Testing the Rehearsal Hypothesis with Two FastTap Interfaces

Carl Gutwin*, Andy Cockburn†, Benjamin Lafreniere*

University of Saskatchewan, Canada*, University of Canterbury, New Zealand†

ABSTRACT

Rehearsal-based interfaces such as Marking Menus or FastTap are designed to enable smooth transitions from novice to expert performance by making the novice's visually-guided actions a physical rehearsal of the expert's feedback-free actions. However, these interfaces have not been extensively tested in real use. We carried out studies of the adoption of rehearsal-based expert methods in two dissimilar applications – a game that directly rewards rapid selections, and a drawing program that has no particular need for urgency. Results showed very different patterns of use for the guidance-free expert method. In the game, participants quickly switched to sustained use of expert selections, whereas few users regularly used the expert method in the drawing program, even after ten weeks and more than 1800 selections. These studies show that rehearsal alone does not guarantee that users will switch to expert methods, and that additional factors affect users' decisions about what methods to use. Our studies also revealed several issues that should be considered by designers of rehearsal-based techniques – such as perceived risk in making selections without visual guidance, the value of guidance that shows possible options in the UI, and training that reminds users of an expert method and motivates its use.

Keywords: Rehearsal; expertise; interaction techniques.

1 INTRODUCTION

Expert interfaces – such as keyboard shortcuts, marking menus, or command maps – provide fast access to an interface's commands, supporting higher performance for experienced users. Previous research has shown that these expert mechanisms enable significant performance gains compared to navigation-based mechanisms such as hierarchical menus or ribbons (e.g., [18, 26, 28]).

These performance gains, however, are not always realized in the real world; there is considerable evidence showing that in everyday use, people do not use most expert mechanisms [5, 6, 13, 22]. Researchers have discussed several reasons for this lack of adoption – including satisficing (the feeling that current methods are good enough) [6], interface inertia (continuing to use a technique out of habit) and local optimality (a slower technique will be faster in any one instance than learning a new technique) [13].

An underlying issue for many expert interface techniques is that there can be substantial costs for the user – in time, effort, and lost productivity – when they switch to a new technique, even if that technique will be faster in the long run. Scarr and colleagues described this issue as the need to undergo a performance *dip* in order to reach a higher performance *ceiling* [29]. For example, switching from hunt-and-peck typing to touch-typing, or switching from a menu-based interface to a command language, may mean that the user becomes slower while they learn the new method, even if they will eventually realize better performance.

The need to suffer a performance dip, however, does not occur with all expert techniques. Researchers have designed several interfaces that provide a smooth transition from novice to expert behavior. A key principle in these designs is Kurtenbach's idea of ***rehearsal*** – that the physical actions taken by a novice to select a command should be the same as those taken by an expert (but with supporting feedback). As the novice learns the commands, they can start to make selections using the expert method, which is faster because there is no need to wait for feedback [18]. Marking Menus are a well-known example – a novice uses the technique with visual guidance from a radial menu, but once they learn the gestures that navigate to different commands, they can simply perform the gestures without waiting for the menu to appear [18].

Rehearsal-based interfaces are therefore intended to provide a smooth transition from a slow but reliable visually-guided mode of interaction to a fast memory-based mode. Proponents of these techniques suggest that ordinary use of the interface (with the novice method) will gradually build up memory that naturally leads to the expert mode, without incurring any performance dip. Several interface methods have been developed to exploit the intended benefits of rehearsal, including Marking Menus [18], FastTap [16], SHARK [34], ExposeHotKey [24], Octopocus [4], Gesture Guides [1], and more (reviewed later). FastTap, for example, uses spatial memory of a grid of commands as the expert selection method. Novices use the visual grid menu to find items, but once command locations are known, users can simply tap on the command's grid cell together with an invocation button to perform a selection [16].

While promising, the benefits of the rehearsal hypothesis have not been strongly validated for realistic interface use. In a limited case study of participants using Marking Menus in a real application, one participant switched to the expert method quickly, and the other did so after about 650 selections. Both participants then primarily used the expert menu, although they sometimes went back to the radial menu after time away [20]. However, this test was limited to two people, and a menu with only six commands.

In addition to there being little validation for the rehearsal hypothesis in realistic settings, previous research has suggested that deliberate practice [11] and mental effort [8] are key aspects of skill acquisition, and that guidance can inhibit skill acquisition [30] – findings that seem to run counter to the rehearsal hypothesis.

In order to better understand novice-to-expert transitions in rehearsal-based interfaces, we carried out two studies that expand substantially on previous evaluations – our studies involved 22 people, up to ten weeks of use, larger command sets, and two realistic applications (a game, and a sketching program). Both applications used a FastTap menu, and we tracked all novice and expert selections, as well as subjective and interview data.

In the first study, 12 participants played a game that involved making ≈300 selections from a 14-item FastTap menu in order to combat enemies that fell from the top of the screen. In the second study, 10 different participants drew a series of 30 pictures, over a ten-week period, in a drawing program that used a 24-item FastTap menu (about 1800 selections overall). There were no special requirements on how quickly the drawings were completed. A key difference between the first study (the game) and the second is the performance requirement for command selection – fast selection is a requirement in the game, but not in the drawing application.

Our results show dramatic differences in terms of adoption rates for the expert selection method between the two studies. In the game, people started using the expert selection method early,

Email: carl.gutwin@usask.ca, andy@cosc.canterbury.ac.nz, ben.lafreniere@gmail.com

quickly increased their rate of use, and maintained a high level of expert use (70% overall, with many participants using the expert method 100% of the time in the later stages). In the drawing study, however, adoption was dramatically lower – overall use of the expert method was only 11% of total selections, and although one person used it regularly, several participants used the expert method only a few times in the ten weeks. Furthermore, there was no clear increase in expert use over the duration of the drawing study.

These results suggest that the broader environment in which an expert interaction method is deployed plays a pivotal role in the adoption of that technique. In particular, it is possible that it is more than rehearsal that drives the level and pattern of adoption of the expert method.

Our work provides two main contributions. First, we provide new empirical evidence about the adoption of expert methods in realistic rehearsal-based interfaces, and we show that in some situations, adoption does not occur quickly or reliably. Second, we identify new factors that may affect adoption of expert methods, including a desire to avoid selection errors, and contextual factors, including the nature and performance requirements of the task at hand. Our results provide designers with important new information about the possibilities and pitfalls in supporting novice-to-expert transitions in realistic interfaces.

2 Related Work

2.1 Interfaces for experts

Understanding skill acquisition has long been a basic objective in psychology – Anderson [2] and Schmidt and Lee [31] provide extensive introductions focusing on cognitive and motor issues. In HCI, numerous techniques have been proposed to help users achieve higher performance, falling into four main groups [7]: *intra-modal improvement*, which aims to boost performance within the current interaction mechanism; *inter-modal improvement*, which involves improvements through switching to a faster style of interaction; *vocabulary extension*, which tries to increase users' knowledge of the commands that are available within an application; and *task mapping*, which involves improving the user's task comprehension or solution strategy.

Many different types of techniques have been suggested in these areas, including different training methods [7, 15], shortcuts for experts [26], memory-based retrieval interfaces [16, 28], adaptive interfaces [33], and task-based customization [3, 5]. Memory-based expert techniques – including keyboard hotkeys, gestural interfaces, command languages, and spatial-memory-based interfaces – have been shown to be particularly fast for experienced users. These techniques are rapid because they involve fewer and faster operations when a user is experienced – rather than navigating or searching for a command like novices do, the expert user can just remember and execute the command.

Importantly, these techniques normally involve inter-modal changes (users must switch from one interaction method to another to increase performance) – for example, switching from mouse-and-menu operation to hotkeys. Rehearsal-based techniques, reviewed below, seek to minimize this inter-modal transition.

2.2 Learning and skill development

Although a full review of learning is beyond the scope of this paper, several concepts are important to the research. First, interface learning can be organized into three stages as proposed by Fitts and Posner [12] – cognitive, associative, and autonomous. During the cognitive phase, users learn what the interface contains, and they rely on visual search to identify commands. During the associative phase, users know what the interface contains, and they begin to focus more on how the execution occurs. They begin to remember where in the UI each command is located, and can move there more and more quickly as they build experience. During the autonomous phase, people attain *automaticity* – they can execute commands quickly, without attention, and in parallel with other activities.

In addition, research has shown that although incidental learning is possible, particularly with spatial locations [27, 28], the depth of mental effort put into learning an interface can be correlated with their eventual memory of the interaction mechanisms (e.g., the gestures in a command set, or the location of items in the UI). Craik and Lockhart's "levels of processing" framework [9] suggests that a deeper, more effortful, mental encoding in memory leads to faster retrieval and longer persistence. The relationship between effort and learning has been demonstrated in research on learning object locations [10] and learning shape-writing [7]. In the motor learning literature, deliberate practice has been identified as a key requirement for acquiring expert performance [11].

2.3 Non-adoption of expert methods

Despite the increased performance ceiling of expert interfaces, several studies of real-world use show a tendency for users to persist with slower, suboptimal methods [5, 6, 13, 22]. Researchers have suggested several reasons for this phenomenon, including:

- *Satisficing.* Users may opt for a strategy that they know is "good enough" for their current purposes, even if they know that a better solution exists [32].
- *Paradox of the active user.* Carroll and Rosson [6] suggest that users who are engaged in ongoing tasks will often continue using known methods rather than learning new ones, and will generally apply known methods to new problem situations.
- *The value of feedback.* Fu and Gray [13] suggest that users can prefer well-practiced novice methods if these provide fast and incremental feedback (particularly in the associative phase).
- *The "guidance hypothesis*." Guidance provided to facilitate learning of an expert technique (e.g., feedback provided during an action) can become relied upon, degrading retention and performance when the guidance is no longer present [30].
- *Local optimality.* For any single action, using a known but slow mechanism is likely to be faster than learning a new one [14].
- *Performance dips.* Switching to a new interaction modality usually incurs a performance dip (as users must learn the new techniques); users may therefore be reluctant to switch because it means a (temporary) reduction in performance [29].

Researchers have considered several methods for helping users over these obstacles – for example, by punishing the use of the novice method [15, 17], by increasing awareness of the expert method [29], by providing feedforward to support expert command execution [4], or by showing the user how much their performance could increase if they switched to the expert method [25]. An alternate approach, however, is to design techniques that do not require overt methods of encouraging or forcing the user to switch to the expert method, and rather provide a natural and gradual transition from novice to expert behavior.

2.4 Avoiding the dip: Rehearsal-based interfaces

Several interfaces have been proposed that attempt to avoid the "performance dip" between novice and expert use. These systems use Kurtenbach's principle of *rehearsal* to enable knowledge transfer from novice to expert methods [18]. The principle states that novices should carry out selection actions in the same way that experts do; therefore, incidental learning will happen through everyday use, and as users gain experience with the interface, they will gradually build up the memory that they need to use the expert method. Feedback and guidance appear for novices, but as users become more experienced, these supports can be removed.

Kurtenbach explored rehearsal in detail with the Marking Menu technique [18, 19, 21]. Novices use this technique as a standard radial menu, in which the menu's visual representation appears a short time after the user holds their stylus down on the screen. As users gain experience with the locations of items in the menu, they can start converting the navigation motions needed to reach the item into a gestural "mark" – which can be performed without needing to wait for the visual guidance of the menu. Once expert, users simply draw the marks that correspond to the items they want to select, which is much faster [18].

Several other techniques have also used the principle of rehearsal. For example, the SHARK text input technique [34] allows users to move from touching individual keys on a virtual keyboard to shapes for words, where the shapes are a fast version of the novice's movement from key to key. Similarly, the ExposeHotKey system [24] allowed people to select toolbar items using the same mechanism as they used for hotkeys; as users learned the key combinations, they used the visual guidance of the toolbar less and less. Finally, the multi-touch FastTap technique [16] provided a spatially stable grid menu that is invoked using a thumb button; once users learn the spatial locations of commands, they can select those commands by touching both the invocation button and the command location with a simultaneous thumb-and-finger touch (which does not show the grid menu at all).

2.5 Real-world studies of rehearsal interfaces

The "rehearsal hypothesis" embodied by the interaction techniques described above has been shown to work in several lab studies (e.g., [4, 16, 18, 24]). However, the hypothesis has not been tested thoroughly in realistic systems and for realistic task scenarios. We are aware of only two studies that focus on real-world settings for these interfaces, both involving Marking Menus.

One study explored the use of several post-WIMP interaction techniques in a Petri Net editor [23]. The study did not track adoption of the expert method for Marking Menus, but did report that the different techniques were better suited for different kinds of tasks. For example, Marking Menus were seen as better for tasks in which a small set of actions were repeated frequently [23].

Another study looked specifically at adoption of the expert selection method in Marking Menus [20]. Two users carried out several hours of real work in a video annotation system, where six of the application's most common commands were included as a Marking Menu. The study showed that one user switched to the expert method very quickly, and used it more than 90% of the time after adoption. The other user took longer to start using the expert method (about 650 selections), but then also used the marking technique most of the time. For both users, however, when there was a time gap in use of the system, they went back to the menu method briefly – the authors suggest that the users needed to refresh their memory of the commands [20].

Although this study provides support for the hypothesis that a rehearsal-based interface can enable a smooth transition to expert use, it is very limited – there were only two users in the study, and the system used a very limited marking menu of only six items. In order to provide a deeper understanding of how rehearsal-based interfaces perform in realistic situations, we carried out two new studies that expand on previous work in terms of the number of participants, the duration of the study, and the size of the command set. It tracked users' transitions to expertise in two dissimilar applications – a game and a drawing application.

3 Study 1: A Game with High Urgency

Our first study used a custom multi-touch Android game called *Shape Slicer*, in which the user gestures to slice through enemy circles that fall from the top of the screen, after selecting the correct color (Figure 1 left). The game used a FastTap grid menu for selecting colors from a set of 12 colors (Figure 1 right). A faint grid on the game screen provided guidance about the location of items in the grid menu. The grid's bottom row contained a permanently-visible menu button showing the currently selected color, and two other commands that were not used for the game.

The interface allowed selection from the menu in two ways:

- *Novice* selections are made by first touching the menu button (Figure 1, lower left of the screen) with the thumb. After a short delay (150ms), the grid of items is displayed and the desired color is selected with a finger;
- *Expert* selections were made by simultaneously touching both the menu button and the desired color's location.

In the expert method of selection, the full menu is not shown at all, but the selected item is shown for 500ms after being touched, to provide feedback on what was selected.

In the game, 300 enemies with randomly-assigned colors were displayed falling from the top of the screen, in three waves of 100 enemies each. In the first wave, enemies fell slowly (about 4 seconds to reach the bottom of the screen); the second and third waves were successively faster (about 3 and 2 seconds to reach the bottom). Multiple concurrent enemies could appear on the display at a time. Once all 300 enemies had fallen, the game was over. The player's score indicated the number of enemies successfully sliced.

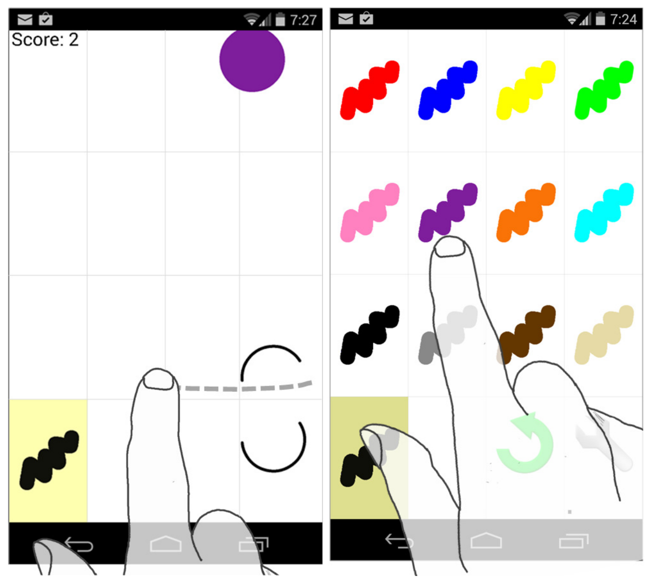

Figure 1: Shape Slicer game (left) and FastTap menu (right)

3.1 Study 1 Methods

3.1.1 Participants, Setup, and Procedure

Twelve participants (10 male, 2 female, mean age 30.4) were recruited from a local university. Participants were all experienced with touch devices (>30 min/day, mean 90 min.), and half had experience with games on touch devices (mean 15 min/day).

The study used a custom-built Android application – Shape Slicer – as described above. Participants carried out the study in a research lab – they first completed a demographic questionnaire and were shown the game (including both methods of using the FastTap menu); they then played the game, and finally filled out a post-session questionnaire. The study took 30 minutes in total.

During this study (and study 2), we exclusively referred to the expert selection method as "one-stage selection" and the novice method as "two-stage selection", to avoid biasing participants toward or away from using one technique over the other.

3.1.2 Design, data collection, and hypotheses

The primary dependent variable is the rate of use of the expert method (i.e., the proportion of selections completed using the expert method). This rate is analyzed using a 10×11 within-participants RM-ANOVA with factors *Block* (1-10) and *Command* (the colors used in the game).

Our primary goal is to test the rehearsal hypothesis, in three parts: first, that people switch to the expert method; second, that use of the expert selection method increases over time; and third, that having switched, users stay with the expert method. Secondary goals were to examine how the factors above varied by participant and command.

3.2 Study 1 Results

The expert selection method was used heavily by all participants from the early stages of the game. Over the entire study, the expert selection method was used for 68.8% of total selections (2582 times out of 3752 selections). The participants' mean rate of expert use was slightly lower at 61.7% (s.d. 45), with the difference stemming from some users not attempting to select certain items, possibly because they were overwhelmed by the game's requirements.

3.2.1 Changes in rate of use of the expert method over time

Participants' use of the expert method began early and increased consistently over the 10 study blocks (30 selections each). As shown in Figure 2, rate of expert use increased from 6.8% in the first block to 87% in the seventh block, giving a significant main effect of Block ($F_{9,99}$=27.53, $p<0.001$, η^2=0.71).

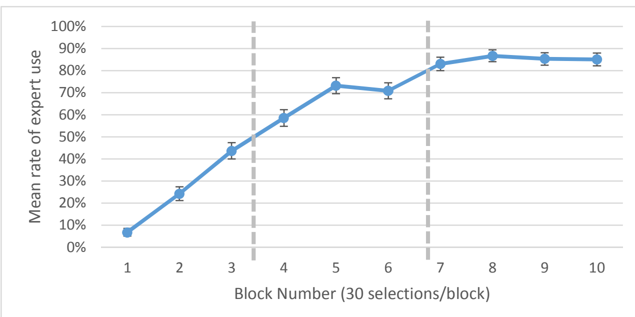

Figure 2: Rate of expert method use (±s.e.), by 30-selection block. Vertical lines indicate the three phases of the game.

3.2.2 Differences in rate of expert use between commands

There were twelve color commands in the game (the bottom row shown in Figure 1 were not used for the game). However, due to a logging error, Brown and Beige were logged as the same color. Our analysis of command differences therefore uses eleven colors.

RM-ANOVA showed a significant main effect of Command on expert usage rate ($F_{10,110}$=2.7, $p<0.01$, η^2=0.19), with mean rates ranging from 50% for Orange to 76% for Black (Figure 3). Command position in the FastTap grid may explain this effect (e.g., corner items may be easier to remember; see Figure 1).

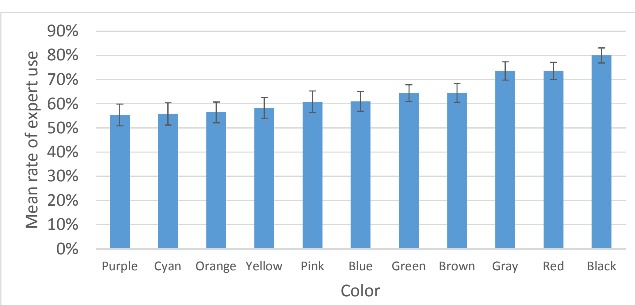

Figure 3: Overall rates of expert use (± s.e.) for each color.

We found no significant interaction effect between Command and Block ($F_{90,990}$=1.05, p=0.359), suggesting that the change in expert use over time was consistent for the different colors.

3.2.3 Individual differences in use of expert method

There were substantial differences in the overall rates at which participants used and adopted the expert selection method, as can be seen in Figures 4 and 5 below. The rate of expert use ranged from a low of 34% (P7) to a high of 85% (P6 and P11).

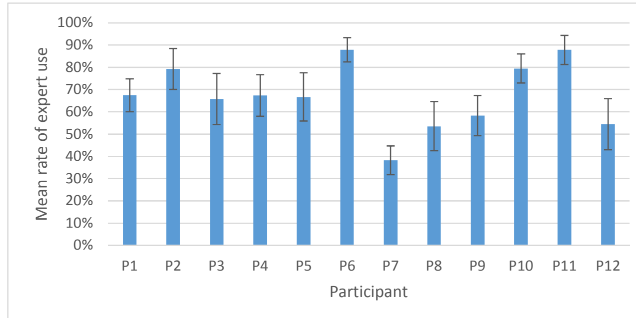

Figure 4: Rate of expert use (± s.e.), by participant.

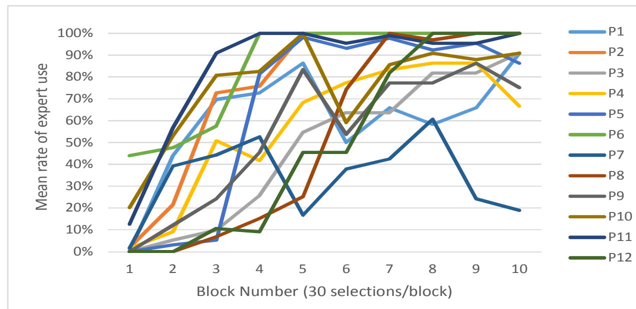

Figure 5: Rate of expert use, by participant and block.

3.2.4 Subjective results and participant comments

Participants' comments at the end of the study reinforce the performance results reported above, but also indicate some of the reasoning behind their behavior. Participants clearly indicated that they recognized the performance requirement in the game, and that they saw the expert ("one-stage") selection method as a way to achieve that performance. For example, in answer to the question "why did you use the one-stage selection method," one person stated "I had to use it when they dropped faster;" another said "I switched because I needed to be quicker," and a third person stated that the expert method "was necessary to improve speed."

It was also clear that participants saw the relationship between memorizing locations and the expert method – when asked about the timing of their switch to the expert method, several participants stated that they only switched once they were comfortable with the locations. For example, one person said "I only used [the expert method] when I was certain that the color was at that location;" another stated "once I was sure where some colors were, and once I started remembering where each one was, I used it more often."

We asked people which method they felt was fastest, which they felt was most accurate, and which they preferred overall. Participants felt strongly that the expert method was faster (12 of 12 people) and that the novice method was more accurate (10 of 12); overall preferences were in favor of the expert method (8 of 12). Reasons for preferring the expert method almost uniformly concerned speed, although one person also stated "I felt more engaged" with the expert method; reasons for preferring the novice method mentioned accuracy and the frustration of not being able to remember locations (e.g., “I was not able to remember all the colors and positions, while the two-stage can also provide accuracy;” “It was a lot more accurate and less frustrating than one-stage”).

4 Study 2: A Drawing App with Low Urgency

The second study involved a drawing program – *FastDraw* – in which participants reproduced three drawings each week for ten weeks (30 drawings total, see Appendix). Each drawing required approximately 60 different selections (including Undo), for a total of about 180 per week. The system was a custom-built Android sketching tool (Figure 6) with 21 drawing commands and 3 system commands, in a FastTap menu.

The FastTap menu worked in the same way as described above for the game study: novice users held down the menu button with their thumb, waited 200ms for the grid of items to appear, and then touched the desired drawing command with a finger; expert users could touch the menu button and the command’s location simultaneously. There were three state variables that could be set through the menu: the drawing tool (paintbrush, line, circle, rectangle, or bucket fill); the color (eight presets, a custom-color option, and a color picker); and the line thickness (from fine to wide). There were two other non-state drawing commands – Undo and Clear.

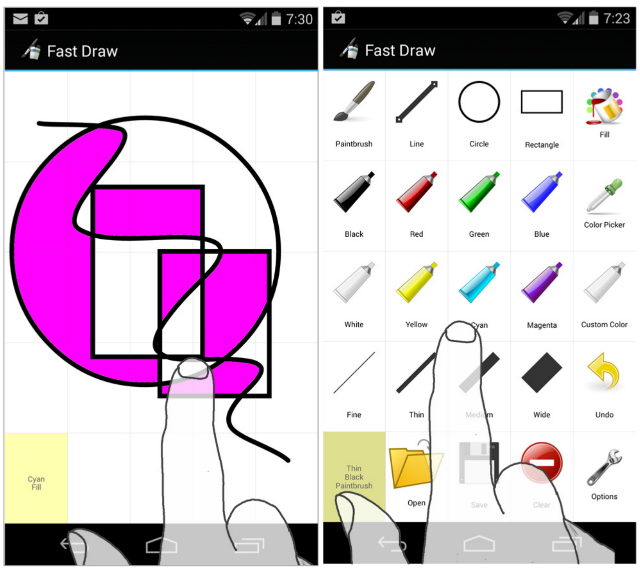

Figure 6: FastDraw drawing (left) and FastTap menu (right).

4.1 Study 2 Methods

4.1.1 Participants, Setup, and Procedure

Ten participants, none of whom participated in study 1, were recruited from a local university (7 male, 3 female, mean age 26.8). Participants were all experienced with touch devices (>30 min/day, mean 120 min.).

At an initial meeting, the experimenter installed the software on each participant’s personal Android device, and explained its use (including showing the participant the novice and expert methods of using the FastTap menu, following the same procedure as in study 1). Participants then carried out the study in their everyday environments on their own devices, completing three drawing tasks each week for ten weeks (see Appendix). The experimenter emailed three new pictures to reproduce at the start of each week, and the participants replied with their drawings and the system’s log files within a week’s time. After weeks 3, 6, and 10, participants completed short questionnaires about their use of the system.

4.1.2 Study design and hypotheses

We organized our analysis by week, using a 10×24 within-participants RM-ANOVA with factors *Week* (1-10) and *Command* (the 24 commands in the menu). The dependent measure was the method used to make selections.

Our goals were the same as for study 1 – to determine whether participants switched to expert mode, their rate of expert use over time, and whether they stayed with the expert method once it had been learned. We also examined how these factors varied by participant and command.

4.2 Study 2 Results

Participants did use the expert method, although far less than in study 1. In contrast to the first study, where expert selections made up more than 68% of total selections, in the drawing study the expert method was used in only 11.5% of selections (2,161 of 18,845). Furthermore, this rate was heavily influenced by one participant who used the expert method frequently – the mean rate of expert use per participant was only 5.1%.

4.2.1 Changes in use of the expert method over time

There were differences in expert use from week to week, but the pattern of adoption (Figure 7) was very different to that of study 1 (Figure 2). As expected, participants initially used the novice method almost exclusively, but gradually increased their use of the expert method until week 5, where expert usage peaked at about 10%. Use of the expert method then decreased to about 4%.

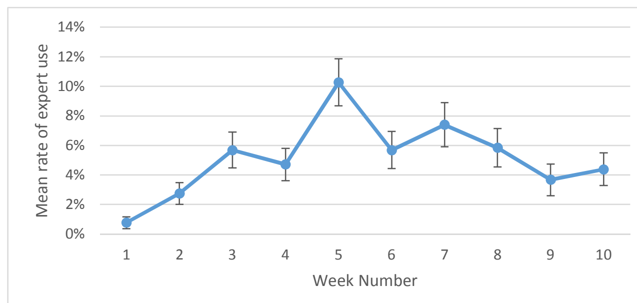

Figure 7: Rate of expert method use (± s.e.), by week.

This weak and variable adoption of the expert method is reflected in the absence of an effect of Week on expert use (RM-ANOVA, $F_{9,81}$=1.22,p=0.29). There was, however, a Week×Command interaction ($F_{207,1863}$=1.47, p<0.001), discussed below.

4.2.2 Differences in expert use between commands

A RM-ANOVA showed a significant effect of Command ($F_{23,207}$=16.32,p<0.001), with expert usage rates varying from 2% with Yellow up to 16.2% with Undo (Figure 8).

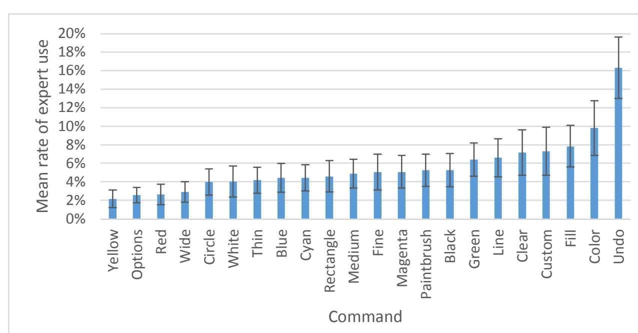

Figure 8: Mean rates of expert use (± s.e.), by command.

A scatter-plot of mean weekly usage of each command (Figure 9) against expert usage rate shows that Undo was used substantially more than other commands, helping to explain its much higher expert use. However, for the other commands there was only a weak relationship between weekly usage count and expert rate (R^2=0.07).

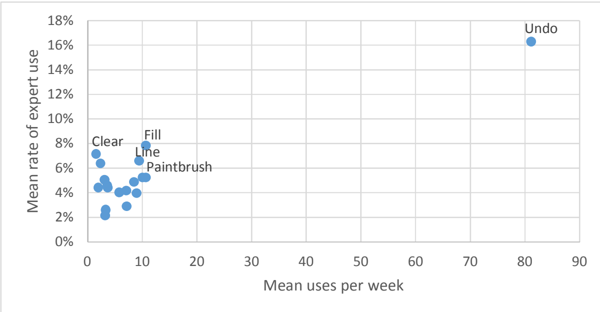

Figure 9: Rate of expert use vs. amount of command use

4.2.3 Individual differences in use of expert method

There were large differences in expert use between participants. For example P8 used the expert method for 33% of selections, P1 for ≈10%, and several participants for ≈1-3%. One participant never used the expert method. In addition, there were large differences between participants in their expert selection usage over time. As shown in Figure 10, participants tried the expert method at different times: P8 showed early and consistent expert use (but with a drop at the end), P1 used the expert method heavily in week 5 and then returned to a lower level of use, P6 showed a gradually-increasing trend in the second half of the study, and P2 had a sharp increase in week 10.

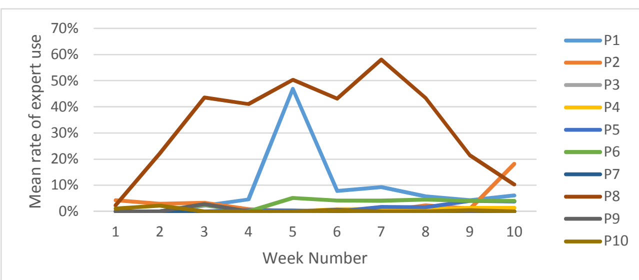

Figure 10: Rate of expert use, by participant and week.

4.2.4 Subjective responses and participant comments

We asked participants to fill out questionnaires about their experience in the study at weeks 4, 7 and 10 (Table 1). Participants stated that they primarily used the novice ("two-stage") selection method (median response of 7 on a 7-point scale) and that they rarely used the expert ("one-stage") method (median of 1/7 for the first two questionnaires, rising to 2/7 at the end). However, it was also clear that participants felt that selecting commands with the novice method was both easy (medians of 6/7 on all three questionnaires) and fast (medians of 5/7, 5/7, and 6/7).

Participants also felt that they were improving with the drawing application, and rated themselves as familiar with the locations. However, some comments suggested that participants were not highly confident about the layout of the menu, as discussed below.

We also asked participants why they preferred one technique over the other. There were several reasons given:

Lack of perceived benefit. Several people said they did not see any real benefit in switching to the expert method – one said "I almost exclusively used the overlay interface, as it wasn't worth it to me to shave off a second from my drawing time to try out the shortcut method." Another stated "I think at the end of the day there was just not enough incentive or motivation for me to try to experiment with and master the shortcut method." Others said "The overlay interface worked just fine for me, and I saw no reason to do the other method;" and "I didn't use the shortcut method because I felt no need to. The overlay method was good enough for the task."

Potential error and effort cost. The lack of a perceived benefit was also coupled with potential costs with the expert method, and several participants indicated that they had considered the tradeoff. In terms of the effort required, one person stated "the mental strain of thinking about where the option is made the time savings not worth it;" another said "It would have been much more work to purposefully memorize all the items only for half an hour a week."

Question and Scale	Median Response W4	W7	End
How much have you improved with the drawing app? (1=None; 7=A large amount)	5	6	6
How familiar are you with the menu and tool locations? (1=Not at all familiar; 7=Very familiar)	5	6	6
How easy is it to select tools? (1=Very difficult; 7=Very easy)	6	6	6
How fast is it to select tools? (1=Very slow; 7=Very fast)	5	5	6
How often do you use two-stage selection? (1=Almost never; 7=Almost always)	7	7	7
How often do you use one-stage selection? (1=Almost never; 7=Almost always)	1	1	2

Table 1: Median responses in week 4, 7, and end of study.

Participants also mentioned error costs – for example, "I was more concerned with accuracy than speed, and making a mistake was more annoying than taking an extra few seconds to complete the task;" and "I wouldn't want to make an error and keep trying because that would most likely take longer than if I just used the simpler overlay method." In contrast, the one participant who used the expert method frequently had a different view on errors: "Making errors was never a problem. The taps were so quick that I found making an error in choice was inconsequential […] the shortcut method was so quick that I was comfortable enough to play guessing games to test my memory of the command locations."

The Save and Send functions, however, had a much higher error cost, and this led to a different calculation of risk. Even the one participant who used the expert method frequently said "I never ever ever used the shortcut method when dealing with the commands to save, clear, or send my data. They were too critical for me to blindly tap, whether I trusted my memory or not."

Using the menu for planning. Two participants stated that they liked having the entire menu open so that they could see all of the tools, and plan what they were going to do next. One participant stated "one reason I used this method more often was the nature of the tasks […] it took a moment to decide which shape/color/size would be the best to use next and it helped to have the menu open to use as a visual reminder of what was available." Another said "if I want something to draw with, I need three things (line thickness, color, and type), therefore it is easier to be able to see all my choices and select the three while the overlay menu is out."

Frequency of use. A few participants stated that they did not use the tools enough to switch – e.g., "I didn't use any one tool often enough for the chord to stick in my head and become habit." The one exception was Undo; several people noted that they used this command with the expert method (e.g., "I only used the shortcut for undo near the end; easy since it is only one thing to select").

Lack of exposure. Although all participants received training on the expert method at the start of the study, some stated that they did not remember it well once they were back in their everyday environments. One said "I didn't really understand what the shortcut is in the beginning. It wasn't until the first survey that I knew about the shortcut. By then I was constantly using the full overlay interface anyway." Others suggested that if they had received more training, they might have been more likely to use the expert method: "another reason [for not switching] was there wasn't a tutorial to practice the specific chords […] I think it would have reinforced the usefulness of the technique in future weeks."

5 Discussion

The key findings from our two studies are as follows:

- In the game, people quickly learned and switched to the expert selection method, and then used it predominantly for the remainder of the time (over 60% use overall);
- In the drawing app, most people did *not* use the expert mode often (5% overall), and use rates did not increase substantially over the ten weeks of the study;
- In both studies, there were differences across participants – but these were much larger for drawing, where one person used the expert method as much as the all the rest;
- Participants in the first study saw switching to expert mode as necessary in order to succeed in the game, whereas in the second study, users saw little benefit to switching.

In the following sections, we discuss possible explanations for these results – in particular, why did rehearsal not lead to adoption (i.e., why was there was such a difference in expert use between the two UIs, both of which used rehearsal) – and then consider the implications of these results for the design of expert interfaces in realistic settings.

5.1 Why was adoption so different in the two UIs, even though both provided rehearsal?

FastTap is designed to exploit the rehearsal hypothesis in order to promote a natural transition to the expert selection method, but the lack of adoption in the drawing study suggests that there is still a barrier at the "switching point" in the transition to expert use, and that the transition is not as smooth as has been suggested. Participants' comments suggest that there was a definite cost to switching between novice and expert selection, and the drawing study suggests that participants perceived the costs of switching as substantial. Together, these findings suggest that users do not become experts simply by happenstance – rather, an expert method must present value to the user that offsets the costs of adopting that method. In this section, we discuss potential costs and contextual factors that may play a role in adoption of the expert method.

5.1.1 Physical vs. mental rehearsal

A potential source of cost for the expert method is the different mental processes required by the two selection modes. In FastTap, the *physical* act of performing a novice selection is a rehearsal of the corresponding expert selection, with the only difference being the time between activating the menu button and selecting the desired item, but the *mental* processes involved in these two selection types are not the same. In a novice FastTap selection, the user does not need to retrieve the location of an item from memory before they open the menu – they open the menu and then perform a visual search. In contrast, an expert selection requires that the user recall the target item's location from memory before or during the selection gesture. This difference in the mental procedure required by the two selection methods may impose an additional cost on switching because a change must be made in how the action is performed mentally. This explanation shares similarities with the "guidance hypothesis", which suggests that the presence of visual guidance can become part of a task being learned, and hinder performance when that guidance is no longer present [30].

In addition to requiring a switch in how the task is performed, there is a learning cost to making the switch; memorizing spatial locations is more difficult than finding an item in a visual space, and although incidental learning of spatial locations does occur [10, 27], our participants did not appear to learn the command grid over the ten weeks of the drawing study without expending explicit effort. The larger command set in the drawing application may also play a role here – the 24-item grid was larger than the game's 14-item grid, and much larger than the study of Marking Menu adoption that only used six items.

5.1.2 Distaste for errors

A number of participants commented that they stayed with the novice selection method because they were more confident with it and were less likely to make errors. This finding is similar to that of past work by Fu and Gray, who found that users preferred methods that provide fast, incremental feedback [13]. The desire to avoid errors is particularly interesting in our study because most of the commands in the drawing application were not destructive – they simply set the properties of the drawing pen (such as color, shape, and line width). Thus, making an error in these settings was easy to fix – as discovered by one user of the expert method who said "the taps were so quick that I found making an error in choice was inconsequential". Despite this, several participants were still wary about selecting an unintended item.

Part of the problem with errors, even if they are easy to fix, is that they add to the cost of the memory-based technique. As participants mentioned, making two or three selections with the expert method (which might be required if memory was not perfect) took longer than the equivalent selection with the novice method (which was essentially error free).

5.1.3 Contextual factors: Performance requirements, satisficing, and local optimality

Perhaps the biggest difference between the two studies was the context in which selections were made – in the context of playing a game or using a drawing application.

In particular, there was a clear difference in the performance requirements for the two studies. In the game, quick selections were needed to do well, particularly as the game progressed and the enemies fell faster. In contrast, in the drawing study there was no real pressure to make selections quickly – although faster selections would mean that participants could get the task done sooner, there was no requirement to do so. This may have motivated participants to put in deliberate effort to learn the expert selection method in the game, but not in the drawing application.

Stated in another way, the novice selection method may have been sufficient for the drawing app context, with the short 200ms timeout for opening the menu not enough to prompt users to adopt a higher-performance method. This was reflected in several comments (e.g., "the overlay method was good enough"). In a low-urgency setting such as the drawing study, having a novice interface that was above people's subjective performance threshold may tip the cost-benefit analysis in favor of staying with the interface that is already known (i.e., the "local optimality" hypothesis).

The context of playing a game versus using a drawing application may have influenced participants' behavior in other ways as well. Games are typically designed to present achievable goals, and participants may have viewed adoption of the expert selection method as part of advancing in the game, motivating them to put in deliberate effort. In contrast, in the drawing study, adoption of expert selection may have been perceived as incidental to the task at hand.

The higher-level task being performed by the user may also have played a role. Two participants in the drawing study mentioned that they preferred the novice method because it showed the entire command set, and this helped them see what was possible. This may suggest that participants were still in an early part of the associative phase of learning [12], and had not fully internalized the range of the UI's 24 commands. It may also suggest that viewing the menu provided additional value for planning out actions in the drawing app. Investigating the influence of higher-level tasks and contextual considerations on the choice of selection methods is an interesting area for future work.

Finally, the game and drawing apps differed in the number of items presented in their menu (14 vs. 24), and the environment in which study tasks were performed (in a lab vs. outside the lab). As mentioned above, the larger command set in the drawing application may have influenced the rate of learning of item locations. The environment may also have had an influence on participants' behavior, potentially encouraging them to put in more deliberate effort in the game condition. In future work, we plan to test the effects of these factors.

5.1.4 "Backsliding"

Our log data shows that even when participants did try out the drawing application's expert method (e.g., in week 5, perhaps prompted by the week-4 questionnaire), they often returned to the novice method afterward. This "backsliding" behavior is surprising; the rehearsal hypothesis suggests that once people transition to using the expert method for an item, their memory for that item will persist and become stronger, allowing the expert mode to become dominant. There are a number of possible explanations for why this did not occur. It may be that participants temporarily experimented with performing the expert method without having learned the item locations. It could also be that, having tried the expert method, they decided that its benefits were not worth the costs in a low-urgency drawing scenario.

5.2 Are expert UIs worth it for low-urgency tasks?

Our study findings suggest that a rehearsal-based UI does not automatically mean that people will adopt an expert method. However, even in situations of limited adoption, there are several reasons why expert interfaces – and rehearsal-based methods in particular – can be valuable additions to applications.

- *No harm in the expert mode.* The expert selection method did not cause difficulty for participants in the drawing study, and so it can be seen as a transparent mode that can be used when desired, without detracting from normal use.
- *Performance requirements can change.* Although our drawing study did not require high performance, other tasks – even with the same application – could make stronger demands. This may have been what occurred in Kurtenbach's study of Marking Menu adoption; the system he tested did not have intrinsic performance requirements, but as it was used for paid work his participants may have perceived a strong performance incentive.
- *Application use is long term.* Our ten-week study showed that incidental learning of spatial locations occurs very slowly if commands are not used frequently – but ten weeks is a small amount of time in the overall use of an application. People have accumulated years of use with many mobile applications and it may be that the slow progress of incidental learning could lead to substantial expert use, even in low-urgency tasks.
- *Some users like the expert mode.* The large individual differences in the drawing study suggests that alternative or advanced techniques are valuable in permitting different people to use the interaction styles that best suit them.

5.3 Directions for future research

There are several questions raised by this research that will lead to further studies. First, would initial practice with the expert mode at the start of the drawing study have led to increased expert use? As one participant suggested, we could have had an intensive training session at the start of the study (e.g., a "Drawing Tool Slicer" game with high performance requirement). Given the quick adoption of expert selection in the game, it is possible that intensive training could quickly overcome the learning costs to adopting expert selection.

Second, will learned expert use decay when performance requirements are reduced? If there is initially a need for high performance that prompts strong adoption of expert methods, will the high rates of use persist when the task becomes less urgent? This could be tested with our game by adding an easy wave of enemies at the end of the game, and observing if participants maintained expert selection, or switched back to novice selection.

6 CONCLUSIONS

Rehearsal-based interfaces such as Marking Menus or FastTap are designed to provide a smooth transition to expert use. However, these interfaces have not been extensively tested in realistic scenarios. We carried out two studies that tracked adoption of expert methods in two very different rehearsal-based applications – a game with a high performance requirement, and a low-urgency drawing program. We found that whereas participants quickly switched to the expert method in the game, very few participants regularly used the expert method in the drawing program – even after ten weeks and 1800 selections. These studies show that the principle of rehearsal alone does not guarantee that users will switch to expert methods – and that there are several factors that affect users' decisions about what methods to use. Designers of rehearsal-based interfaces, and memory-based interfaces more generally, should think about the issues raised in our research – such as user perception of the risk of making errors with a memory-based technique, the value of showing the entire interface to users who want to see possible options, and training that reminds users of the expert methods (and possibly provides practice sessions with high performance requirements).

REFERENCES

1. Anderson, F. and Bischof, W. Learning and performance with gesture guides. *ACM CHI 2013*, 1109-1118.
2. Anderson, J. *Learning and Memory*. Wiley, 1995.
3. Banovic, N., Chevalier, F., Grossman, T., Fitzmaurice, G. Triggering triggers and burying barriers to customizing software. *ACM CHI 2012*, 2717-2726.
4. Bau, O. & Mackay, W. OctoPocus: A dynamic guide for learning gesture-based command sets. *ACM UIST 2008*, 37-46.
5. Bhavnani, S. and John, B. The Strategic Use of Complex Computer Systems. *HCI 15* (2000), 107-137.
6. Carroll, J. and Rossen, M. Paradox of the active user. in Carroll, J. ed. *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*, MIT Press, 1987, 80-111.
7. Cockburn, A., Gutwin, C., Scarr, J., and Malacria, S. Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys 47* (2): 31:1-36.
8. Cockburn, A., Kristensson, P., Alexander, J. and Zhai, S. Hard Lessons: Effort-Inducing Interfaces Benefit Spatial Learning. *ACM CHI 2007*, 1571-1580.
9. Craik, F. and Lockhart, R. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior 11* (1972), 671-684.
10. Ehret, B. Learning where to look: Location learning in graphical user interfaces. *ACM CHI 2002*, 211-218.

11. Ericsson, K. A., Krampe, R.T., and Tesch-romer, C. The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 1993, 363-406.
12. Fitts, P. and Posner, M. *Human Performance*. Brookes/Cole, 1967.
13. Fu, W. and Gray, W. Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science 28*, 6 (2004), 901-935.
14. Gray, W., Sims, C., Fu, W.. and Schoelles, M. The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review 113* (2006), 461-482.
15. Grossman, T., Dragicevic, P., Balakrishnan, R. Strategies for accelearating on-line learning of hotkeys. *ACM CHI 2007*, 1591-1600.
16. Gutwin, C., Cockburn, A., Scarr, J., Malacria, S. Faster command selection on tablets with FastTap. *ACM CHI 2014*, 2617-2626.
17. Krisler, B Alterman, R. Training towards mastery: Over-coming the active user paradox. *NordiCHI 2008*, 239-248.
18. Kurtenbach, G. *The Design and Evaluation of Marking Menus*, PhD Dissertation, Computer Science, University of Toronto, 1993.
19. Kurtenbach, G. and Buxton, B. The Limits of Expert Performance Using Hierarchic Marking Menus. *INTERCHI 1993*, 482-487.
20. Kurtenbach, G. and Buxton, W. User learning and performance with Marking Menus. *ACM CHI 1994*, 258-264.
21. Kurtenbach, G., Sellen, A., Buxton, W. An empirical evaluation of some articulatory and cognitive aspects of Marking Menus. *HCI* 8,1 (1993), 1-23.
22. Lane, D., Napier, H., Peres, S. and Sandor, A. Hidden costs of graphical user interfaces: Failure to make the transition from menus and icon toolbars to keyboard shortcuts. *IJHCI 18*, 2 (2005), 133-144.
23. Mackay, W. Which interaction technique works when?: Floating palettes, marking menus and toolglasses support different task strategies. *AVI 2002*, 203-208.
24. Malacria, S., Bailly, G., Harrison, J., Cockburn, A. and Gutwin, C. Promoting hotkey use through rehearsal with ExposeHK. *ACM CHI 2013*, 573-582.
25. Malacria, S., Scarr, J., Cockburn, A., Gutwin, C. and Grossman, T. Skillometers: Reflective widgets that motivate and help users to improve performance. *ACM UIST 2013*, 321-330.
26. Odell, D., Davis, R., Smith, A. and Wright, P., Toolglasses, marking menus, and hotkeys: A comparison of one and two-handed command selection techniques. *Graphics Interface 2004*, 17-24.
27. Robertson, G., Czerwinski, M., et al. Data Mountain: Using spatial memory for document management. *ACM UIST 1998*, 153-162.
28. Scarr, J., Cockburn, A., Gutwin, C., Bunt, A. Improving command selection with CommandMaps. *ACM CHI 2012*, 257-266.
29. Scarr, J., Cockburn, A., Gutwin, C. and Quinn, P. Dips and Ceilings: Understanding and Supporting Transitions to Expertise in User Interfaces. *ACM CHI 2011*, 2741-2750.
30. Schmidt, R. Frequent augmented feedback can degrade learning: Evidence and interpretations. Requin, J. ed., *Tutorials in motor neuroscience*, 1991, 59-75.
31. Schmidt, R. and Lee, T. *Motor Control and Learning: A Behavioral Emphasis*. Human Kinetics, 2011.
32. Simon, H. Satisficing. in Eatwell, J., Millgate, M. and Newman, P. eds. *The New Palgrave: A Dictionary of Economics*, Stockton Press, 1987, 243-245.
33. Weld, D., Anderson, C., Domingos, P., Etzioni, O., Gajos, K., Lau, T. and Wolfman, S. Automatically personalizing user interfaces. *IJCAI 2003*, 1613-1619.
34. Zhai, S. and Kristensson, P. Shorthand Writing on Stylus Keyboard. *ACM CHI 2003*, 97-104.

APPENDIX: WEEKLY DRAWING TASKS FOR STUDY 2