Masters of Control: Behavioral Patterns of Simultaneous Unit Group Manipulation in StarCraft 2

Eddie Q. Yan
Electrical Engineering
UCLA
eddieyan101@ucla.edu

Jeff Huang
Computer Science
Brown University
chi@jeffhuang.com

Gifford K. Cheung
Information School
University of Washington
giffordc@uw.edu

ABSTRACT
Most user interfaces require the user to focus on one element at a time, but *StarCraft 2* is a game where players often control more than a hundred units simultaneously. The game interface provides an optional mechanism called "control groups" that allows players to select multiple units and assign them to a group in order to quickly recall previous selections of units. From an analysis of over 3,000 replays, we show that the usage of control groups is a key differentiator of individual players as well as players of different skill levels—novice users rarely use control groups while experts nearly always do. But players also behave differently in how they use their control groups, especially in time-pressured situations. While certain control group behaviors are common across all skill levels, expert players appear to be better at remaining composed and sustaining control group use in battle. We also qualitatively analyze discussions on web forums from players about how they use control groups to provide context about how such a simple interface mechanic has produced numerous ways of optimizing unit control.

Author Keywords
video games; skill; control groups; player behavior

ACM Classification Keywords
K.8.0 Personal Computing: Games

INTRODUCTION
Games are interesting to study because of how they are free to evolve independently from and also influence conventional software interfaces [15]. Many competitive video games are currently growing in popularity. Often referred to as electronic sports, or esports, these games are defined by competitive international tournaments that draw millions of spectators with prize pools reaching millions of dollars [18]. Esports are also characterized by being entertainment not only for the player, but also for the spectator [4, 18]. In this work, we focus on a unit grouping interface feature in a popular esport title, *StarCraft 2*, a real-time strategy (RTS) game. *StarCraft 2*, like many other esports, present players with an intricate and demanding task that favors rapid context switching and mastery of the game interface.

CHI 2015, April 18–23, 2015, Seoul, Republic of Korea.

http://dx.doi.org/10.1145/2702123.2702429

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 videogames 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 RTS games, these features include sophisticated ranking systems that automate player matchmaking in online ranked games and the ability to record gameplay as sequences of commands performed by each player. For many games, a near-complete record of game state, user keystrokes, and mouse clicks are available from players at a wide variety of skill levels in the form of files known as "replays."

These games provide 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 as they jump around the map, better players use an interface feature called control groups to bind groups of units to single keys, and thus can issue commands to numerous units quickly. We aim to understand the characteristics of control group use among players of different skill levels in *StarCraft 2*.

The main contributions of our work include a quantitative and qualitative analysis of a variety of player control group behaviors at different skill levels. These behaviors include warmup, producing units, and rebinding units under both normal and pressured situations. We also investigate control group habits among players and use our understanding of control group behaviors to build two classifiers: one for player skill and another for identifying expert players. Finally, we discuss the characteristics of control groups that we consider relevant to theories in human-computer interaction and interface design.

RELATED WORK

Analysis of Gameplay Data and Player Skill

Many other modern multiplayer video game titles ship with the ability to record gameplay data and incorporate matchmaking features for players. Halo: Reach is an example of such a game and has been studied extensively in [8, 13]. As in Huang et al. [8], we explore the explanatory power of gameplay data for user skill. Like Halo: Reach, *StarCraft 2* also has built-in skill estimation capabilities that are used for matchmaking, which we relate to the features defined in *Metrics and Features*.

StarCraft 2 replay data has been studied previously to identify which characteristics explain a player's skill level [19]. Several features that were highly 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. We also extend this work by using our data to build a trinary classifier of player skill following the binary classifiers introduced in Thompson et al. [19]. 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 demonstrated a process for opponent modeling through data mining by analyzing *StarCraft: Brood War* replays [20]. We demonstrate a process of identifying players from anonymous replays which can enhance opponent modeling as it enables the focused study of a particular player's game to predict strategy.

In this paper we investigate whether it is possible to identify online players using their control group configurations as a practical application of understanding individual control group usage differences. This is similar to the question posed by Lauradoux [11] about identifying chess players from their in-game behaviors extracted from replays.

Understanding Task Transfer in the Real World

The predecessor to *StarCraft 2*, *StarCraft: Brood War*, has been used as a proxy for real world situations to improve communication, collaboration, and understanding in the process of collaborative task transfer [10]. In Kirsh and Rebotier [10], the effectiveness of methods such as annotated stills and annotated video in "passing the bubble" (switching control from one player to another in the middle of a match) was evaluated by measuring win/loss rates, game score, and knowledge of the game.

In our work, we investigate the differences among novice users (those uninitiated or potentially on the receiving end of the task transfer) and expert users (those potentially generating the information to facilitate task transfer). Our results suggest that task transfer is more awkward in the case of expert-to-expert task transfer, whereby each expert has already developed his or her own unique signature of gameplay that may be difficult to consciously modify.

Keystroke-Level Models

Keystrokes have been studied more generally in software applications. Regarding user identification, past work on keystroke-level modeling (KLM) has shown that skilled individuals develop characteristic patterns of input when typing text on a keyboard or tapping characters via a telegraph [12]. In Leggett et al. [12], digraph latencies were demonstrated to be promising features for both static and dynamic identity verification of users performing tasks using a keyboard. We relate this to players of an RTS game by investigating whether professional gamers also tend to develop distinctive styles of keyboard use. However, our analysis focuses on features that are subtly different from those used in KLM (see *Metrics and Features*).

For KLM-based analysis of interfaces, two of our major concepts promise to be especially relevant for time-pressured use cases [6]. First, we show that there is a marked distinction between experts and novice performance during time periods of high time-pressure. A KLM analysis of general user interfaces should anticipate similar drop-offs from novice users in the "heat of battle." Second, we investigate the idea of warmup actions in *StarCraft*, a concept that is especially interesting when applied beyond games. The idea of meaningless key presses is not addressed by the KLM model. The typical KLM analysis of an interface is based on the expectation that each key press is either meaningful or erroneous. However, in a game where victory or defeat are defined by micro-second precision, we find that meaningless warmup actions are applied regularly and that the best players are those who start keying without purpose and are able to seamlessly transition that rhythm of tapping from meaningless to meaningful game actions. This implies that the traditional KLM model can deeply benefit from a closer look at the relationship among habit, tempo, expertise, and meaningless actions.

CONTROL GROUPS

The class of commands we focus on in this analysis are those that apply to control groups. Control groups are commonly referred to within the RTS community as "unit bindings" or more colloquially as "hotkeys." This naming occurs because control groups are used by players to efficiently control and manage diverse groups of units within the game. Control 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 their *current selection*, a single buffer containing references to units currently controlled by a player. In order to control a unit not currently selected, players must update their current selection to include the desired unit before issuing commands.

Control groups are convenient because they allow users to rapidly switch their current selection to previously defined selections of units. Players can modify control groups by adding ("binding") additional selected units to a control group number or by replacing its selection with the current selection of units. Players can also recall the units assigned to a specific control group, which will update their current selection with the units assigned to the selected control group. Use of control groups is not required to play *StarCraft 2*, as players can simply manually select units each time using the mouse, but allows for faster context switching and command execution within the game. An example of a control group mapping is shown in Figure 1.

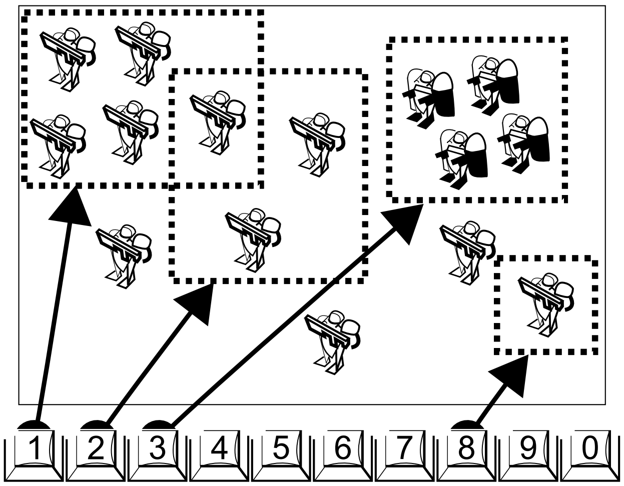

Figure 1: Control groups are bound to keyboard keys 0–9 and represent groups of units. Different groups can include the same units, or no units at all.

Comparisons to Other Applications

Control groups match a familiar interface design pattern: they are a set of buffers that can be updated, reset, and recalled. These interfaces are present in many applications outside of games. The ubiquitous copy-and-paste clipboard matches this pattern. Just as control groups allow a user to manage multiple sets of data, Microsoft's Office Clipboard history allow users to set and recall multiple clipboard items. Researchers continue to innovate on this clipboard: Faure et al. [5] introduce a keyboard-based interaction technique for navigating clipboard history via a long pressed Ctrl-V followed by repeatedly pressing the letter V to cycle through the items. This technique maps almost directly onto focus-cycling shortcuts in *StarCraft 2*. Other applications with a similar model include: desktop window managers where adding and removing windows from workspaces is analogous to adding and resetting control groups; advanced text editors such as Emacs with buffers and marks; and the group functionality in vector-based image programs (e.g. Adobe Illustrator, Microsoft Powerpoint).

METHOD

We use a mixed-methods approach involving quantitative analysis of data from replay collection websites, and qualitative analysis from online discussion forums about control group use. Using these two methods allows us to not only find out large-scale patterns of control group use, but also why players behaved in a particular way.

Quantitative Analysis

Replays of *StarCraft 2* games were retrieved from two sources and included player information that allowed us to discern the skill level of the source players.

To investigate how differences in player skill affect control group use, we examined data from players of three different skill levels {novice, proficient, expert}. We consider players in the bronze and silver leagues (roughly 0–30th percentile) of Blizzard's rating system as the novice class. We consider players in the platinum and diamond leagues together (roughly 60th–90th percentile) as the proficient class and players in the grandmaster league (top 0.5% of players and above) as the expert class.

For the novice and proficient classes, we downloaded replays from a popular replay aggregator website, GGTracker. We also collected other replays of players in the gold and master leagues from GGTracker, but these were excluded in our classification of player skill to avoid the problem of trying to classify players in adjacent leagues[1]. 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 representing the expert class were obtained from season 2 of the 2013 World Championship Series (WCS) tournament replay pack released by Blizzard [1].

Inclusion criteria for the replays obtained from GGtracker website:

- The replay has two players (1 versus 1 mode)[2]
- Both players have a profile located at us.battle.net[3]
- Both players must be a member of the expected league during the collection[4]
- Both players must have usernames not consisting entirely of I's, `|`'s, or l's [5]
- The match length must be at least 300 seconds (≈420 in game seconds)[6]

[1]Attempting to separate adjacent leagues of skill rating into distinct classes adds to the difficulty of the classification problem due to players are between adjacent leagues. We adopt Thompson et al.'s approach [19] and avoid differentiating between adjacent leagues.

[2]We are estimating 1v1 ratings; Blizzard's rating system assigns a different rating for each mode of gameplay e.g. 2v2, 3v3, etc.

[3]Different regions have separate ranking systems, and therefore potentially different definitions of leagues.

[4]We used the "Highest Career Finish" provided by the player's `battle.net` profile and replay metadata to determine this.

[5]A common tactic to obscure one's identity is to use a "barcode" username e.g. `|I|I|I|I|I|I|I`. While we demonstrate a classifier that is capable of de-anonymizing expert players, we avoid replays from barcodes because of the difficult in establishing truth labels to evaluate classification performance.

[6]We wish to avoid shorter games that may only exhibit transient control group behavior.

Inclusion criteria for the player data obtained from the 2013 WCS Season 2 replay pack:

- The player has at least two recorded matches in the same replay pack
- The player has an identifiable name (same as above)

In total, 3,316 replays representing 138 expert players, 800 proficient players (platinum: 400, diamond: 400), 802 novice players (bronze: 401, silver: 401) as well as 400 gold and 400 master league players were included in this analysis. We consolidated aliases or variations of usernames of known players. The data was parsed from the replay files using the sc2reader Python library [9].

Metrics and Features

We consider metrics to explore based on the variety of user behavior they expose. Studies on KLM have examined features such as digraph latencies [12] or the amount of time it takes a user to perform a specified command (e.g. keystroke or mouse click) [3]. However, input for *StarCraft 2* and many other video games is often flexible in terms of individual keystroke sequences. That is, commands involving control groups can be repeated, executed out of order, or sometimes not at all with no direct influence on the outcome of the game. For example, players have the freedom to essentially select a control group as many times as they want, as this does not affect the game unless an additional command is issued. In considering the freedom that such interface flexibility confers upon the users, we choose to invert the concept of latency used in KLM and consider the frequency of control group commands instead. This frequency is computed based on real time as opposed to game time (real seconds are longer by a factor of 1.4).

Our qualitative analysis indicates that *StarCraft 2* 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 control group 5 whereas another may always choose control group 3. Considering these tendencies, it seems appropriate to consider frequencies for each of the control groups separately, as control group usage may depend on how often the units bound to a control group need to be selected.

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

Together, these combinations yield 30 features per player per game, as we consider three types of control group commands: add, set, get, with 10 possible key bindings $\{0–9\}$ per command. Each feature was therefore the frequency that a specific control group and action combination was used in the game. These features are used for the classification tasks described later in the paper.

Defining Battles and Warmup

Some additional constraints were implemented to observe how players used control groups when in battle and during the warmup phase.

For our analysis, we defined battles as multiple units dying within a short period of time to include both small skirmishes and large engagements, both of which can tax a player's attention. To detect battles, we segmented the replay into periods of 10 in-game seconds (about 7 real seconds) and recorded the number of units owned by a player that were killed by an opposing player. Segments with more than two unit deaths were treated as times in which battles occurred. The window size and threshold on unit deaths were decided by reviewing replays at a variety of skill levels and was intended to capture as wide of a range of battles as possible. Peacetime is defined as the set of time windows that were not detected as battles. In some cases, replays were excluded from peacetime vs. battle analysis as either no battles were detected in them or because the replays had corrupted data. Players are excluded from the boxplots shown later in the paper if their ratios of command rates involve a divide by zero.

For battles and peacetime, we considered two subsets of control group use. We studied how often players selected the structures used to produce and upgrade units (macro) in and out of battle and also how often players modified or rebound their control groups, e.g. to manage newly produced units during battles and peacetime.

We define the warmup period as the first 120 in-game seconds ($\approx$86 real seconds). During this time period, players have only a few units to control. Still, players can choose to bind these units to control groups and rapidly cycle through them to warmup. We compared their warmup to their non-warmup (120+ seconds until the end of the game) control group usage.

Qualitative Analysis

Additionally, we conducted a qualitative analysis of control groups through a phenomenological lens that prioritizes the personal experience of the players as they reported in online discussions. This analysis was conducted in parallel to the data modelling to validate our choices of quantitative features and to inform our interpretations of the data and how to generalize our findings. The quotations throughout this paper are from this analysis and are used to illustrate the context of the features.

We collected forum posts about control group usage from the online social website Reddit. 50 posts with 776 total comments were analyzed via grounded theory [17] (Figure 2). A technical report comprising the data and coding scheme from the qualitative analysis is available online at `http://hci.cs.brown.edu/StarcraftReport.pdf`. All of the quotes listed as forum posts in this paper are taken from this collection.

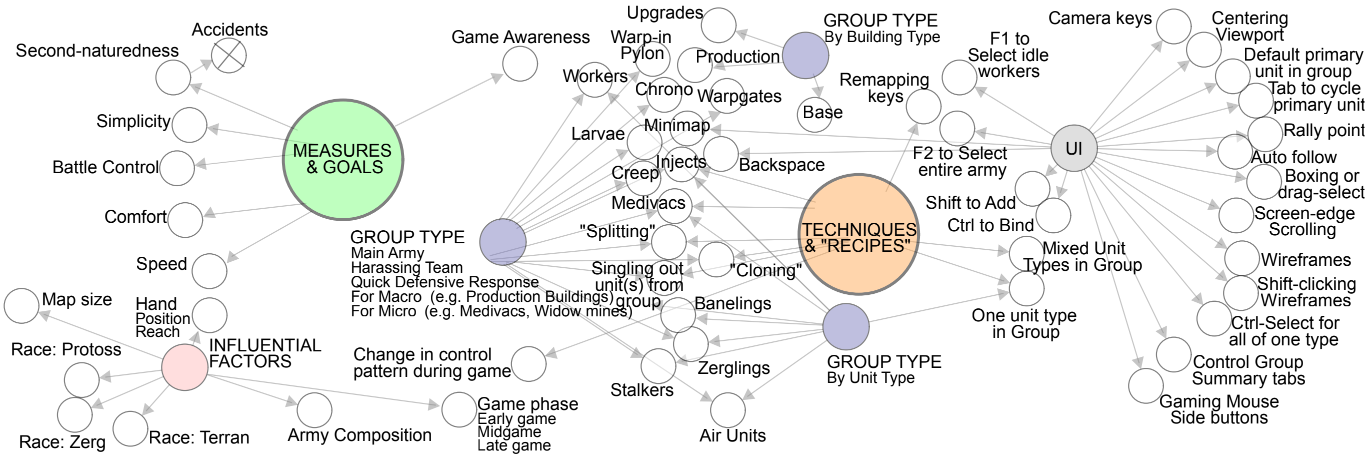

Figure 2: Diagram of themes and selected sub-categories from our qualitative analysis. This graph was manually created during the analysis. It shows major themes of habit, high technical skill, and diversity in technique.

DIFFERENCES BETWEEN SKILL LEVELS

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 nothing is really going on. The worst thing you can do is just to sit and watch ur base with 0 APM when nothing is needed to be done." —P1

Figure 3 shows the relationship between warmup and non-warmup control group use. The ratio between warmup control group use and non-warmup control group use peaks in players below the expert class. In the lower leagues, a few players bind units to control groups at the start of the game and then stop using them entirely. Other players, and in greater number, exhibit similar behavior: they use control 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 game by spamming excessively yet lack the ability to sustain the control group use rates thoughout the game as their attention is taxed. These players may be attempting to integrate control groups into their play (see warmup trends in Table 1) but have not yet sufficiently mastered them.

"I constantly spam [control groups] 5 and 6 checking my queens energy and only stop when I'm moving guys or building units." —P2

Aggregate Control Group Use

Overall, we find that players at higher skill levels tend to use control groups more frequently. Distributions of aggregate control group use are shown in Figure 4 and steadily move towards greater control group usage with increasing skill. We find that in both the novice and proficient classes, there remains a substantial proportion

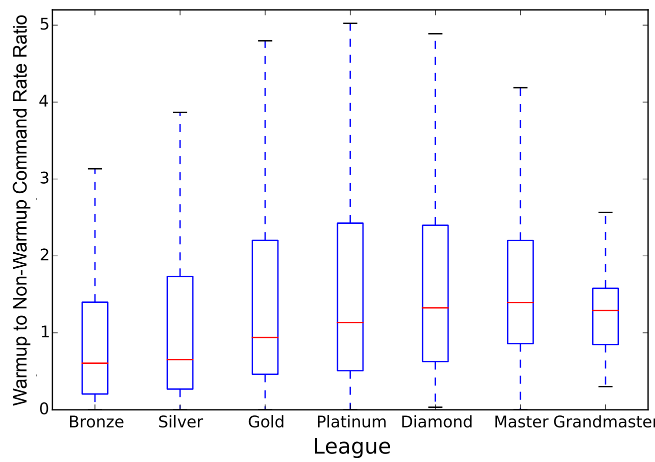

Figure 3: Warmup to non-warmup control group usage ratio at different skill levels. Some lower-skilled players use control groups at the beginning of the game but stop using them later. Expert players show the most consistent usage throughout the game.

of players who essentially do not use control groups at all. Within the expert class, everyone uses control groups to some extent, with the majority executing around two control group commands per second.

War and Peace, Macro and Micro

To gain a better understanding of control group usage and its relationship to player skill, we focused on finer-grained features that represent two distinct control group usage: macro and micro. Macro actions maintain the player's economy to keep income and production optimal: continually ordering new workers, buildings, and attack units. Micro actions optimize the effectiveness of individual units as they scout, position, harass, and fight. We investigate these features in two different forms of time-pressure in the game: battle and peacetime. Because it is

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 1: 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 control group usage throughout the game.

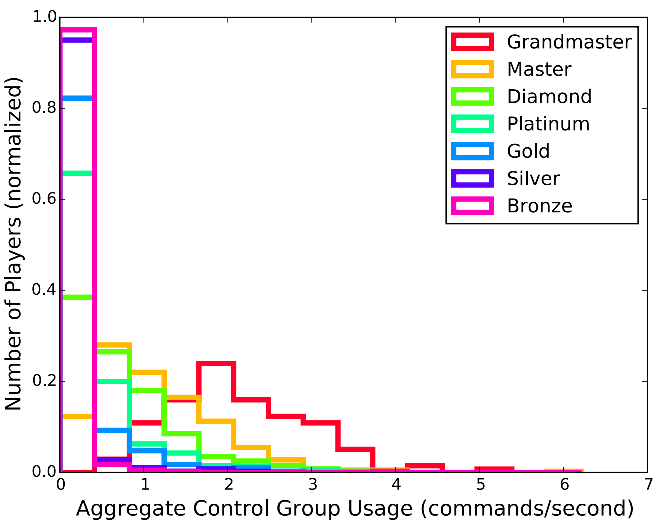

Figure 4: Distribution of control group usage among the leagues. A significant proportion of players in the novice classes do not use control group commands at all. Increasing rates of control group usage are seen with increasing skill ratings up to the expert class, where the average is around two commands per second.

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.

Use of Control Groups to Produce Units Under Pressure

"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]" —P3

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, expert players show the most frequent use of control groups to select production structures both in and out of battle, with lower usage rates in the lower leagues (Figure 5). Interestingly, the median event rate for expert players in battle is quite close to that of master league players during peacetime. Performing excessive or spamming selections of production structures can be helpful to monitor queues as it can ensure that the idle time of buildings is minimized. In the quote above, P3 habitually checks his different units and buildings in rotation. He also has trained himself 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.

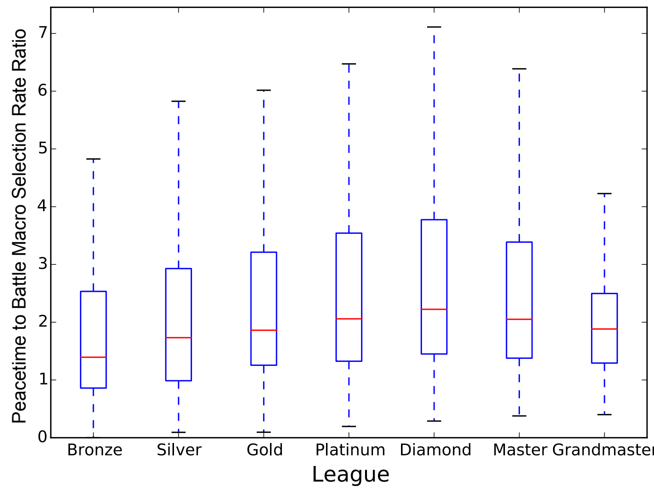

Figure 5: Ratios of macro selection rates via control groups during peacetime vs. battle. Nearly all players perform macro control group actions more frequently during peacetime.

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

Rebinding Control Groups in Battle

"want to reinforce [your army in the middle of a battle]? [Press these keys:] 1->s->zzzzzzzzz->shift+control+left click [eggs]->shift+[army control group] Blam .... the reinforcements will pop out immediately under your control. This is VITAL to Zerg play" —P4

StarCraft 2 is a fast-paced game, and during battles the number of units bound to a player's control 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 control 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 expert class 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 (Figure 6). This suggests that 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.

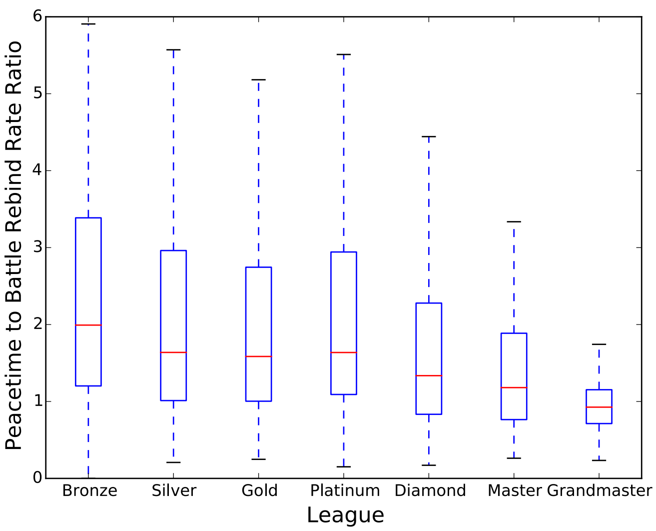

Figure 6: Ratio of control group rebind rates during peacetime vs. battle. Most expert players rebind control groups as frequently in battle as during peacetime. Players are able to better match their peacetime rebind rates in battle with increasing skill.

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 3: Median control group rebind rates during battle and peacetime (in units of commands/second). Rebind rates are highest in the expert class.

CONTROL GROUP HABITS

Our qualitative analysis found that players use control groups because they want to execute commands faster, to improve their ability to focus, grow a greater awareness of game events, and precisely manipulate game units. These mastery goals require interface control to become habitual and unobtrusive, and players practice heavily with the goal of making it natural to them.

As they improve, players develop a personal and routinised control style that fits their game strategy (race, opponent race, play style) and their physical needs (left/right-handedness, hand span). This relationship is maintained in two ways. Players configure their control to better execute their style of play; they depend on that configuration to keep them focused on their preferred style.

Frequently Used Keys

"Whatever you're comfortable with and can get to quickly. I basically don't use control groups past 5, simply because it's further away." —P5

The default keyboard shortcuts in *StarCraft 2* are heavily biased towards the left half of a QWERTY keyboard. Given this arrangement of shortcuts, we expected and found that keys {1–5} are the most frequently used while keys {6–9, 0} were less frequently used. This corresponds with our qualitative analysis which shows that players prioritize the closer keys to maximize their physical comfort when they expect to click on the keys for certain control groups at a high rate. Conversely, players will assign units to far away keys if they plan to recall them only on occasion.

How Habit Uniqueness Changes with Skill Level

Naturally, expert players are dissimilar or distant from players in the novice or proficient classes simply because experts tend to use control groups more frequently. However, this leads to a follow-up question: since experts use control groups more frequently, do they use them in similar ways? Or, are their styles of use unique?

Comparing player to player distances at different skill levels answers this question. We consider the Euclidean distance between players computed from the features described previously as a perspective on how similar two players are in their control group habits (Table 4). From the perspective of our features, expert players have the *most* distinct control group habits. This trend also appears in lower skill levels: the average distance between two gold players is also less than the average distance between two diamond players, and so on.

SKILL CLASSIFICATION

Classifying Skill from Control Group Usage

Our findings show that control group features contain predictive information about the skill levels of players. Applying this knowledge, we constructed machine learning classifiers using the support vector machine (SVM) implementation provided by the scikit-learn Python library [16].

League	Mean	Median	SD	Min	Max
Bronze	0.076	0.036	0.188	0.000	2.317
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 4: 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 control group usage habits.

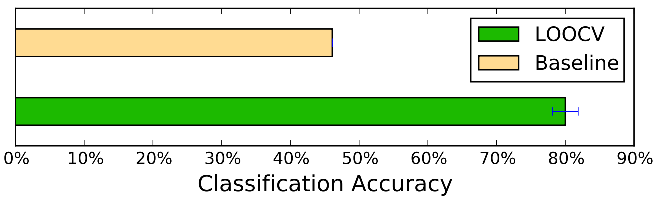

Figure 7: Skill level classification performance of {novice, proficient, expert} players estimated via LOOCV compared to baseline classification performance from choosing the most frequent class. Here, the error bar represents the 95% confidence interval for LOOCV accuracy.

Data was scaled before training and testing of the classifiers to have zero mean and unit variance by the preprocessing function for SVMs provided by the scikit-learn library. In the case of skill level classification, players who appear in multiple replays had their data averaged together such that each player is only represented once during the process. This averaging is done to help avoid "twinning" of the data [7], where cross-validation produces an overly optimistic estimate of accuracy due to highly similar instances of data being present in both the training and testing set. The default implementation of multi-class classification (one-against-one) in scikit-learn was used. The performance of our classifiers was evaluated using leave-one-out cross validation (LOOCV).

With an SVM-based classifier, we were able to achieve a LOOCV classification accuracy of 80.0% (95% CI 78.1%-81.9%) in the case of the trinary problem {novice, proficient, expert} compared to a baseline accuracy of 46.1% when choosing the most frequent class (Figure 7).

The results show that control group usage patterns have considerable predictive power to classify the skill level of players. The trinary classification performance when estimating players as {novice, proficient, expert} is comparable to the binary classification performance in [19]. This classifier enables a rapid (requires only a single match) and coarse estimation of skill that can supplement existing methods of matchmaking or skill estimation for online players. Instead of undergoing a series of placement matches to estimate player's initial skill rating, our method can be used to match players at a similar skill level after a single game.

Individual Variation and Expert Player Identification

"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." —P7

Our features show that players in the expert class tend to develop unique patterns of control group use. Additionally, the intra-cluster distance (distance between two points generated by the same player in two different matches) was on average significantly lower (mean = 0.359, SD = 0.272 commands/second) than the inter-cluster distance (distance between two players). Therefore, expert players not only tend to develop unique patterns of control group use, but also they remain reasonably consistent from game to game. In other words, expert players have signatures of control group behaviors that can be used to identify them.

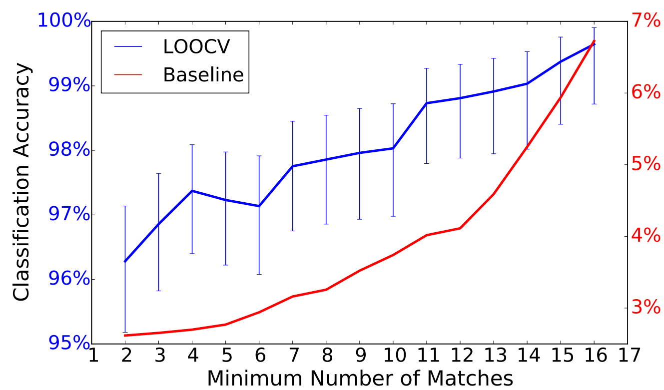

Figure 8: Classification performance of identifying expert players from control group features estimated via LOOCV compared to baseline classification performance. Here, error bars represent the 95% confidence intervals for the LOOCV accuracy. The minimum number of matches is an additional inclusion criteria for players.

Since our data includes expert players with varying number of examples (number of matches played), we consider our LOOCV classification performance as we increase the minimum number of matches required for a player to be included. With a minimum of 2 matches required per player, we were able to achieve a 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[7]. In general, classification performance improved as the number of games required for a player to be included was increased (Figure 8). As we increased the minimum number of games to 16 matches required per player, we were able to achieve a LOOCV classification

[7]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 et al. [2]

accuracy of 99.6% (95% CI 98.7%–99.9%) compared to a baseline accuracy of 6.7%.

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 a barcode username to obscure their identities. Our performance shows that accurate and rapid identification of experts is possible using our features. As online tournaments for esports increase in popularity as qualifying stages for larger events, we believe that 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 [20]

DISCUSSION: HABIT, WARMUP, AND TIME-PRESSURE

Many players practice gaming competitively, optimizing the user interface so to be as unique and idiosyncratic as themselves. This nature is reflected in narrow tasks such as rebinding new units into one's existing control groups. Doing so is not considered a creative or playful task in-game, yet personalization exists. Qualitative reports show multiple ways to do this one task and our replay data accurately fingerprints these users. The same tendencies are found in other games, but also in the repeated, habitual interactions for all software.

If these affordances in games reward practice, personalization, routinisation, and warming-up, interaction designers can look for similar affordances elsewhere. *StarCraft* allows players to warmup without consequences. Warmup sequences can be employed that are harmless and valid (the user interface responds to each warm-up command giving subtle cues that they are doing the right thing). How might an air traffic controller warm up when he starts his work? Perhaps, like the *StarCraft* player, he needs to go through the motions of selection, de-selection, binding, rebinding; or in the controller's case, when he arranges his flight strips, his communications equipment, how is he creating and maintaining a tempo of readiness?

Our findings about these habits explore several points: is spamming to warm up, defined as useless keystrokes to engage muscle memory, a form of practice that enables players to regain skill with an interface rapidly? Does staying active in periods of idle time improve performance when our abilities are taxed? The most frequently used keys are similar across players, but experts can still differ significantly from each other in their total control group command rates. Community norms such as using the first few control groups for armies also exist. Therefore we interpret this variation to be a result of personal spamming habits throughout the game. That is, does one player prefer to repeat the sequence "1,2,4" throughout the game while another repeats "1,3" over and over again? How frequently do they do so? It is not clear that spam actions, when ingrained to the point of habit, can be easily unlearned or adjusted.

The consistency that exists among the usage patterns of expert players is interesting because these players are not executing the same strategy (e.g. build orders) without fail for sixteen games in a row in a tournament setting. They are constantly forced to adapt their play to their opponent's race, style, and tendencies, yet the same usage patterns hold. This behavior suggests that experts shoehorn their control groups to their current build order and composition, perhaps as away of coping with the need to play quickly. If control group use were 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 control group habits, and their control 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.

We see promising future work in finding the value in meaningless warmup actions that, in actuality, allow users to exercise familiarity and to establish a routine or flow that they will then apply towards the real work that they intend to do. Software interfaces that do not afford warmup may themselves be obstacles to work they are meant to accomplish.

Returning to our related work, we can also comment on task transfer. We have identified features that should be addressed. Because experts are so personalized in their style, we cannot expect two expert-level users to transfer tasks without a drop in performance. Instead, we might expect wasted time as experts have to "wipe the slate clean." Alternatively, if two experts are similar in their personalization style, we may predict a smoother transition, meaning that in organizations where we have enough experts, our classifier can save time in transitions between job-shifts by recommending the best replacements for outgoing experts.

Furthermore, our work informs a time-critical perspective of interaction design [6] (and also, [14]'s concept of automatic negative affect). This is not limited to formal crisis management. Even everyday interactions have "peace"-time and "battle." A brother who is coordinating a plane ticket purchase with his and his sister's family is under pressure when is given only minutes to secure a good price. His expertise at navigating the purchase, researching, and communicating is taxed. Here we may understand that a user who has had less practice with the booking interface would have trouble multi-tasking, dealing with the micro-level intricacies of flight booking while processing the demands of his family. As we have found, habit, expertise, and time pressure are deeply intertwined.

CONCLUSION

We introduced a set of features in game replay data to describe control group use, then provided a broad overview of how players use control groups at various skill levels and a qualitative context for these trends. In addition, we showed that players at higher skill levels diverge in terms of their habits despite following overall usage trends. We then demonstrated applications for these features, including the rapid classification of player skill and the identification of expert players, the latter of which has significant implications for professional gamers. Finally, we turn to implications of these features on the area of interaction design in general, identifying a connection with warming-up, task transfer, and time-critical interactions.

As an optional user interface component for organizing units, control groups may seem to the outside observer to be nothing more than programmable keyboard shortcuts. However, when players are issuing commands as fast as they can physically do so and their human motor abilities reach their limits, mastery of control groups through habitual play extend the players' capabilities so they can perform menial tasks like managing their economy in the midst of battle. These findings offer guidance towards other interactive scenarios where our human abilities are taxed (even if for one, critical moment) and additional capacity may come from minor details in the user interface.

ACKNOWLEDGMENTS

The authors would like to thank the contributors of the sc2reader (Graylin Kim et al.) and GGTracker (David Joerg et al.) projects. The authors would also like to thank Robbie Nakamura of teamliquid.net for providing guidance on expert player identities.

REFERENCES

1. Blizzard. Wcs season 1 and 2 replays. `http://media.blizzard.com/sc2/replays/WCS-Season1-and-2-Replays.zip`. Accessed August 30, 2014.
2. Brown, L. D., Cai, T. T., and DasGupta, A. Interval estimation for a binomial proportion. *Statistical Science 16*, 2 (2001), pp. 101–117.
3. Card, S. K., Moran, T. P., and Newell, A. The keystroke-level model for user performance time with interactive systems. *Commun. ACM 23*, 7 (July 1980), 396–410.
4. Cheung, G., and Huang, J. Starcraft from the stands: understanding the game spectator. In *Proceedings of CHI* (2011), 763–772.
5. Faure, G., Chapuis, O., and Roussel, N. Power tools for copying and moving: useful stuff for your desktop. In *Proceedings of CHI* (2009), 1675–1678.
6. Fields, B., Wright, P., and Harrison, M. Time, tasks and errors. *ACM SIGCHI Bulletin 28*, 2 (1996), 53–56.
7. Forman, G., and Scholz, M. Apples-to-apples in cross-validation studies: Pitfalls in classifier performance measurement. *SIGKDD Explor. Newsl. 12*, 1 (Nov. 2010), 49–57.
8. Huang, J., Zimmermann, T., Nagapan, N., Harrison, C., and Phillips, B. C. Mastering the art of war: How patterns of gameplay influence skill in halo. In *Proceedings of CHI* (2013), 695–704.
9. Kim, G. sc2reader. `http://sc2reader.readthedocs.org/`, 2011.
10. Kirsh, D., and Rebotier, T. Passing the bubble: Cognitive efficiency of augmented video for collaborative transfer of situational understanding. Tech. rep., DTIC Document, 2003.
11. Lauradoux, C. Show me how you play and I will tell you who you are. In *Atelier Protection de la Vie Privée* (Soreze, France, June 2011).
12. Leggett, J., Williams, G., Usnick, M., and Longnecker, M. Dynamic identity verification via keystroke characteristics. *International Journal of Man-Machine Studies 35*, 6 (1991), 859 – 870.
13. Mason, W., and Clauset, A. Friends ftw! friendship and competition in halo: Reach. 375–386.
14. Norman, D. A. Emotional design. *Ubiquity 2004*, January (Jan. 2004), 1–1.
15. Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., and Fuller, T. User-centered design in games. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications* (2003), 883–906.
16. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830.
17. Strauss, A., and Corbin, J. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory.* SAGE Publications, 1998.
18. Taylor, T. *Raising the Stakes: E-sports and the Professionalization of Computer Gaming.* Mit Press, 2012.
19. Thompson, J. J., Blair, M. R., Chen, L., and Henrey, A. J. Video game telemetry as a critical tool in the study of complex skill learning. *PLoS ONE 8*, 9 (2013), e75129.
20. Weber, B., and Mateas, M. A data mining approach to strategy prediction. In *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on* (Sept 2009), 140–147.