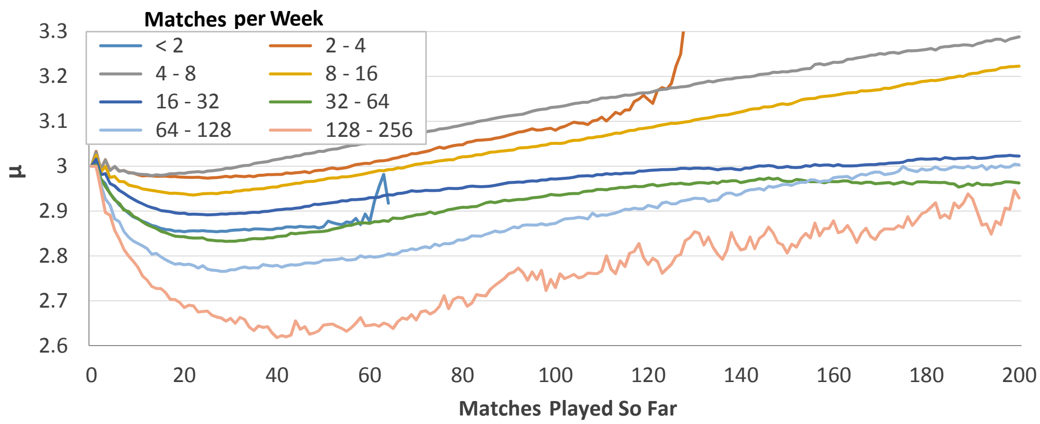


Figure 2. Players who play different numbers of matches per week gain expertise at different rates, generally trending towards higher skill for each match played. Note that each line is aggregated from thousands of players, and that it takes around 15 matches before μ more accurately reflects a player's skill.

Breaks in Play. Further investigating the idea of distributing practice, we can look at breaks from playing in the patterns of players' gameplay behavior. Players commonly took breaks of days, weeks, or months due to vacation, to play other games, real life distractions, or just temporary boredom with one game. In other performance tasks studied in the past, this has been referred to as a 'warm-up' decrement (Adams (1952)). The effect these breaks have on skill after a single match, 3 matches, 5 matches, or 10 matches can also be measured from the data. Here we look at skill changes between matches when the player returns to understand how much skill is lost during a break and how long it takes to recover practice time.

Figure 4 exhibits a few behaviors that players exhibit after breaks. The change in skill from before the break to after the break is illustrated by the 4 lines representing the next 1, 3, 5, and 10 matches after the break. When players are not taking breaks (breaks of 0 days), skill generally increases, evidenced by the climbing intercepts on the y-axis. Breaks of 1–2 days correlate with a small drop in skill in the next match played after the break, but has little long-term effect. In short, the loss of proficiency occurring due to short breaks is likewise small.

Longer breaks correlate with larger skill decreases, but the relationship does not appear linear (as a counterexample, 60 day breaks do not reduce skill twice as much as a 30 day break). More concretely, a 30 day break correlates with a skill drop of 10 matches of

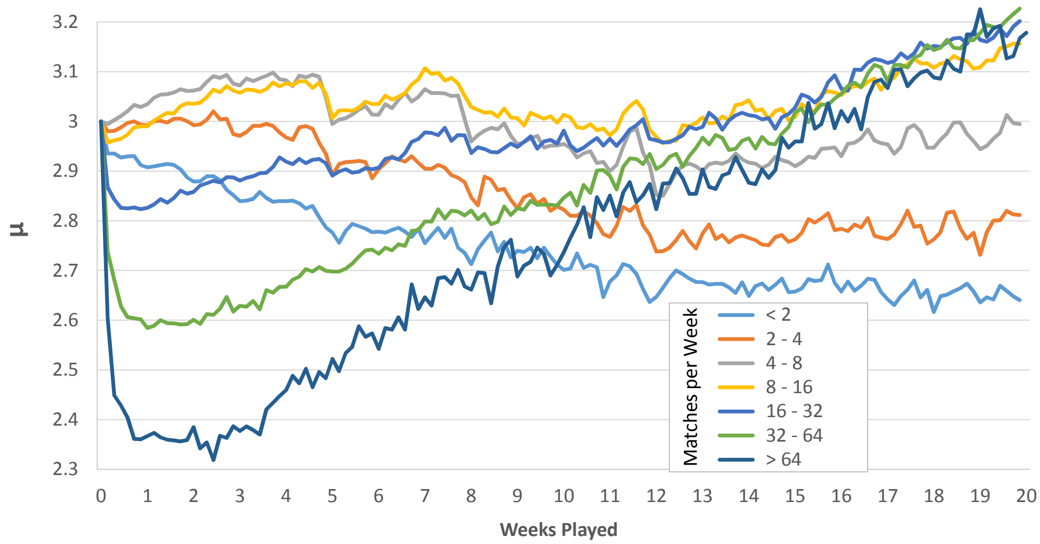


Figure 3. When looking at the skill over the number of weeks played, the more frequent players gain skill faster (evidenced by the higher slope of the skill), even though it takes more matches to reach that level. Here it's clear that it takes about a week for μ to accurately reflect a player's skill.

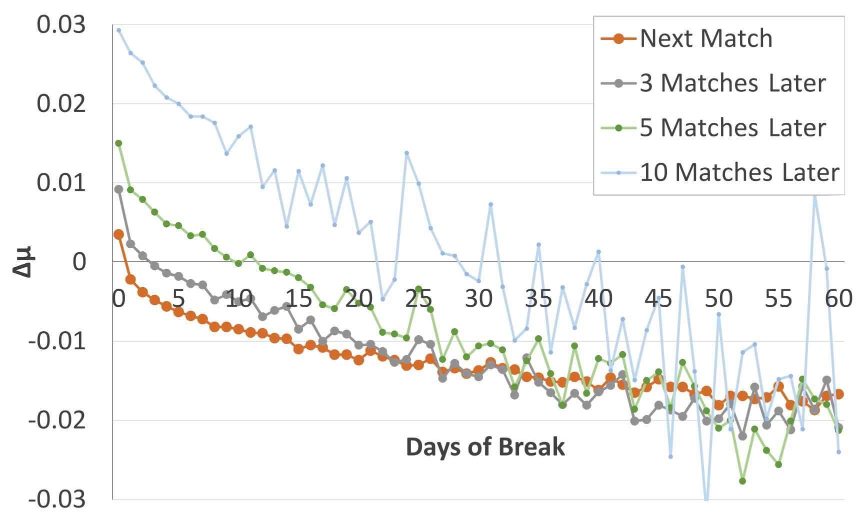


Figure 4. Skill change (plotted on the y-axis) between the match before and after a break. The x-axis represents instances of different lengths of breaks, with the change measured for the next match after returning from the break, 3 matches after, 5 matches after, and 10 matches after. Larger drops in skill typically follow longer breaks, but players can catch back up quickly.

play (10 matches later, the skill returns to the value before the break, i.e. $\Delta\mu = 0$); this is shown by the intersection of the '10 Matches Later' line with the x-axis. Thus, the amount of time required to regain skill following a 30 day break is only about 3 hours of gameplay.

These findings supplement those in the earlier section about play intensity. It appears that playing too frequently prevents the player from optimally earning skill per match, but taking too long of a break results in a loss of skill when the player returns. Thus, a player who is most efficient in gaining skill is one that plays occasionally without long breaks in between matches.

Compared to retraining in physical sports, this catch-up time is short; this may be because there is little physical catch-up required. The player only has to reacquaint themselves with the controls, and regain the mindset of their previous play.

"When I return after a prolonged absence my aim is less sharp and I play rubbish for a while which is obviously less fun. I sometimes get the added bonus of my creaky brain forgetting the buttons which is never fun either!" —P1

Progression

Skill Change Across All Players. While the median player's skill increases over time, this is not true for every player. We can classify different players' skill change over time and look into each group more closely. We converted the skill time-series into a symbolic representation of 4 levels and 4 time segments (4×4) using SAX (Symbolic Aggregate approXimation) (Lin, Keogh, Lonardi, and Chiu (2003)). SAX is a popular algorithm for discretizing time-series data. A player's skill over time is normalized, and divided into equal segments; each segment is then converted into a symbol depending on how much it deviates from the expected mean (Figure 5 shows an example). The segments used in our study were the four periods between the first match and the 100th match (to control for the same number of matches per player). Applying SAX to the skill over gameplay data allowed us to aggregate the different patterns of skill change from multiple players.

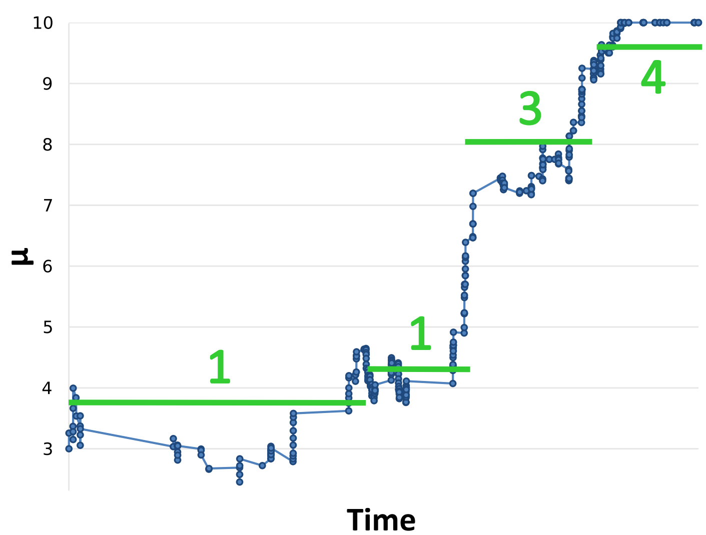


Figure 5. The skill plotted over time for an example individual player. The SAX representation is overlaid, creating the sequence "1134" indicating the player improved drastically in the second half of their matches, eventually becoming one of the best players.

Table 1 shows that the most common pattern in skill change was a slow steady increase in skill. The second most common pattern showed the opposite trend—a slow decline in skill

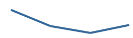
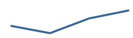
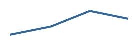
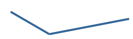
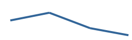
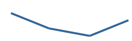
Pattern	# Players	Total Matches
	61791	217
	45814	252
	36320	257
	27290	219
	22759	216
	22452	253
	20659	260
	20633	222
	19858	247
	19292	216
	17573	219
	17454	245
	17389	260
	15670	215
	13692	236
	12516	239

Table 1
Skill change patterns (top 15 most common patterns) in the first 100 matches for players that played at least 100 matches. There are 4 segments in each pattern, of which there are 4 possible levels. The total matches is the mean number of matches played in the entire 7 month period that we have data, with higher than average total matches highlighted in green and below average total matches highlighted in red.

in the first 100 matches. Additionally, numerous other patterns were common, including sharp rises and drops in skills, and improvement followed by decline and vice versa. The most surprising finding is that players who improved in the first 100 matches actually ended up playing fewer total matches in the entire 7 month period than players with declining skill. We believe two factors play a role in this effect: 1) players who improve are more aggressive and hardcore gamers; 2) a skill improvement is not obvious to the player, but they do notice themselves performing worse against (unknown to them) stronger opponents, and the additional challenge may cause additional stress and frustration, provoking them to play fewer games.

Skill Change in Top Players. Next, we sought to examine how the progression of skill in the top players occurred, especially in relation to the average player. These were the 100 players with the highest TrueSkill rating at the end of the 7 month period in our dataset. 3 players were removed from this skill progression analysis because they did not complete at least 100 matches.

Again, we converted the time-series data of skill ratings of the top players into a SAX 4×4 discretization. Unsurprisingly, steady or fast climbs in skill levels in the first 100 matches were common patterns for the top players. Nearly all top players were able to achieve a high skill level by the end of their first 100 matches. Compared to the entire

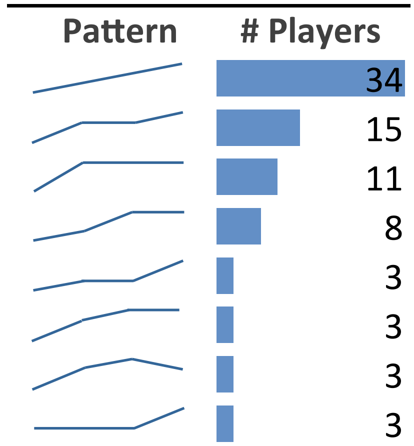


Table 2
Common skill change patterns in the first 100 matches for the top players. Again, there are 4 segments in each pattern, of which there are 4 possible levels. Less frequent patterns feature dips, but the dips do not occur for most top players.

population of players, top players exhibited more large leaps in skill levels (Table 2).

Top players improve more consistently. Dips in the first 100 matches only occurred in 14% of top players, compared to dips occurring more than half the time across all players. Since dips can be signals of technique or strategy shifts that eventually improve performance (Gray and Lindstedt (in press); Scarr, Cockburn, Gutwin, and Quinn (2011)), the lack of dips may suggest that top players already have the habits and innate tools to succeed from the start, and do not encounter substantial strategy shifts. This finding provides an orthogonal perspective to the theory of deliberate practice from Ericsson et al. (1993), where natural ability should be overwhelmed by practice that makes a focused effort to improve one's skill. Because our data is captured at the onset of the game's release, no player had a chance to practice beforehand. However, it is quite possible that the player's age or other related activities may have influenced the ability for one player to progress faster than another, as Campitelli and Gobet (2008) hypothesized in a longitudinal study of chess expertise.

Study 2: Forming Habits and Routines in *StarCraft 2*

The second study focuses on a unit grouping interface feature in a popular real-time strategy (RTS) game, *StarCraft 2*, where instead of controlling a single soldier, players command up to two hundred individual units. At the start of a game, players have a single building and a handful of units to command; the opposing player has the same. They then compete to gather resources, build infrastructure (e.g. barracks, factories, starports), and train armies to destroy their opponent. Like many other competitive video games, *StarCraft 2* presents players with an intricate and demanding task that favors rapid context switching and mastery of the game interface.

StarCraft 2 provides an ideal case for exploring how the behaviors of expert players have been optimized to efficiently multi-task and strategize. *StarCraft* and other real-time strategy games require skilled players to control and manage hundreds of units at once, from soldiers in battle to resource harvesting units to production buildings and builders. While novice players struggle to keep up with the increasing number of demands on their attention, better players use unit groupings to bind groups of units to single keys, and thus can issue commands to numerous and different sets of units quickly. In the span of seconds, an expert player can use his pre-set unit groupings to command his army (e.g. group 1) to assault the enemy base, but also to send a specialized force (e.g. group 2) to cut off enemy units who are en route. All the while, he is checking the production of new units at his home base, assigned to group 5. Because unit groups expand the player's capacity to multi-task, we can better understand the different player skill levels by studying the characteristics of unit group use. Replays of *StarCraft 2* matches were retrieved from two sources and included player information that allowed us to discern the skill level of the source players. The game servers place players into different leagues based on skill: Bronze, Silver, Gold, Platinum, Diamond, Master, and Grandmaster—which we examine for differences in unit group usage.

Unit Groups

Unit groups are commonly referred to within the real-time strategy gaming community as 'control groups' or more colloquially as 'hotkeys.' They are used by players to efficiently control and manage diverse groups of units within the game. Unit groups are generally referred to and accessed via keys {0–9} on the keyboard and store selections of units within the game. This ability is important as during a game, players can only issue commands to a *working set*: a single buffer containing references to units currently controlled by a player. In order to control a unit not in their working set, players must update their working set to include the desired unit before issuing commands.

Unit groups offer the convenience of allowing allow players to rapidly switch their working set to previously defined selections of units. Players can modify unit groups by adding (or "binding") additional selected units to a unit group number or by replacing its selection with the current working set of units. Players can also recall the units assigned to a specific unit group, which will update their working set with the units assigned to the selected unit group. Use of unit groups is not required to play *StarCraft 2*, as players can manually select units each time using the mouse, but allows for faster context switching and command execution within the game. An example of a unit group mapping is shown in Figure 6.

For non-Grandmaster league players, we downloaded replays from a popular replay aggregator website, GGTracker. Players upload replays to aggregator websites such as GGTracker to share them with others, or to access the analytics of replay data (e.g. actions per minute, resources collected, etc.) that such websites provide. Replays were included in our study if they comprised a 1 versus 1 match of at least 5 minutes in the North America region, where the players' usernames were not obscured and their skill league during data collection matched their skill league during the match. Replays representing the Grandmaster class were obtained from season 2 of the 2013 World Championship Series (WCS) tournament replay pack released by Blizzard (2013).

Distributions of aggregate unit group use are shown in Figure 7, illustrating that increasing skill correlates with steadily more frequent unit group usage. Here, we measure

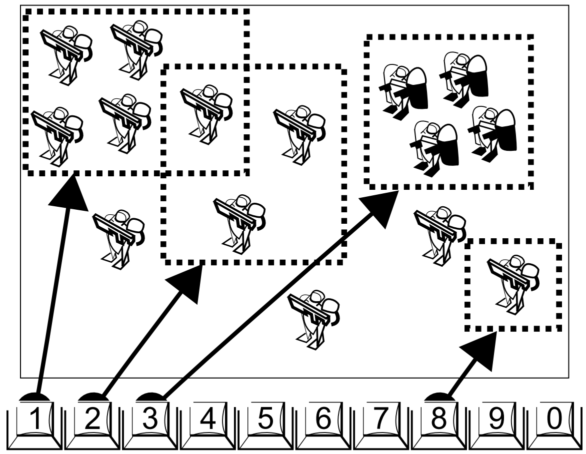


Figure 6. Unit groups are bound to keyboard keys 0–9 and represent groups of units. Different groups can include the same units, or no units at all.

how often a player uses unit groups in terms of commands issued by the player per second, a popular type of metric for real time strategy games. In both the less skilled leagues, there remains a substantial proportion of players who essentially do not use unit groups. Within the higher skill levels, all players use unit groups to some extent, with the majority executing around two unit group commands per second.

Unit Group Features. Players often choose to assign units of different types in habitual yet distinct ways. For example, one player may always choose to bind production structures to unit group 5 whereas another may always choose unit group 3. Considering these tendencies, it seems appropriate to consider frequencies for each of the unit groups separately, as unit group usage may depend on how often the units bound to a unit group need to be selected. In this work, we focus on the rates at which players execute unit group actions as the features for our analyses.

We also distinguish between the types of commands that can be issued to unit groups so that a first collection of command rates is obtained for setting a unit group to the current selection, a second for adding the current selection to a unit group, and a third for recalling the selection specified by a unit group. This differentiation into three types of commands is potentially useful again because of the freedom players have in executing unit group commands—how often players repeatedly rebind or update a unit group is user dependent.

Together, these combinations yield 30 features per player per game, as we consider three types of unit group commands: set, get, rebind, with 10 possible key bindings $\{0–9\}$ per command. Each feature was therefore the frequency that a specific unit group and action combination was used in the game. At a high level, each unique player of the game is represented as a vector of features $[f_1, f_2, f_3, ..., f_{30}]$ corresponding to the frequency of each unit group action.

Forming Habits

Warmup. *"Something that really made me play better was spamming, getting your hands warm and fast will make it possible in the later stages of the game for you to multitask and just play alot faster. Also try tapping between armies, scouting units, bases even if*

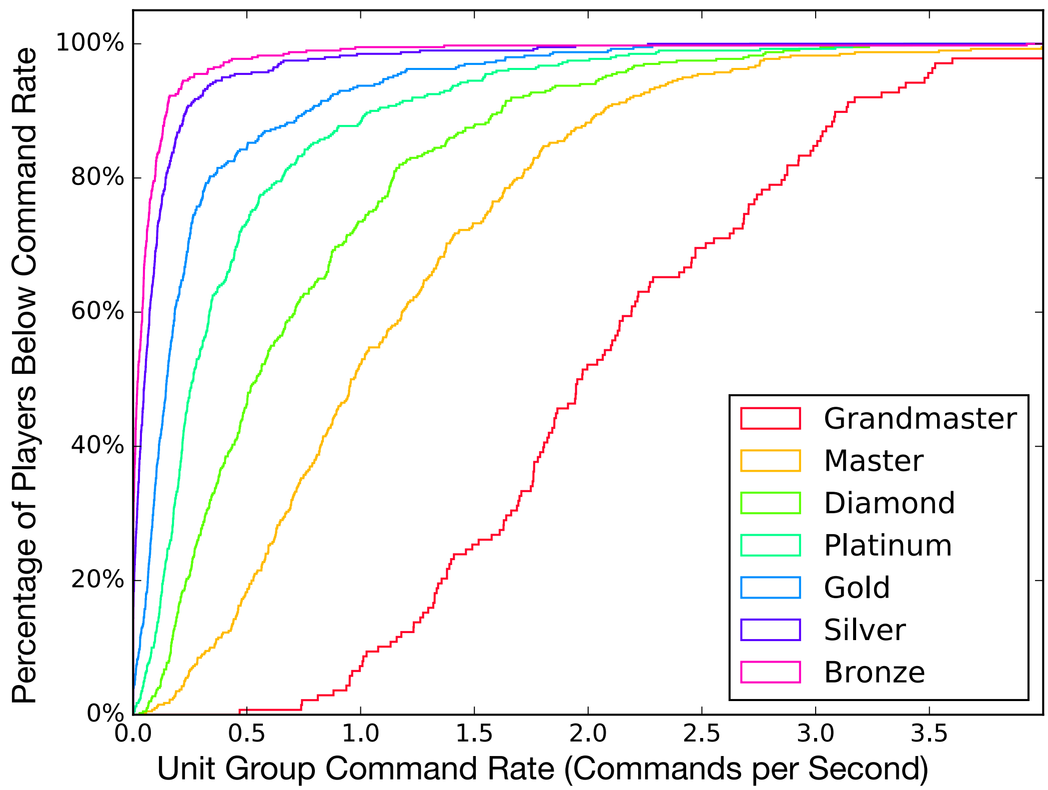


Figure 7. Distribution of unit group usage among the skill leagues, measured as command rates (average number of commands issued per second). A significant proportion of less skilled players use unit groups infrequently or not at all. Increasing rates of unit group usage are seen with increasing skill levels up to the Grandmaster league, where the average is around two commands per second.

nothing is really going on. The worst thing you can do is just to sit and watch ur base with 0 [actions] when nothing is needed to be done.” —P2

During professional *StarCraft* matches, players can be seen tapping their keyboards at very high speeds. These actions are registered at rates of 200 actions per minute and up. Understandably, this allows professionals to issue a high amount of commands in order to control hundreds of units at a time. Yet, even in the first minutes of a game, when players have only one building and a small number of worker units, they are already selecting and re-selecting unit groups at these same rates of hundreds of actions per minute. Players at all league levels can be found doing the same kind of re-selections during the introductory seconds of a match.

Why do players do this? There is no effect on the behavior of the units; in the first seconds, units are moving automatically to collect resources without any input needed from the player. Nor is there is an effect in re-selecting the same group hundreds of times over. So, is it pointless activity, “spam” as it is called by the community[1]? Are they enacted merely to show off fast hand movement speed and to inflate the reported APM at the end

[1] http://wiki.teamliquid.net/starcraft2/APM

of a game? Are less-skilled players just emulating professional players mindlessly? Or, as P2 suggests, it is a "warmup" action that enhances the players hand speed and mentally prepares them to win?

This warm-up-effectiveness question exists in physical sports, too, where similar warmup activities have undergone empirical study. Like competitive *StarCraft* gamers, athletes want an advantage through warmup activities. They seek direct performative benefits. Zois, Bishop, Ball, and Aughey (2011) report on the increased performance for soccer players when they replace their usual warm-up routine with high-intensity leg presses and game-like activities (passing, shooting and ball-control). Transferring these concepts to *StarCraft* suggests that there may be direct physical benefits to mashing unit selections ("APM spamming"), which itself it a high-intensity activity.

Another body of research in the field of sports has coined the term "Pre-performance routines" (PPR), defined as "a sequence of task-relevant thoughts and actions which an athlete engages in systematically prior to his or her performance of a specific sports skill." Cotterill (2010)'s summary of experimental studies show a positive impact of routines in basketball, golf, bowling, tennis, water polo, rugby, gymnastics, darts, and volleyball. Cotterill lays out a broad set of potential benefits of PPR that extend beyond physiological advantages to include mental preparation, emotional control (e.g. avoiding "choking" under pressure (Mesagno, Marchant, Morris, et al. (2008))), tuning the reflexes, and more. Additionally, Cotterill briefly highlights a connection between PPR and individualization, arguing that routines are more effective when they are tailored to the needs of the individual performer. We also see that the unit group patterns of individuals are distinct—see our later section on Individual Habits.

This literature suggests that *StarCraft* players would derive performative benefits from "spamming"—although its benefits at lower leagues may be questionable, not because the warmup is pointless, but rather, because novices to the game also lack the knowledge of how best to warm up. Looking to our dataset, we define the warmup period as the first 120 in-game seconds. During this time period, players have only a few units to control. Still, players can choose to bind these units to unit groups and rapidly cycle through them to warmup. We compared their warmup to their non-warmup (120+ seconds until the end of the game) unit group usage.

In the lower leagues, a few players bind units to groups at the start of the match and then stop using them entirely. Other players, and in greater number, exhibit similar behavior: they use unit groups more than three times as frequently during the warmup phase than they do outside of it. This result could be attributed to less skilled players who attempt to mimic the behavior of expert players at the beginning of the match by spamming excessively yet lack the ability to sustain the unit group use rates throughout the match as their attention is taxed. These players may be attempting to integrate unit groups into their play (see warmup trends in Table 3) but have not yet sufficiently mastered them. Expert players show almost identical unit group use rates during warmup and non-warmup phases of the game. In summary, lower-league players exhibit low effectiveness in translating their warm-up actions into the actual match.

"I constantly spam [unit groups] 5 and 6 checking my queens energy and only stop when I'm moving guys or building units." —P3

League	Median Warmup Command Rate	Median Non-Warmup Command Rate
Bronze	0.012	0.020
Silver	0.023	0.052
Gold	0.125	0.143
Platinum	0.307	0.229
Diamond	0.782	0.482
Master	1.377	0.901
Grandmaster	**2.360**	**1.907**

Table 3

Median warmup and non-warmup command rates (in units of commands/second). As skill ratings increase, warmup and non-warmup command rates increase. Players at higher skill levels seem better at sustaining unit group usage throughout the match.

Under Pressure. To gain a better understanding of unit group usage and its relationship to player skill, we focused on finer-grained categories that represent two distinct unit group usage: macro and micro. Macro actions maintain the player's economy to keep income and production optimal: continually producing new units and commanding new workers or soldiers. Micro actions optimize the effectiveness of individual units as they scout, position, harass, and fight. We investigate these two categories in two different forms of time-pressure in the game: battle and peacetime. Because it is easy to neglect unit production and resource harvesting during battle, the ability to maintain economic efficiency (macro) while using units appropriately during battle (micro) is a trait of a skilled player.

"Get a macro rotation.... Every time you warp in, check money, check supply... Every time you start a colossus [a unit that requires 8 food supply], ... build a pylon [a building that provides 8 supply]" —P4

Battles require a lot of focus from players as they try to manage dozens of fighting units. More skilled players are still able to multi-task during these battles and continue to execute macro commands. In replays, Grandmaster players show the most frequent use of unit groups to select production buildings both in and out of battle, with lower usage rates in the lower leagues (Table 4). Interestingly, the median event rate for Grandmaster players in battle is quite close to that of Master league players during peacetime. Performing excessive or spamming selections of production buildings can be helpful to monitor production queues as it can ensure that the idle time of buildings is minimized. In the quote above, P4 habitually checks his different units and buildings in rotation. This player also has a tendency to pair the training of an expensive unit with the construction of the food supply that it consumes. This allows him relegate some of his macro work to pure reflex.

StarCraft 2 is a fast-paced game, and during battles the number of units bound to a player's unit groups can diminish rapidly as units are eliminated from the game. Unless they are given explicit orders, newly produced units do not automatically join these groups. To maintain these groups, skilled players rebind their unit groups by either setting them to new selections of units or adding additional units to them. In the replay data, these actions can be connected to player skill.

We find that players rebind units most in the Grandmaster league both in and out of battle, and that players in the Grandmaster and Master leagues have the most similar rebind rates (compared to themselves) in and out of battle (Table 5). This suggests that

League	Median Peace Selection Rate	Median Battle Selection Rate
Bronze	0.011	0.002
Silver	0.036	0.020
Gold	0.114	0.058
Platinum	0.216	0.098
Diamond	0.422	0.161
Master	0.752	0.317
Grandmaster	**1.332**	**0.712**

Table 4

Median macro selection rates during peacetime and battle (in units of commands/second). As skill ratings increase, production structure selection rates increase. Many Grandmaster players select production structures via unit groups in battle as often as players in lower leagues do during peacetime.

higher skilled players are more vigilant about managing newly produced units, and that lower skilled players are distracted by battles so they perform fewer rebinds than during peacetime. In essence, rebinding is a type of unit management and organization task which requires more cognitive load to perform, especially at a level of skill where this is not yet a habit.

League	Median Peace Rebind Rate	Median Battle Rebind Rate
Bronze	0.006	0.000
Silver	0.010	0.000
Gold	0.019	0.004
Platinum	0.027	0.009
Diamond	0.037	0.022
Master	0.054	0.047
Grandmaster	**0.088**	**0.097**

Table 5

Median unit group rebind rates during battle and peacetime (in units of commands/second). Rebind rates are highest in the Grandmaster league.

Individual Habits

Routine Transitions. Professional *StarCraft* players warming up can often be observed rapidly cycling through unit groups without necessarily issuing any commands to their selected units or structures. In terms of actual keystrokes, this warmup can resemble repeated sequences such as '123123123123' or '456456456.' We can estimate the transition probabilities between two unit groups {A, B} by counting the number of times a player selects unit group B following unit group A and dividing by the total number of times a player selects any unit group after selecting unit group A.

The transition probabilities themselves can reveal interesting differences among players. For example, Figure 8 shows a transition matrix of a player who has a tendency to make repeated selections of certain unit groups, namely unit group 1. This pattern of transitions indicates that this player puts together the units that they frequently check into unit group 1.

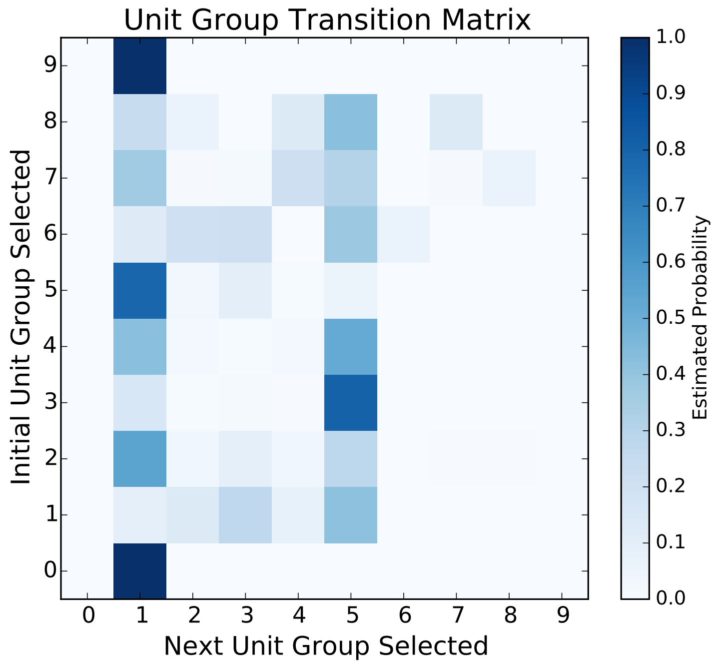


Figure 8. The transition matrix for a single player who makes few repeated selections of unit groups and appears to use unit group 1 most frequently.

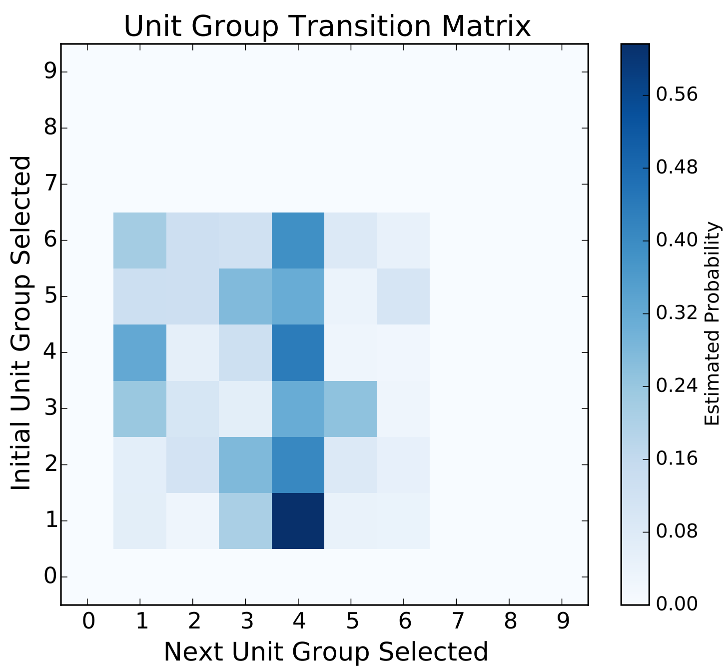


Figure 9. The transition matrix for a single player who makes frequent repeated selections of unit group 4, with transitions between groups 1 and 4 and vice-versa being common.

Some players select a single unit group repeatedly, such as the player shown in Figure 9. Unlike the player in Figure 8 who has relatively few repeated selections, the player in Figure 9 selects unit group 4 even when that group is already selected. Repeated selections of the same unit group—as shown on the diagonal of the matrix—represent a higher proportion of some players' transitions than others. These repeated selections suggest a play style that incorporates more "spam" actions. Note that selecting a new unit group does not refocus or change the player's view. Thus, these transition probabilities act as individualized fingerprints that also provide clues to the specific play style of each player.

Habits Change With Skill. Naturally, more skilled players are dissimilar to less skilled players simply because better players tend to use unit groups more frequently. However, this leads to a follow-up question: since better players use unit groups more frequently, do they use them in similar ways? Or, are their styles of play truly unique?

League	Mean	Median	SD	Min	Max
Bronze	0.076	0.036	0.188	0.000	2.318
Silver	0.118	0.057	0.204	0.000	1.710
Gold	0.261	0.139	0.283	0.000	1.818
Platinum	0.396	0.254	0.379	0.002	2.853
Diamond	0.581	0.500	0.390	0.006	2.861
Master	0.754	0.676	**0.430**	0.039	**3.801**
Grandmaster	**0.955**	**0.914**	0.378	**0.096**	2.452

Table 6

Player to player distance statistics at different skill levels (in units of commands/second). The higher the skill level of players, the greater the distance becomes between any two players. This trend suggests that as skill level increases, players tend to diverge in terms of their unit group usage habits.

Comparing player to player differences at varying skill levels answers this question. Here, each player is represented by a vector where each value is the frequency that they *set*, *get*, *rebind* each unit group for each of the 10 possible unit groups, resulting in a total of 30 values in each vector (as described previously in Unit Group Features). So the distance between two players P1, P2 is $\sqrt{(f_{1P1} - f_{1P2})^2 + (f_{2P1} - f_{2P2})^2 + ...}$ for the vector for each player. We consider the Euclidean distance between players as a perspective on how similar two players are in their unit group habits. (Table 6). From the perspective of these features, Grandmaster players have the *most* distinct unit group habits. This trend also appears in lower skill levels: the average distance between two Gold level players is also less than the average distance between two Diamond players, and so on.

Uniqueness. *"You just have to worry about doing the same thing every time, regardless of the situation, so it becomes muscle memory and a reaction. ... whatever you're doing needs to be consistent every time so it can be written in your memory and you yourself will become consistent." —P5*

Table 6 shows that players in the Grandmaster league tend to develop unique patterns of unit group use. Additionally, using the same definition of "distance" as in the previous section, the intra-individual distance (distance between two matches from the same player in two different matches) was on average significantly lower (mean = 0.359, SD = 0.272 commands/second) than the inter-individual distance (distance between two players).

Therefore, expert players not only tend to develop unique patterns of unit group use, but also they remain reasonably consistent from match to match. In other words, expert players have signatures of unit group behaviors that can be used to identify them.

The fact that uniqueness in unit group habits should exist among players is not obvious, as many players will often employ similar strategies in game that require control of the same unit types. Some strategies are so typical among players that they are referred to the community as "standard play." Along the same lines, the fact that consistency should exist within a given player's games is also not obvious, as players can draw from a wide range of strategies when playing—especially when playing multiple matches against an opponent in a tournament setting. These strategies will often rely on using different unit types that need to be controlled differently, yet players adapt their unit group habits around these different strategies.

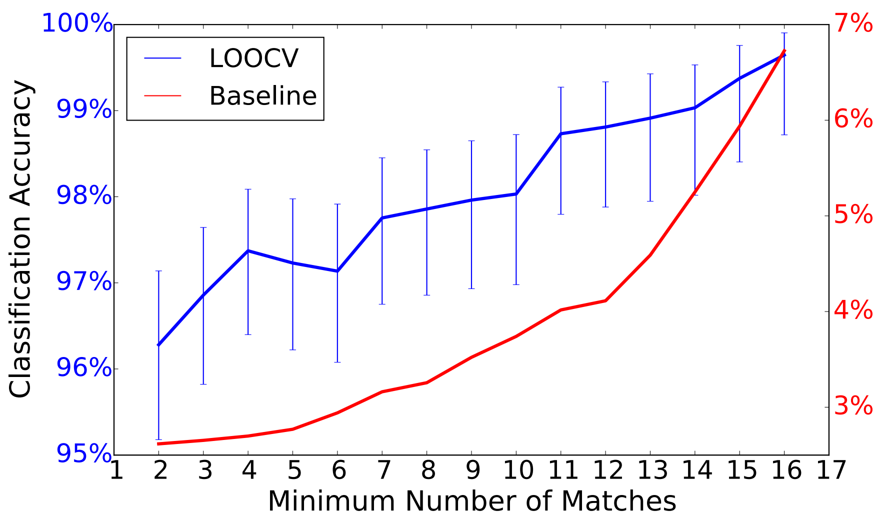


Figure 10. Accuracy of identifying of expert players from unit group habits estimated via LOOCV compared to baseline identification performance. Here, error bars represent the 95% confidence interval for the LOOCV accuracy. The minimum number of matches is an additional inclusion criteria for players.

Since our data includes expert players with varying number matches played, we measure their uniqueness by attempting to identify them based only on their unit group usage. With a minimum of 2 matches required per player, we were able to achieve a leave-one-out cross-validation (LOOCV) classification accuracy of 96.3% (95% CI 95.2%–97.1%) compared to a baseline accuracy of 2.6% when choosing the most frequent class[2]. In general, classification performance improved as the number of matches required for a player to be included was increased (Figure 10). As we increased the minimum number of matches to 16 matches required per player, we were able to achieve a LOOCV classification accuracy of 99.6% (95% CI 98.7%–99.9%) compared to a baseline accuracy of 6.7%. The uniqueness of unit grouping habits leads to high-confidence identification of a player after only a few matches.

[2]Due to the relatively high value of $\hat{p}$ in these cases, we compute the 95% CI using the Wilson interval recommended and defined according to Brown, Cai, and DasGupta (2001)

Experts are particularly concerned with hiding their identity when sparring on public ranked matches as it prevents opponents from gaining advantages by studying replays and understanding one's unique tendencies, strengths, and weaknesses. At the time of writing, more than 70 out of the top 100 ranked players in the world were using an obfuscated username[3] to hide their identities. Our performance in Figure 10 shows that accurate and rapid identification of experts is possible using our features. While online tournaments for competitive video games increase in popularity as qualifying stages for larger events, the ability to identify players or detect mismatches in identity will become increasingly valuable.

"Professional gamers are known to study the replays of an opponent before an important match, much like a chess grandmaster preparing for a match." —Weber and Mateas (2009)

Discussion

Practice and Progression

Our higher-level analysis of skill through practice and progression illustrate several results. Players gain skill at a faster rate (skill per match) when they play a moderate number of matches. However, those who play more matches will gain skill quicker if compared to less intense players, due to the sheer number of matches they play even if their rate of skill gain is lower. We also notice that sustained practice is necessary but not sufficient; taking breaks of several days quantifiably impact a player's skill when they return, which requires recovery time of several matches to return to the skill they had before the break. The longer the break, the longer the recovery period. Thus, for practice with the goal of increasing skill, a player may consider long periods of consistent and sustained play. This compares to Donovan and Radosevich (1999) which emphasizes spaced practice, but here our data also suggests that the time interval should not be too wide.

After analyses of the different factors and looking individually at the top players in *Halo Reach*, what have we learned about practice and progression? Certainly, the concept of expertise varies for every task, as telegraphy or sports training are very different activities from blasting enemies on a screen. Some factors we examined span multiple activities: frequency of performing the activity correlate with higher performance up to a point and a catch-up period follows long breaks. But among the top players, skill is acquired differently: some players gain skill rapidly the moment they start playing, while others lose some skill to gain it back again later. Also intriguing are those who reach a plateau (as evidenced by the SAX patterns in their first 100 matches) only become the best after time away from the game. Such plateaus that are transcended may be evidence of overcoming a suboptimal strategy (see Gray and Lindstedt (in press)), which may be further examined in the future by corroborating long-term skill patterns (like we did for *Halo Reach* with in-game analytics (as with *StarCraft 2*.

Our work corroborates the study of skill trajectory in an online game by Stafford and Dewar (2014) in a couple of ways. First, Stafford and Dewar divide players into *goers* who play more frequently within a fixed time-frame and *resters* who take longer breaks between matches. Like our findings, players who play less intensely earn more skill per match. Therefore, it appears there is some value gained in the period between matches if

[3] Obfuscated usernames are created by mixing together a sequence of the characters 'l', '1', and 'I'

the breaks are brief. However, we also note that our findings about longer multi-day breaks show that when players are away from a game for many days, they will need to recover skill when they return. Second, Stafford and Dewar suggest that the top players have some sort of advantage from the very beginning due to their higher score compared to other players, and their ability to increase this lead over time. We note in our skill progression analysis that the top players have a trajectory comprising few drops in skill compared to the overall player pool. These comparable findings hold even when the type of game differs (a competitive player-versus-player game compared to a singleplayer game), and the skill metric differs (TrueSkill requires one player to lose rating when another gains, while the score from Stafford and Dewar (2014) has no relation to the performance of other players).

Expert Habits, Unique and Consistent

Experts retain a consistent set of habits and routines that allow them to perform at exceptional levels. Furthermore, they exhibit uniqueness and idiosyncrasy in these routines. (*StarCraft* unit group usage is unique enough to fingerprint individuals).

The consistency that exists among the usage patterns of expert players is interesting because these players are not executing the same strategy (e.g. the order in which they prioritize building units) multiple matches in a row in a tournament setting. They are constantly forced to adapt their play to their opponent's in-game race, style, and tendencies, yet the same routines remain. This behavior suggests that experts adapt their unit groups to their current build ordering and composition, perhaps as a way of coping with the need to play quickly. If unit group use was defined by game events and outcomes, (e.g. only checking on an army when being attacked, selecting production only when idle), we would see little consistency among players. Our results lead to the hypothesis that the relationship between habit and performance is cyclic: experts are capable of sustaining consistently high performance because of their unit group habits, and their unit group habits exist as a result of this consistent performance. Overall, our work speaks to the power of habit and importance of adaptation in the face of diverse and time-pressured situations.

The primacy of these idiosyncratic habits carry implications for how the broader picture of practice and progression might be explained. For example, one might ask how a habit may or may not deteriorate between gaming breaks. Or, if seen as a strategy, a particular habit might be a high performing habit or suboptimal habit with a ceiling of effectiveness. This view of habit offers an explanation of SAX patterns such as the plateaus in skill since our analyses show that players ingrain habits into themselves, producing a well-worn consistency that intersects all of the matches that they play. Perhaps the consistency in a habit explains both why it is possible to achieve skill improvements with high use and why heavy practice that is not reflectively "deliberate" (Ericsson et al. (1993)) can result in plateaus.

Warmed Up and Ready to Perform (or Practice)

A common thread through both of these studies is warmth, to be warmed up after a long break and to warm up the fingers for quick play. Furthermore, we've identified a type of warming-up that involves meaningless actions that prepares experts for performing when it counts. We believe that is particularly applicable in other skilled domains.

For example, researchers in the field of construction (Chen, Golparvar-Fard, and Kleiner (2014); Zhao, Thabet, McCoy, and Kleiner (2012)) have designed virtual environments to train construction workers in managing electric hazards. A virtual simulation of electrical tasks in hazardous situations allows trainees to "recognize hazards, strengthen proper working memory and transfer the relative experience into real life work" —Zhao et al. (2012). In medical practice, 3D replicas of patient's hearts, skulls, livers and other organs allow surgeons to rehearse a tricky surgery (Rengier et al. (2010)). These practitioners are exercising meaningless, safe actions to ingrain a skillset within themselves. In an emergency, that skillset will be readily available.

Our findings have a number of implications for these kinds of training. First, the success of this training is intertwined with its frequency and intensity. Of course, the more frequent and intense the training sessions, the more effective it will be. But, there may be an upper bound. Following our results, we can recommend less intense training sessions when they are sufficiently frequent. Thus, a frequent safety training program can be kept effective despite reducing the duration of individual sessions. Second, infrequent sessions need time for practitioners to recover their forgotten skills. There might be a number of ways to do this. One would be to draw on our findings of meaningless warm-up actions and to recommend a series of warm-up activities before resuming a session. Third, when there actually is a crisis or surgery, we should also ask if respondents or surgeons need to warm-up before jumping in to action. For example, a common instruction for performing CPR (cardiopulmonary resuscitation) is to follow the rhythm of "Staying Alive" by the Bee Gees. To warm-up would be to sing a few bars of the song before beginning the resuscitative action. In the domain of surgery, a doctor would not repeat an entire rehearsal surgery on another 3D print. But, we would ask if there are particular hand motions in the surgical act that can be done to "warm up." We predict that those particular motions are idiosyncratic, just as they are for *StarCraft* players.

Overall, our findings emphasize the importance of attending to the frequency and intensity of training or rehearsal sessions; and, we recommend taking a closer, micro-focus on the habits of the hand—to see how they are engrained, warmed-up, and available for the 'real' performance.

Returning to Ericsson et al. (1993), we might even consider "deliberate practice" itself as a skilled activity that requires its own "warming up." Ericsson et al. paint a picture of successful practice as as influenced by larger environmental and contextual factors of motivation, years of engagement, and more. Our different models of skill progression for play intensity, breaks in play, and the theme of warming up all together suggest that there is more to discover on what an optimal "deliberate practice" session consists of. Perhaps, there is a time of physical warmup required to bring the body of the player to a position of readiness. Or perhaps, a warmup exercise is required to overcome an extended break in practice. Then, one might ask, what is the proper level of intensity given the time allotted for practice or deadline for achieving mastery. We look forward to seeing the picture of deliberate practice further understood, down to the moments before a session begins.

Limitations to this Approach

There are several limitations to studying large-scale naturalistic gameplay data. However, many of these limitations were reduced due to our use of cohort analysis. For example, if players in a cohort that started later stick with the game for longer, this is a signal that the game has become more compelling to play for longer. For each of the studies presented in this paper, players are grouped by either start date or by skill level. Players grouped by start date allow us to observe their progression as they begin their initial matches. Rather than examining data within a particular calendar date range, a start date based cohort would naturally make the first day of play for each player comparable. This approach also avoids many confounding factors due to changes in the game itself, changes in the game's culture or overall player base, or even world events like holidays or popular sporting events that can change the gameplay demographic. Players grouped by skill level allow us to compare between these cohorts, to identify if there are any behaviors such as gameplay habits that allowed the players to acquire higher skill levels.

One common limitation of post-hoc data analysis is the inability to understand causality since the game variables cannot be manipulated to create a controlled experiment. When two variables such as gameplay intensity and skill are correlated, it cannot be determined whether increased gameplay intensity caused an increase in skill or whether some external factor caused both factors to increase; for example, this external factor may be that those players who play more frequently are naturally capable of gaining expertise quicker. There are numerous demographic factors that may also confound the relationship, such as age which affects gaming reflexes and leisure time available to practice (Campitelli and Gobet (2008)). The third possibility, that an increase in skill causes more gameplay intensity is also possible, but is unlikely in our study since the player's skill rating is not shown to the player. Regardless, we caution claiming that particular factors will cause an increase in skill, but rather our findings describe the nature of players who have higher skill.

Another instance of this limitation is in the case of warmup effectiveness—while players that executed more warmup actions were generally favored in our data, it is not clear whether warmup actions are merely an indicator of greater skill. Executing more warmup actions could be an indicator that a player is comfortable playing quickly throughout all phases of the game rather than a form of practice that improves skill. How a study of warmup effectiveness in games could be potentially designed is an interesting question itself, as it becomes difficult to separate the effectiveness of warmup from enforcing an unnatural play style on participants. That is, even if warming up leads to better performance—does this benefit overcome the potential drawbacks if novice players find it unnatural to do so?

Unlike non-competitive games where players earn a specific score based on their performance (e.g., Stafford and Dewar (2014)), Elo-like rating systems start with initial priors which take some time to converge to a new skill level. The lag may result in an inaccurate rating in the beginning (due to the priors) or if a player's skill changes substantially. In *Halo Reach*, we observed the median TrueSkill decreasing from the starting value $\mu = 3$ initially—probably because a player's actual skill is lower than the starting value. It was not until about 30–35 matches later that the median TrueSkill μ rose.

User-reported data from 300 players when they signed up for an opt-in player experience panel showed that 18 of them (6%) reported sharing their Xbox live account with other people. When those sharing an account play the same game, and particularly the same playlist in *Halo Reach*, their different skills will confuse the rating system. The better player may raise the skill rating when they are playing, while the worse player will tend to lower the skill rating, causing it to be highly variable. Thus, during matchmaking, the skill rating may not accurately reflect the skill of the current player. Additionally, online gaming accounts can be handed off to another person, resulting in a similar inaccurate reflection of skill; it would take a number of matches for the rating to recalibrate to the new player's skill.

Conclusion

We identify how habits and practice of competitive video game players affect their skill in games. Thousands of game replays from *StarCraft 2* and match histories from *Halo Reach* are used to understand gameplay behaviors at a macro- and micro-scale. On a micro level, unit groupings in *StarCraft 2* called unit groups are a key differentiator of individual players as well as players of different skill levels; novice players rarely use unit groups while experts nearly always do. While certain unit group behaviors are common across all skill levels, expert players appear to be better at remaining composed and sustaining unit group use in battle. But even among experts, routines and habits are unique: both the frequency in which they use unit groups, and the order they cycle between groups. Broadly, play intensity, breaks in play, and skill change over time affect a player's skill in *Halo Reach*. Players with the most efficient skill gain are likely to play with moderate frequency, and avoid long breaks between matches. They are only surpassed by more frequent players due to the sheer number of additional matches that those players play. The best players in the 7-month period have varied skill patterns that often run counter to the trends seen for typical players; they have an innate advantage where they start at a higher skill level and also increase skill with greater velocity and encounter fewer dips in skill.

Gaming skill forms from both deeply engrained individualized habits due to time pressure, and sustained and intense practice that can result in bursts of improvement. For the studious gamer, they may seek to practice patterns of routines, which build muscle memory for time-sensitive situations. For the casual gamer, they may be satisfied in knowing that they are likely to be gaining skill at a faster rate per match than someone who plays more intensely.

Our work presents evidence that supplements existing studies in cognitive science and human-computer interaction. Gameplay data provides us an opportunity to find patterns of players' using cohort analysis. By grouping together players by start date or skill level, we are able to determine differences between groups of players to extract signals of behavior out of noisy naturalistic data. The signals from our two studies say "practice consistently, stay warm."

References

Adams, J. A. (1952). Warm-up decrement in performance on the pursuit-rotor. *The American Journal of Psychology*, *65*(3), 404–414.

Anderson, J. R., & Schunn, C. (2000). Implications of the ACT-R learning theory: No magic bullets. *Advances in Instructional Psychology: Educational Design and Cognitive Science*, *5*, 1–33.

Blizzard. (2013). *WCS Season 1 and 2 Replays.* `http://media.blizzard.com/sc2/replays/WCS-Season1-and-2-Replays.zip`. (Accessed August 30, 2014)

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, *16*(2), pp. 101-117. Retrieved from `http://www.jstor.org/stable/2676784`

Bryan, W. L., & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. *Psychological Review*, *4*(1), 27-53.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, *6*(4), 345-375.

Campitelli, G., & Gobet, F. (2008). The role of practice in chess: A longitudinal study. *Learning and Individual Differences*, *18*(4), 446–458.

Chen, A., Golparvar-Fard, M., & Kleiner, B. (2014). Saves: An augmented virtuality strategy for training construction hazard recognition. In *Construction Research Congress 2014: Construction in a Global Network* (pp. 2345–2354).

Cotterill, S. (2010). Pre-performance routines in sport: Current understanding and future directions. *International Review of Sport and Exercise Psychology*, *3*(2), 132–153.

Dangauthier, P., Herbrich, R., Minka, T., & Graepel, T. (2007). Trueskill through time: Revisiting the history of chess. In *Proceedings of NIPS* (pp. 337–344).

d_ijk_stra. (2012). *Statistical bonjwa theory: the history of bw.* Retrieved August 1, 2012 from `http://www.teamliquid.net/forum/viewmessage.php?topic_id=357124`.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795.

Elo, A. E. (1978). *The rating of chessplayers, past and present.* Batsford.

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.

Gobet, F., & Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology*, *43*(1), 159.

Gray, W. D., & Lindstedt, J. K. (in press). Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance. *Cognitive Science*.

Herbrich, R., Minka, T., & Graepel, T. (2006). Trueskill$^{\mathrm{TM}}$: A bayesian skill rating system. In *Proceedings of NIPS* (p. 569-576).

Hock-koon, S. (2012). Affordances of elliptical learning in arcade video games. In *Proceedings of DiGRA Nordic.*

Huang, J., Zimmermann, T., Nagappan, N., Harrison, C., & Phillips, B. C. (2013). Mastering the art of war: How patterns of gameplay influence skill in halo. In *Proceedings of CHI* (pp. 695–704). Retrieved from `http://doi.acm.org/10.1145/2470654.2470753` doi: 10.1145/2470654.2470753

Hullett, K., Nagappan, N., Schuh, E., & Hopson, J. (2012). Empirical analysis of user data in game software development. In *Proceedings of ESEM* (p. 89-98).

Keller, F. (1958). The phantom plateau. *Journal of the Experimental Analysis of Behavior*, *1*(1), 1-13.

Lee, M. J., & Ko, A. J. (2011). Personifying programming tool feedback improves novice programmers' learning. In *Proceedings of ICER* (pp. 109–116).

Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of SIGMOD Workshop on Research Issues in Data Mining and Knowledge* (pp. 2–11).

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions a meta-analysis. *Psychological Science*, *25*(8), 1608–1618.

Mason, W., & Clauset, A. (2013). Friends ftw! friendship and competition in halo: Reach. In *Proceedings of CSCW* (pp. 375–386). Retrieved from `http://doi.acm.org/10.1145/2441776.2441820` doi: 10.1145/2441776.2441820

Mesagno, C., Marchant, D., Morris, T., et al. (2008). A pre-performance routine to alleviate choking in "choking-susceptible" athletes. *The Sport Psychologist*, *22*(4), 439–457.

Reeves, S., Brown, B., & Laurier, E. (2009). Experts at play understanding skilled expertise. *Games and Culture*, *4*(3), 205–227.

Rengier, F., Mehndiratta, A., von Tengg-Kobligk, H., Zechmann, C. M., Unterhinninghofen, R., Kauczor, H.-U., & Giesel, F. L. (2010). 3d printing based on imaging data: review of medical applications. *International Journal of Computer Assisted Radiology and Surgery*, *5*(4), 335–341.

Scarr, J., Cockburn, A., Gutwin, C., & Quinn, P. (2011). Dips and ceilings: understanding and supporting transitions to expertise in user interfaces. In *Proceedings of CHI* (pp. 2741–2750).

Stafford, T., & Dewar, M. (2014). Tracing the trajectory of skill learning with a very large sample of online game players. *Psychological Science*, *25*(2), 511–518.

Thompson, J. J., Blair, M. R., Chen, L., & Henrey, A. J. (2013). Video game telemetry as a critical tool in the study of complex skill learning. *PLoS ONE*, *8*(9), e75129. doi: 10.1371/journal.pone.0075129

Weber, B., & Mateas, M. (2009, Sept). A data mining approach to strategy prediction. In *IEEE Symposium on Computational Intelligence and Games* (p. 140-147). doi: 10.1109/CIG.2009.5286483

White, R. W., Dumais, S. T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. In *Proceedings of WSDM* (pp. 132–141).

Xu, Y., Cao, X., Sellen, A., Herbrich, R., & Graepel, T. (2011). Sociable killers: Understanding social relationships in an online first-person shooter game. In *Proceedings of CSCW* (pp. 197–206).

Yan, E. Q., Huang, J., & Cheung, G. K. (2015). Masters of Control: Behavioral Patterns of Simultaneous Unit Group Manipulation in StarCraft 2. In *Proceedings of CHI* (pp. 3711–3720).

Zhao, D., Thabet, W., McCoy, A., & Kleiner, B. (2012). Managing electrocution hazards in the us construction industry using vr simulation and cloud technology. In *eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2012.* CRC Press Leiden, Netherlands.

Zois, J., Bishop, D. J., Ball, K., & Aughey, R. J. (2011). High-intensity warm-ups elicit superior performance to a current soccer warm-up routine. *Journal of Science and Medicine in Sport*, *14*(6), 522–528.