Investigating the Post-Training Persistence of Expert Interaction Techniques

BENJAMIN LAFRENIERE, Autodesk Research
CARL GUTWIN, University of Saskatchewan
ANDY COCKBURN, University of Canterbury

Expert interaction techniques enable users to greatly improve their performance; however, to realize these advantages, the user must first acquire the skill necessary to use a technique, then choose to use it over competing novice techniques. This article investigates several factors that may influence whether use of an expert technique persists when the context of use changes. Two studies examine the effect of changing performance requirements, and find that a high performance requirement imposed in a training context can effectively push users to adopt an expert technique, and that use of the technique is maintained when the requirement is subsequently reduced or removed. In a final study, performance requirement, high-level task, and environment of use are changed—participants played a training game to learn the menu for a drawing application, which they then used to complete a series of drawings over the following week. Participants exhibited a somewhat surprising "all-or-nothing" effect, using the expert technique nearly exclusively or not at all, and maintaining this behavior over a range of qualitatively different tasks. This suggests that switching to an expert technique involves a global change by the user, rather than an incremental change as suggested by previous work.

CCS Concepts: • **Human-centered computing → User studies**; **HCI theory, concepts and models**; **Graphical user interfaces**;

Additional Key Words and Phrases: Novice to expert transitions, expert interaction techniques, training

ACM Reference format:
Benjamin Lafreniere, Carl Gutwin, and Andy Cockburn. 2017. Investigating the Post-Training Persistence of Expert Interaction Techniques. *ACM Trans. Comput.-Hum. Interact.* 24, 4, Article 29 (August 2017), 46 pages.
https://doi.org/10.1145/3119928

1 INTRODUCTION

Expert interaction techniques, such as keyboard shortcuts, gestures, and command languages allow users of interactive systems to greatly improve their performance by substituting fast, memory-based interactions for slower interactions based on visual search and direct manipulation

This work is supported by the Natural Sciences and Engineering Council of Canada.
Authors' addresses: B. Lafreniere, Autodesk Research, 210 King St. East, Suite 500, Toronto, Ontario M5A 1J7, Canada; email: ben.lafreniere@autodesk.com; C. Gutwin, Department of Computer Science, University of Saskatchewan, 110 Science Place, Saskatoon, Saskatchewan S7N 5C9, Canada; email: gutwin@cs.usask.ca; A. Cockburn, Department of Computer Science and Software Engineering, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand; email: andy@cosc.canterbury.ac.nz.

https://doi.org/10.1145/3119928

(Malacria et al. 2013a; Lane et al. 2005; Odell et al. 2004). However, a key challenge for these techniques is that they require an initial investment of time and effort to acquire the memory and skill necessary to use them, and this can act as a barrier to their adoption (Lane et al. 2005; Scarr et al. 2011), particularly when the expert technique exists alongside novice techniques (i.e., the user can choose to use either the novice or the expert method to perform a given action). To address this challenge, training systems have been developed that teach expert interaction techniques, and allow users to practice them to develop the required memory and skill. Previous research has explored the use of games as a way to teach touch gestures (Cockburn et al. 2007b; Kristensson and Zhai 2007), and a number of commercial products exist to help people learn expert techniques. For example, ShortcutFoo[1] provides flashcard-style training in keyboard shortcuts and commands for a variety of applications, including Visual Studio, Excel, and Photoshop, and VIM Adventures[2] provides a game in which players navigate and complete challenges in a game world using keyboard shortcuts and commands for the VIM editor.

While a number of training systems have been proposed, little is known about how well training in expert interaction techniques actually works—that is, whether practicing expert interaction techniques in a training context will translate into increased use in a real-world setting, displacing novice techniques. Although a great deal of research has been carried out to examine general transfer of learning (i.e., the degree to which knowledge and skills learned in one context are used in another context), few studies have looked specifically at transfer when the training involves an expert technique that is intended to supersede a novice technique that the user already knows. We call this particular type of transfer *post-training persistence*—that is, whether use of a learned technique persists when the user leaves the training setting, and has the option of using alternative techniques.

Improving the persistence of expert interaction techniques would lead to increased use of these superior methods, improving the overall level of performance with interactive systems. However, the post-training persistence of expert interaction techniques has not been widely studied, and we know little about the factors that influence whether or not a user will maintain use of an expert interaction technique when the context of use changes. Investigating these questions is the main focus of this article.

It may seem odd that a user who has learned a faster expert technique would ever revert to using a slower novice method. However, after reviewing the literature on learning transfer, we identified several factors that may prevent an expert technique from persisting. These include differences in *performance requirement* between the training task and real use; *technique-task compatibility* (the degree to which tasks in the usage context are able to exploit the expert technique); and *risk aversion* (the willingness of the user to make errors in the training versus the usage context).

Performance requirement. Tasks used in training may have different performance requirements than the tasks carried out in everyday use. For example, to encourage practice of the expert technique, a training system may use a demanding task that necessitates the use of the faster expert method. If this level of performance is not required in the real-world usage setting, the user may feel no pressure to continue using the expert technique. For instance, in a previous work by Gutwin et al., participants quickly adopted an expert selection method while playing a training game, but another set of participants showed little uptake of the same expert method over 10 weeks of using a drawing application, which may suggest that a high performance requirement is a key driver of use of the expert method (Gutwin et al. 2015).

[1] https://www.shortcutfoo.com.
[2] http://www.vim-adventures.com.

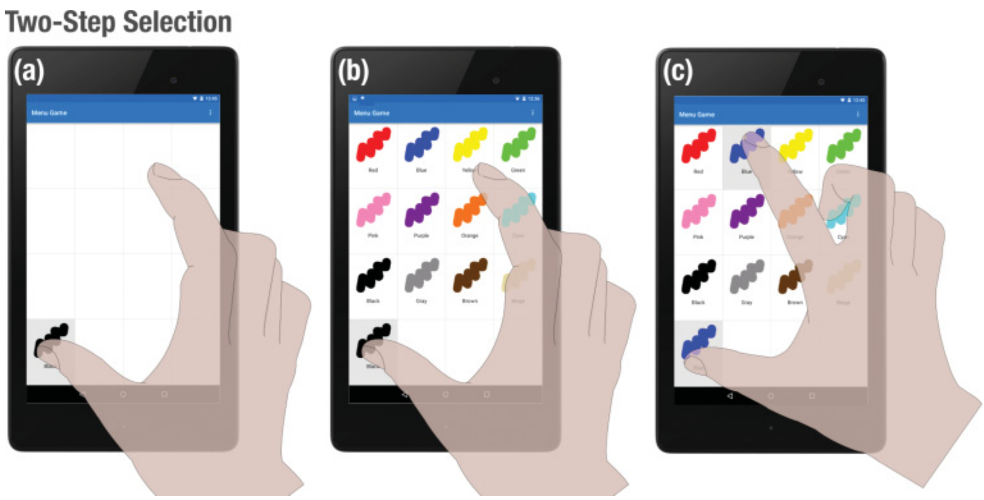

Fig. 1. Two-step selection in a FastTap menu. The user touches the menu button (a), after a short delay the menu items are displayed (b), and the user selects the desired item (c).

Technique-task compatibility. Second, the training context may be particularly well-suited to performing the expert technique, while the usage context may not. Previous work on touch-typing has demonstrated that training using non-words or random sequences of letters gives a relative advantage to touch-typing—possibly because this nonsense text is difficult to remember, forcing people to spend more time looking at the screen and less time looking at their hands and the keyboard. This artificial characteristic of the training environment can improve adoption of the expert technique, but in real-world usage with text that is easily held in the user's memory, it becomes easier to revert to "hunt-and-peck" typing (Yechiam et al. 2003).

Risk aversion. Third, differences in the consequences of errors and a user's willingness to make mistakes may affect the persistence of an expert technique. For example, a user may be willing to guess at memory-based shortcuts in a training game, but not during a work task.

It is important to establish the conditions under which use of expert interaction techniques persists, because this information can guide the design of training systems. In particular, there are advantages to training in a different context than that in which a technique will be used. For example, a tutorial can use representative tasks that allow the learner to focus on understanding the technique, and a training game can employ techniques to promote rapid learning (such as optimizing rehearsal intervals (Zhai et al. 2002; Zhai and Kristensson 2003), increasing mental effort (Cockburn et al. 2007a), or encouraging use of an expert technique by imposing a high performance requirement (Gutwin et al. 2015)). Moreover, these goals can be achieved without the risk of the user making errors on an important document. Finally, the alternative—that people will simply learn the expert method through continued use of the software—has been shown to be generally less successful than dedicated training (Gutwin et al. 2015).

To investigate some of the factors affecting the post-training persistence of expert techniques, we carried out three studies using FastTap menus (Gutwin et al. 2014), a type of spatial menu system that supports both a novice method and an expert selection method. In the novice (two-step) selection method, the user first touches a menu button with their thumb (Figure 1(a)). After a short delay (~250ms) the items in the menu are displayed (Figure 1(b)), and the user can then select the desired item with another finger (Figure 1(c)). In the expert (one-step) selection method,

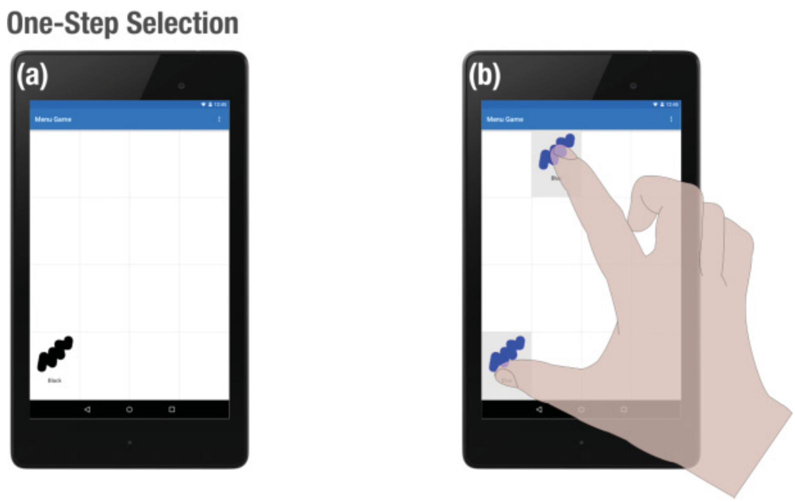

Fig. 2. One-step selection in a FastTap menu. The user taps the menu button and the location of the desired item together (b), without waiting for the menu to be displayed.

the user taps the menu button and the location of the desired item together (Figure 2(b)), without waiting for the menu to appear.

FastTap menus are an ideal technique for studying post-training persistence because their two selection modes epitomize the difference between novice and expert interaction techniques—two-step selection is a closed-loop action involving visual search, whereas one-step selection is a significantly faster open-loop action based on memory of hand postures and spatial locations.

In each of our three studies, we first trained people on both the novice and expert methods of selecting commands in an application, and then changed aspects of the context to see whether the change would affect use of the expert method. The first two studies examined the effect of changing the *performance requirement* of the user's task—that is, the degree to which people need to use the faster expert method in order to successfully complete their task. We created a game in which the player had to select commands (see Figure 3 in Section 4), and as the game progressed they had to make their selections more and more quickly. This high performance requirement pushed people to switch to the expert technique (to take advantage of its greater performance). Once they had learned the expert method, we reduced the performance requirement (in Study 1) or removed it altogether (in Study 2), to see whether people would continue using the expert technique.

The main result from these studies was that reducing or removing the performance requirement did *not* reduce the use of the expert technique—that is, use of the expert selection method was maintained even after the need to make fast selections was removed. Even when we reduced the speed of the game to zero—meaning that participants could take as much time as they wanted to make a selection—more than 97% of selections were still made using the expert technique. This is an important result because performance requirement is a factor that can be easily manipulated in a training context, but is largely out of the control of an application designer in a usage context.

The third study tested a more drastic change in context. Specifically, we changed the *performance requirement, high-level task*, and the *environment of use*. As in the previous studies, we trained participants using a game, but this time they learned a menu of drawing tools taken from a touch-based drawing application that we had also developed. We then installed the drawing application on participants' personal smartphones, and sent them a series of drawing tasks to complete over the next 7 days. The drawings were done in the participants' everyday environments, with minimal experimental control imposed on how the tasks were completed. We logged all use of the interface in order to determine whether participants continued using the expert technique—even in a very different usage setting, with a different application and completely different tasks, and no intrinsic requirement that participants make fast selections.

Given the number of changes made between the training and usage environments, and the limited overall success of transfer that is shown in previous work (e.g. (Gist et al. 1990; Anderson and Bischof 2013)), we did not expect that the expert selection technique would completely persist. We expected that people would only partially maintain their use of the expert technique—for example, for commands that were frequently used—and that we might observe decay in use of the techniques over the week of tasks. However, our findings were counter to these expectations. For five of the ten participants, the expert technique was used nearly 100% of the time for the drawing tasks, with essentially no reduction over the 7 days, and across a range of qualitatively different types of drawing tasks (novel, repeated, requiring creativity, etc.). For another four of the ten participants, the expert technique did not persist at all—they returned to the novice technique, and used two-step selection almost exclusively for the drawing tasks. Only one participant used a mixture of the novice and expert methods, and a post-study interview revealed that this was the result of a deliberate choice on his part after thinking about the relative advantages of the two selection methods.

This "all or nothing persistence" phenomenon has not been observed before, and suggests that some expert techniques require a global change on the part of the user, rather than being gradual (for example, as suggested by Kurtenbach's rehearsal hypothesis (1993)). We discuss several potential explanations for these results, including simple individual differences in users' interest in using expert methods (the fact that someone has learned an expert method does not mean that they have an interest in using it), and the possibility that users are more affected by potential negatives of a new technique than they are by the potential positives (i.e., a negativity bias (Quinn 2016; Kahneman 2013)).

Overall, our results provide valuable new insights into the post-training persistence of expert techniques, and demonstrate that there is considerable complexity in this phenomenon. It is clear that more work needs to be carried out to confirm and expand on these studies, but our work nevertheless provides several important contributions. First, we identify post-training persistence as a unique issue in systems with competing novice and expert techniques, and identify key contextual factors that could impact persistence when moving from a training context to a usage context. Second, we show that changing the performance requirement alone does not appear to reduce the use of an expert method (when the application and environment are held constant). Third, we show that games can be an effective training context for memory-based expert techniques—in less than 20 minutes of gameplay, most participants learned the expert technique for a 12-item spatial menu. Finally, we show that expert technique use learned in a game can persist across large contextual changes (to the task, application, environment, and performance requirement)—although more work is needed to determine why transfer occurred for some participants but not for others.

2 BACKGROUND AND RELATED WORK

We begin with a review of related work on expert interaction techniques, approaches that have been explored to support the transition from novice to expert performance in interactive systems, and general work on training and skill development. We then present the concept of post-training persistence, contextualized within this body of existing work.

2.1 Expert Interaction Techniques

User interfaces often provide more than one mechanism for performing a given action. A common scheme is to provide both *novice interaction techniques* that can be easily performed with little or no training (e.g., hierarchical menus), and *expert interaction techniques* that require up-front training to perform, but which enable higher performance by taking advantage of the richness and power of human perception, memory, cognition, and motor actions (e.g., keyboard shortcuts).[3]

A common characteristic of novice interaction techniques is that they are operated in a closed-loop fashion, where sensory information and feedback are used to adjust actions as they are carried out. For example, in a hierarchical menu, the user will perform a visual search for the desired item. Even if the user has an idea of the position of the item, the visual feedback provided by the menu will be used during target acquisition, and selection time will be some function of the number of items present (Cockburn et al. 2007a). The visual guidance provided by these techniques makes them easy to learn and use, and likely explains their near ubiquity. However, these benefits come at a cost in terms of performance—the richness and power of human perception, cognition, and motor action are constrained by the closed-loop nature of these techniques.

In contrast, expert interaction techniques are often memory-based and include open-loop actions. For example, once a user has learned that Ctrl + C is the keyboard shortcut for Copy, and has practiced executing this shortcut on the keyboard sufficiently, the command can be performed rapidly and without attending to visual feedback. This enables expert techniques to achieve significantly higher performance as compared to their novice counterparts (Malacria et al. 2013a; Lane et al. 2005; Odell et al. 2004; Card et al. 1983).

A range of expert interaction techniques have been developed and studied. Keyboard shortcuts are a consistent feature of GUIs and their performance benefits are well established (Lane et al. 2005; Card et al. 1983). Gestures and marks, performed with pen, mouse, or touch input, have also been explored as rapid command invocation techniques (Kurtenbach and Buxton 1991; Appert and Zhai 2009), and some evidence suggests that gestures can be easier to learn than keyboard shortcuts (Appert and Zhai 2009). Command languages, most commonly used in command-line interfaces and quick application launchers (e.g., Quicksilver,[4] Apple Spotlight), have also been explored as additions to GUI applications (Hendy et al. 2010; Scarr et al. 2011). More recently, spatially stable menus have been investigated as a way to enable faster interaction with interactive systems (Scarr et al. 2012), and as an expert interaction technique for multi-touch devices (Gutwin et al. 2014, 2015).

Though a range of different expert interaction techniques exist, and the learning characteristics of some of them have been studied, there has been little work investigating the transfer of these techniques from a training context to a usage context (i.e., their post-training persistence). In this work, we examine these issues through spatially stable menu systems such as FastTap. While these menus are not as common as keyboard shortcuts, they have several advantages for a first examination of the issues surrounding post-training persistence, as discussed in Section 1. In addition, because this type of menu is not widespread, participants in our studies are less likely to have pre-conceptions on the merits of particular techniques (in contrast, there are many people who are strong advocates of command-line tools or keyboard shortcuts).

[3]Note that "novice" and "expert" here do not refer to the general level of experience possessed by the user—a novice user of an application may still use keyboard shortcuts, while an expert user may still use hierarchical menus. Instead, these terms refer to the fact that expert techniques require some expertise to be used effectively, while novice techniques do not.
[4]http://qsapp.com/.

2.1.1 Barriers to Adoption. It is well established that users tend to plateau at mediocre performance with interactive systems (Carroll and Rosson 1987; Fu and Gray 2004; Gray et al. 2006; Bhavnani and John 2000; Lane et al. 2005; Gray and Lindstedt 2016), and a number of the proposed explanations for this phenomenon provide insights into barriers to adoption of expert interaction techniques as well.

The "paradox of the active user," a phrase coined by Carroll and Rosson (1987), refers to a hesitance among users to take time away from a current task to learn new or improved ways of doing things, even if time could be saved through an initial period spent learning. This suggests that one barrier to adoption of expert techniques is the initial learning they require. Fu and Gray (2004) built on Carroll and Rosson's work, and found that people prefer to maintain procedures that are well-practiced, generally applicable, and that provide fast, incremental feedback. Since expert interaction techniques achieve high performance by reducing or removing the need to attend to feedback, this preference by users may also explain the hesitance to use these techniques. Conversely, this preference may also explain why the expert techniques that *are* frequently used tend to be those that are generally applicable (e.g., Ctrl + C, Ctrl + X, and Ctrl + V for Copy, Cut, and Paste, respectively, which are supported in many applications).

Scarr et al. (2011) examined issues of transitioning to expertise in user interfaces, and developed a framework that characterizes expertise development in terms of *intramodal* development with a particular interaction technique, and *intermodal* development (i.e., switching from one technique to another). Critically, this framework suggests that people will suffer a performance dip when switching to a new modality, even if that modality offers an ultimately higher performance ceiling, due to the time spent familiarizing themselves with the new technique. This dip is important, because it may deter switching to the expert technique because continuing to use the novice technique will be faster for any one action (a phenomenon referred to as "local optimality" (Gray et al. 2006)), and the dip may cause the user to form a negative impression of the expert technique.

Recent work suggests that a barrier to expert adoption exists even for rehearsal-based techniques where the novice and expert interaction techniques share the same physical actions (e.g., Marking Menus and FastTap menus). Though rehearsal-based interaction can take advantage of incidental learning to ease transitions from a novice to an expert technique, work by Gutwin et al. (2015) suggests that there is still a barrier at the switching point.

Complementary to the work just discussed, Gray and Lindstedt (2016) have proposed a framework for understanding how individuals discover and invent new methods of skilled performance, by examining dips, plateaus, and leaps in their performance over time. They distinguish between performance "plateaus," in which an individual's performance is stable, but could be improved with better methods, and performance "asymptote," in which the current method is optimal for the given task in the given environment. They also suggest that performance dips can be used as a behavioral signal of periods of experimentation, discovery, and trial and error.

Taken together, the previous work suggests that training tools are important for encouraging the adoption of expert techniques. If training can be designed such that an expert technique does exhibit post-training persistence, then training can be performed separate from the user's primary task, sidestepping the paradox of the active user and minimize or negating the performance dip when switching between techniques. In this article, we establish a theoretical and empirical grounding for this type of dedicated training, by investigating the post-training persistence of expert interaction techniques.

2.1.2 Techniques to Support Adoption. Existing research on supporting the adoption of expert interaction techniques has primarily focused on in situ training techniques that operate within the usage context. These techniques variously promote awareness of expert methods, motivate their use, or try to create a more natural path from novice to expert usage.

Grossman et al. (2007) proposed a variety of schemes to assist learning of keyboard shortcuts within an application, including using visual and audio feedback when a command is invoked to expose the user to its shortcut; deterring use of the GUI by making the system unresponsive for 2 seconds after each menu selection; and forcing shortcut use after each menu selection. An experiment showed that forced use and audio feedback were effective at encouraging shortcut use, and that subjective responses were not adversely affected. Along similar lines, HotKeyCoach (Krisler and Alterman 2008) displays a dialog box when an item is selected with the mouse, requiring either a click to dismiss the dialog, or for the user to perform the keyboard shortcut before proceeding, which was shown to be effective in increasing keyboard shortcut knowledge. Though experiments have shown that these methods are successful, forced use and audio feedback may be unacceptable to users in real-world usage situations.

Subtler techniques have also been proposed to promote learning of keyboard shortcuts and command languages in an application. The Blur system (Scarr et al. 2011) uses "calm notifications"—transparent windows that are temporarily displayed on the corner of the screen—to display command line alternatives to actions performed in the GUI. Commercial products KeyRocket[5] and EVE[6] provide similar functionality for keyboard shortcuts, displaying a notification with the shortcut for each action the user performs.

To encourage users to improve their efficiency, Malacria et al. (2013b) proposed Skillometers—a dedicated widget that displays information about the user's performance during interaction, as well as an estimate of time savings that could be achieved through use of expert techniques. A comparative evaluation on a simple task suggests that this approach can increase shortcut use.

Finally, expert interaction techniques have been developed with novice-to-expert transitions in mind. Kurtenbach developed Marking Menus based on three design principles: self-revelation, guidance, and rehearsal (Kurtenbach 1993; Kurtenbach and Buxton 1994). In these menus, the physical actions for making an expert selection are a sped-up version of the visually guided actions of making a novice selection. The intention is to create a more natural transition from the novice to the expert selection method, because novice selection is a rehearsal of expert selection. More recently, FastTap menus have been developed as a touch-screen interaction technique based on the same design principles (Gutwin et al. 2014). While these techniques have been shown to ease learning of expert techniques, recent work has suggested that there remains a barrier at the switching point between the novice and expert methods (Gutwin et al. 2015). This suggests that some level of intentional learning and practice is required even for interaction techniques designed to promote a natural transition to expert use.

Though a range of in-situ training techniques have been explored, we are unaware of work that has examined the persistence of expert techniques after training is finished. This is important because a number of the in-situ training techniques that have been explored may not be acceptable to users in real-world usage scenarios (e.g., audio feedback, UI delays, or forced use of shortcuts). As well, work in the psychology literature has shown that guidance, in the form of feedback provided during an action, can become relied upon, degrading retention and performance when the guidance is removed (Schmidt 1991). In this article, we conduct an investigation into the factors that influence whether use of a technique persists once guidance is removed, and our findings have the potential to inform the design of in-situ training techniques.

[5]https://www.veodin.com/keyrocket/.
[6]http://www.hotkey-eve.com/.

More generally, there has been little research on dedicated (i.e., outside the application) training for expert interaction techniques. In the commercial space, ShortcutFoo[7] provides dedicated training in keyboard shortcuts and commands in a wide range of applications, including Visual Studio, Excel, and Photoshop; and VIM Adventures[8] provides training for keyboard shortcuts and commands for the VIM editor. These systems variously use games, competition, challenges, and interval training to help users learn expert interaction techniques. However, the effectiveness of these approaches has not been well established. In this article, we investigate specific factors that can affect whether dedicated training on expert interaction techniques will transfer to use of those techniques in subsequent usage contexts. As a basis for our exploration, the next section reviews the existing body of work on training and skill development from the psychology and motor learning literature.

2.2 Training, Transfer, and Retention

A great deal of research has been carried out over many decades on learning and training (e.g., see Schwartz et al. (2001) for a general review). In situations where training occurs outside of the everyday use of learned skills and knowledge (for example, classroom training to teach new skills for a job), the value of a training scheme involves more than just the amount of learning that occurs during training itself. As a result, researchers have also considered the concepts of *transfer* and *retention* as important parts of the overall success of training. Transfer is the degree to which trainees apply the knowledge or skill acquired in a training context to another setting (such as a real-world job) (Baldwin and Ford 1988), and retention is the length of time that knowledge or skills are remembered or used after training has ended (Wang 2009). Our particular concern in this article is with transfer, although issues in both training design and retention are relevant at several points.

Many researchers have observed that transfer is a difficult problem, and that typically only a low percentage of knowledge and skills is ever translated back to the transfer context (i.e., the setting where the skills and knowledge are to be applied) (Gist et al. 1990). As stated by Ford and Weissbein (1997), “the ‘transfer problem’ in organizational training [is] that much of what is trained fails to be applied in the work setting” (p. 22). Researchers in motor learning have also noted that achieving good performance in the training environment is not enough to guarantee transfer—as stated by Anderson and Bischof (2013), “While it is commonly believed that increased performance on practice trials is indicative of learning, [this is] only evident through retention and transfer tests” (p. 1110).

There are several factors that can affect the success and failure of transfer, and these can be generally grouped into four categories: the design of the training program, the characteristics of the transfer setting, the characteristics of the trainee, and the characteristics of the learned skill. Here, we review several factors in these categories that are specifically relevant to the transfer of expert interaction techniques.

2.2.1 Design of the Training Program. There are many different ways of training people in a new skill; although a full review of training approaches is outside the scope of our work (see Compeau et al. (1995), Schmidt and Lee (2011), or Salas and Cannon-Bowers (2001) for surveys), there are several aspects of the way a training program is designed that have been shown to affect both transfer and retention.

[7] https://www.shortcutfoo.com
[8] http://www.vim-adventures.com

Similarity of training and transfer settings. Several prior results in psychology suggest that retrieval is easier when the transfer setting is similar to the conditions under which the skill or knowledge was first learned (e.g., Tulving and Thompson's (1973) encoding specificity principle, and Woodworth and Thorndike's (1901) identical-elements theory). Researchers have investigated how similarity of cues and stimuli affect retrieval (e.g, (Arthur et al. 1998; Baldwin and Ford 1988)), and the effect of different procedures in which those cues appear (Machin 2002). We build on this past work, identifying specific differences between training and usage contexts that may affect the post-training persistence of expert interaction techniques (Section 3.1), and then test several of these in three studies.

Stimulus variability and underlying principles. If the transfer setting provides a variety of situations in which the new skill is applicable, transfer can be improved if training uses a range of stimuli representative of these situations (Gist et al. 1990). Similarly, if there is a general rule that governs the use of the skill or knowledge, it can be valuable to focus training on building an understanding of these underlying principles (Rohrer et al. 2005).

Conditions of practice. Several studies have looked at the way in which the organization of the training sessions affects learning rate, retention, and transfer. Researchers have considered the distribution of practice (e.g., massed versus grouped practice trials, and the amount of time between groups (Cepeda et al. 2006)), whether training involves whole or part concepts (Adams 1987), and the type of feedback given (e.g., knowledge of performance versus knowledge of results (Gentile 1972)).

Part versus whole training. Research from the 1950s and 1960s investigated whether a complex task could be broken down into parts that could be learned more efficiently than the task as a whole, motivated by the desire to efficiently train people in a relatively difficult body of material (e.g., training pilots to fly airplanes and execute missions). In general, researchers found that while part-skill training does effectively transfer to performance of a whole task, whole-task training is typically better, and when part skills must be time-shared with other activities in a final task, some additional integrative whole-task practice is required (see overview in Adams (1987)).

Guidance. Providing guidance during training supports a learner (cognitively or physically) in the execution of a skill (e.g., guiding the learner's motions as they practice a golf swing). Guidance can improve learning, but may also impair transfer: the *guidance hypothesis* (Schmidt 1991) suggests that augmented feedback that improves early performance may impair performance once the guidance is removed.

Effort and deliberation. Studies of memory have shown that retention can be improved if learners are deliberate in their practice (Ericsson et al. 1993), and if they need to work harder during training. For example, Craik and Lockhart's (1972) levels of processing framework suggest that the strength of a memory is a function of the depth to which the stimulus is analyzed. In human-computer interaction (HCI) research, this idea has been explored by Cockburn et al. (2007b), who showed that spatial memory of an onscreen keyboard was improved when users could not see the letters on the keys.

Overlearning. When a training program uses overlearning, it asks learners to continue practicing even after they have demonstrated initial mastery of a skill (e.g., past their first error-free trial) (Ebbinghaus 1913; Driskell et al. 1992). The intention of overlearning is to develop automaticity of the skill, and studies show that overlearning can substantially improve retention (Driskell et al. 1992) (although after longer time periods the positive effects of overlearning may be reduced (Rohrer et al. 2005)).

2.2.2 Factors in the Transfer Setting. Prior research has also identified several characteristics of the transfer setting that can substantially affect transfer (e.g, (Baldwin and Ford 1988; Ford and Weissbein 1997; Vosburg 2000)).

Need for the new skill. If a learned skill is not needed for the successful completion of a job or task, it is less likely to be transferred. Although most previous research assumes that training provides valuable content, several researchers indicate the importance of a needs assessment at the start of the training process (e.g, (Rouiller and Goldstein 1993)). This characteristic is similar to our idea of "performance requirement" as described in Section 3.

Usage opportunity. Similar to the effects of need as described earlier, the number of opportunities to use the learned skill can affect transfer. Researchers have found that increased opportunities to apply learned skills in the work environment imply improved likelihood of retrieving those skills (Baldwin and Ford 1988; Ford et al. 1992).

Time between training and use. Although time is typically used as a factor in retention studies, it can also affect transfer. The general effect is the same in both cases—longer times between training and use lead to reduced retention, and thus reduced transfer (e.g, (Ebbinghaus 1913; Driskell et al. 1992)), due to increased possibilities of memory interference and simple forgetting (Wang 2010).

Support from organizations and people. Several studies have shown the importance of a supportive work environment for transfer of training—for example, researchers have considered the effects of supervisors' attitudes toward the training (Wexley and Latham 2001), how a non-supportive organizational climate affects transfer (Broad and Newstrom 1992), and budgetary and scheduling support (Peters and Waterman 1982).

Specific transfer support. A variety of post-training strategies for enhancing retention and transfer have been investigated, including programs to set transfer goals, monitoring of skill performance, and specific "relapse-prevention" techniques (Wexley and Baldwin 1986; Foxon 1993; Machin 2002).

2.2.3 Characteristics of the Trainee. Individual differences can also have a substantial effect on transfer, although these can be much more difficult for a training system to assess or control. Baldwin and Ford (1988) state that "trainee characteristics consist of ability or skill, motivation, and personality factors" (p. 64). Of these, motivation appears to be the strongest factor affecting transfer—for example, desire to succeed and belief in the value of the training have been shown to lead to improved transfer of skills to job settings (Tubiana and Ben-Shakhar 1982; Baumgartel et al. 1984). Researchers have also studied techniques for increasing motivation, such as gamification (McGonigal 2011; Landers and Armstrong 2017) or displays that show users how much better they could perform with new skills (Bateman et al. 2012; Malacria Scarr et al. 2013).

A trainee's post-training motivation can also affect transfer. In particular, choosing to use strategies such as goal-setting or self-management once back in the work setting can lead to improved transfer negotiation and supervision of skills (Wexley and Baldwin 1986; Gist et al. 1990).

2.2.4 Characteristics of the Learned Skill. There are many different kinds of skill that can be taught by a training program—for example, typical training programs involve cognitive skills (e.g., memorization), physical skills (e.g., sports or psychomotor movements), and interpersonal skills (e.g., negotiation). Although researchers have considered issues of transfer for all of these skill types, a few studies have looked specifically at whether different skills transfer differently (Wang 2010). In addition, it is difficult to consider how certain factors could be controlled for some types of tasks (e.g., overlearning is much easier to implement for a memory skill than for an interpersonal skill) (Baldwin and Ford 1988).

One specific task characteristic that has been tested is the complexity of the learned skill. A meta-analysis by Wang (2010) showed better transfer for tasks of higher complexity, possibly because the added complexity leads to deeper processing. An exception was found in situations where the higher complexity involved multiple ways to carry out the task. This added variability actually led to reduced transfer.

2.2.5 Skills Transfer in Human-Computer Interaction. In addition to the work reviewed earlier, the HCI literature contains a number of studies on skills transfer for specific input techniques—see Cockburn et al. (2014) for a recent review of interaction techniques designed to support novice to expert transitions. For example, touch typing skill with QWERTY-layout keyboards has been shown to transfer to a number of variant keyboard designs (Green et al. 2004; Isokoski and Raisamo 2000; Matias et al. 1996), and gesture skills learned with one hand have been shown to transfer to the other hand (Annett and Bischof 2013). Finally, reuse of gestures and chunking has been recommended as a means of reducing cognitive load and improving learnability in gesture-based systems (Buxton 1995; Shen et al. 2006; Wu et al. 2006).

There has been relatively little research on skill transfer for expert interaction techniques specifically, with the exception of some work on touch typing. Touch typing can be considered to be a memory-based expert interaction technique that exists alongside an alternative novice technique (i.e., visually guided or "hunt and peck" typing). Yechiam et al. (2003) studied the phenomenon of people undergoing training in how to touch type, and then reverting to visually guided (novice) typing in everyday use. They theorized that this failure of transfer was due to a particular difference between the training and usage contexts, combined with a tendency to allocate behavior based on what produces better immediate performance (a phenomenon called "melioration" (Herrnstein et al. 1993)). Specifically, touch typing training is often based on typing non-words or random sequences of letters without a semantic structure, which makes visually guided typing more difficult by forcing the user to frequently alternate between looking at the monitor and the keyboard, and also makes it more difficult to reacquire a position in the text. Thus, visually guided (novice) typing is at a disadvantage during training, but advantaged during everyday use. The authors argued that because of this difference, touch typing trainees can reach a point where touch typing is more efficient than visually guided typing during training, but less efficient during everyday use (in the terminology of this article, the expert technique does not persist, because of the differences between the training and usage contexts).

2.2.6 Models and Frameworks of Skill Acquisition. Several researchers have also sought to identify and formalize the principles that underlie learning and skill acquisition, and these frameworks can have explanatory power for the issues of expertise and transfer that we consider here. In the following text, we present an overview of several frameworks.

Phases of skill acquisition. Fitts and Posner (1967) stated that learning progresses in three distinct stages. The "cognitive" stage is the earliest, and is characterized by the initial development of strategies and understanding of the task. In the "associative" stage, learners refine their performance based on feedback, and build associations between specific stimuli and appropriate responses. The associative phase is where people move from focusing on what is done in the task, to how it is done, and can last for a long time. Finally, the "autonomous" stage is characterized by task elements becoming automatic and less subject to cognitive control. Although Fitts and Posner did not explicitly discuss transfer, it is reasonable to expect that transfer to new contexts becomes more likely at later stages (i.e., in the advanced associative stage or autonomous stage).

Strategy selection and microstrategies. Another early view of skill acquisition suggests that practice and time allow people to make better decisions about which of the many available strategies should be used to complete a task (Crossman 1959). This view focuses on the selection of strategies rather than on strategy refinement. A more recent version of this theory that is particularly applicable to interactive systems involves "microstrategies." In this work, Gray and Boehm-Davis [2000] observe that there are often several ways that a unit of interaction can be carried out through simple mouse and keyboard actions, and go on to show that people recognize small time differences between different strategies (as small as 150ms), and choose more efficient strategies when tasks are stable. The idea of strategy selection suggests that users' choices about how to complete a task are not simply based on the frequency of different past experiences, but that people are also sensitive to the efficiency of different methods.

Cognitive architectures. Several theories of learning and skill development are based on the idea that cognition can be modeled by production rules; two of the best-developed frameworks are the ACT-R architecture (e.g, (Anderson 1983)) and the SOAR framework (e.g. (Newell 1990)). These models have been used to describe and explain many elements of learning, with considerable success. Transfer of training has been explored using ACT-R (Singley and Anderson 1989)—researchers suggest that the similarity between the production rules in the two different contexts determines the degree of transfer between those contexts (echoing Woodworth and Thorndike's early identical-elements theory (1901)).

The idea that increasing practice strengthens certain production rules can be used to explain several aspects of learning (including the Fitts and Posner's three stages of skill acquisition). However, some research has shown that real users do not behave in ways that are predicted by these models. One set of studies by Fu and Gray (2004) is particularly relevant to our work—in this work, Fu and Gray observe that plateaus in real-world performance (i.e., the paradox of the active user) can be explained using cognitive-science frameworks. As described earlier, this work proposes that users are not always focused on optimization, but instead sometimes follow patterns of behavior that are "stable but suboptimal" because of local data (e.g., interactive feedback that reduces mental workload).

There are thus several theories and frameworks that could be used to explain different elements of skill transfer when a user learns a more-efficient technique for an interactive task such as command selection. The different theories, however, can present competing outcomes—for example, the work on microstrategies suggests that people are highly sensitive to improving efficiency, whereas Fu and Gray's work suggests that people may fail to adopt more-efficient techniques even when they are known. Later in the article, we return to the issue of how the different theories can explain the results of our studies, but first we build on the existing body of work on skill training, transfer, and retention, to look more closely at the issue of post-training persistence. In the next section, we present a detailed definition of this concept, and identify key differences between training and usage contexts for expert interaction techniques that we believe will affect their post-training persistence.

3 THE POST-TRAINING PERSISTENCE OF EXPERT INTERACTION TECHNIQUES

We define the *post-training persistence* of an expert interaction technique as the degree to which, once acquired in a training context, the technique transfers to a usage context and is used in preference to novice interaction techniques capable of performing equivalent actions. As a special case of skills transfer, many of the factors identified earlier in past work are likely to impact post-training persistence as well (e.g., design of the training program, training and transfer settings, characteristics of the trainee, etc.) While all of these are potentially interesting areas for study, in this article, we focus on characteristics of the training and transfer contexts, and differences between these contexts. This decision was pragmatic—we attempted to identify factors that would commonly differ between training and usage contexts for expert interaction techniques, and particularly those that could be manipulated easily in training environments.

In the rest of this section, we discuss specific differences between training and usage contexts that may influence the persistence of expert interaction techniques, and in the sections that follow, we describe the studies we conducted to test some of these factors experimentally.

3.1 Training and Usage Contexts

By considering past work on transfer, and a range of different potential training contexts for expert interaction techniques (e.g., tutorials, flash cards, classroom instruction, and training games), we identified seven key differences between training and usage contexts that are likely to affect the post-training persistence of an expert technique.

3.1.1 Retrieval Triggers. To successfully perform an expert technique in a usage context, the user must first think to use the technique, and then retrieve from memory the information required to perform the technique (e.g., the specific keyboard shortcut for a command). Differences between the training and usage contexts may make this difficult by removing triggers that helped with retrieval in the training context. For example, consider a user that has learned keyboard shortcuts for Microsoft Word using a flash-card game. During training, the user may be able to effectively indicate Ctrl + B when presented with the word "Bold" written on a card, but it is not clear how directly this will translate into use of the shortcut during regular use, where the "Bold" stimulus is absent, replaced by an intention to make a current selection of text bold. This may prevent the user from thinking to use the expert technique, or may make it more difficult to retrieve the relevant shortcut.

3.1.2 Execution Actions. Once the user has retrieved the necessary information to use the expert technique, this information may need to be translated in some way before the technique can be executed. For example, a user who has been verbally quizzed on the keyboard shortcuts for Photoshop may be able to confidently state that "Ctrl + D" is the shortcut for the Deselect command, but it may take extra effort for the user to translate this into the muscle movements to perform the keyboard shortcut. As a more extreme example, memorizing the basic chord chart for a guitar is not the same as being able to play the actual chords. This process of translation knowledge about how to perform a new method into the cognitive, perceptual, and motor operations that constitute that method—referred to as "method development" in past work (Gray and Lindstedt 2016)—may hurt the persistence of a technique by imposing an additional cognitive cost to executing the necessary actions in the usage context.

This difference between training and usage contexts is most relevant for systems that do not give the learner a chance to practice the execution actions for an expert interaction technique (e.g., learning keyboard shortcuts through a "Top Photoshop Keyboard Shortcuts" webpage).

3.1.3 Performance Requirement. A third potential difference between training and usage contexts is the performance requirement that each imposes on the user. In a training context, a high performance requirement can be easily imposed with a timer, or a fast-paced game, and these approaches have been shown to be effective in motivating users to quickly begin using an expert interaction technique (Gutwin et al. 2015). In contrast, in a usage context, the performance requirement is largely out of the control of the application designers, and depends much more on the wider context in which an application is used (e.g., if the user has a major deadline, or is casually typing up a grocery list). Pragmatically, we would like an expert technique adopted in a training context with a high performance requirement to persist when the performance requirement is removed in a real-world usage context.

The effect of the performance requirement placed on the user is also important to study because it is a particularly powerful technique—imposing a high performance requirement can all but force the user to make use of the expert technique by rendering the novice technique insufficient to meet the goals set in the training context (i.e., it can act as an implicit requirement that the user use a particular technique). Given this power, it is particularly interesting to ask what happens when this requirement is removed—will use of the expert technique established in a setting with a high performance requirement be maintained when the performance requirement drops below the threshold where a novice technique could be used in its place?

With the aforementioned issues as motivation, we conducted two studies to test post-training persistence in the face of changing performance requirements, and showed that participants *do* maintain use of the expert technique, even when all need for performance is removed. Confirming this transfer effect then set the stage for our third study, where we tested the effects of high-level task factors (discussed later) in addition to a change in performance requirement.

3.1.4 High-Level Task. A particularly interesting question is whether differences in the high-level tasks performed by the user in the training versus usage contexts will affect post-training persistence, potentially by changing retrieval triggers or the effort required to execute actions. This is important to understand because to some extent it is unavoidable for the high-level task to be different in a training context versus a usage context—in the former the user is training, while in the latter they are working on their real-world tasks. More generally, there is the question of whether working on unfamiliar tasks, repeated tasks, or tasks requiring creativity will affect persistence.

There are a number of arguments for why differences in high-level tasks might inhibit post-training persistence. First, the high-level task the user is performing may be an important retrieval trigger. For example, without the prompts and cues provided by the training system, the user may not think to use the expert technique. As well, during everyday usage, the user's cognitive resources may be consumed with additional considerations, such as planning the next action to be performed in an interface, devising a strategy for creating a particular effect, or creative thinking about what to do next, and as a result retrieval triggers may be overlooked. These additional cognitive costs could also hurt the user's ability to translate what they have learned (e.g., the association of the word "Copy" in a training game with "Ctrl + C") into actions for executing the expert technique (performing the keyboard shortcut).

The high-level task being performed during real-world usage could potentially affect the choice of interaction technique in other ways as well. During a training game for learning keyboard shortcuts, a user may be willing to guess at keyboard shortcuts for which they are not completely confident, because there are a few negative consequences for an incorrect guess. In contrast, guessing at a keyboard shortcut while working on an important document may not be deemed acceptable, because it could cause delays or a loss of important work.

It is important to understand how differences in high-level tasks affect persistence, because for training it is desirable to give the user a "safe" environment where they can try out a technique and practice it without fear of losing important work. In our third study, we test whether different types of higher-level tasks (e.g., repetitive, unfamiliar, and creative) affect the persistence of expert selections using FastTap menus.

3.1.5 User Interface and Presence of Novice Techniques. Differences between the interfaces presented to the user in the training and usage contexts may also affect post-training persistence. For example, differences in the interface could alter or remove retrieval triggers that were present in the training context, preventing the user from thinking to use a technique. The interface in the usage context could also contain additional methods of performing a given action that are not present in the training context. For example, a training system for keyboard shortcuts is unlikely to include a hierarchical menu alternative, which will be available in the usage context. This could act as a barrier to persistence because the user will not be trained to choose the expert technique in the presence of a novice alternative, and may simply revert to using the familiar novice technique once they are in the usage context.

Differences in user interface could even have an effect when training is performed in situ within an application's interface, because removing the guidance or restrictions provided by these techniques may change retrieval triggers or affect the cost versus benefit of using an expert technique. For example, when techniques for encouraging expert interaction techniques such as the explicit time penalties explored in Grossman et al. (2007) are removed, the user may revert to novice techniques.

3.1.6 Extrinsic Motivators. Training and usage contexts may also differ in how they reward or punish particular behaviors, and this could affect persistence. In particular, training contexts often include elements that extrinsically motivate the user to adopt and use an expert technique by rewarding its use (or punishing the user for not using it (Grossman et al. 2007)). Similarly, training undertaken in a classroom or experimental setting may place an expectation on the user to adopt and use the technique, whereas in the usage context, there may be no such extrinsic motivator. If a user is only making use of a technique because of these external forces, then the technique may not persist when moving to a usage context.

3.1.7 Physical Environment. Finally, differences in the physical environment or workspace in which training and usage are undertaken could also influence post-training persistence. For example, when undertaking training, the user may set up the environment specifically for performing the expert technique—participants learning keyboard shortcuts may keep their hands on the home row while training, even if they would typically have one hand on the mouse during regular use. The physical environment could also be seen as changing available retrieval triggers and required execution actions between the two contexts.

3.2 Summary

There are several factors related to the differences between a training context and a usage context that can affect the post-training persistence of an expert technique. The studies presented in the following sections how persistence is affected by some of these. In particular, we first conducted two studies to test persistence when the task's performance requirement changes. We then test a scenario that is closer to real-world usage, in which performance requirement, high-level task, physical environment, and extrinsic sources of motivation are all varied.

4 STUDY 1–REDUCING THE PERFORMANCE REQUIREMENT

We are interested in testing whether expert techniques persist across realistic differences between the training and usage contexts, with the goal of creating dedicated training systems. As a first step, we conducted two studies testing whether adoption of an expert technique learned under a high performance requirement will be maintained when the performance requirement is reduced or dropped entirely. As mentioned in the previous section, performance requirement is a particularly important factor to consider because it can be easily manipulated in training, but is largely out of the designer's control in real-world usage. Specifically, our first study was designed to answer the following research question:

Controlling for all other factors, is use of an expert technique (learned under a high performance requirement) maintained when the performance requirement is reduced to the level where the expert technique is no longer needed?

Answering this question is important because we want to show that use of the expert technique is not simply a result of a high performance requirement at the moment when a technique must be used. Past work has demonstrated that performance requirement is a particularly powerful way to motivate users to adopt and use an expert interaction technique in a short period of time, but it is unknown whether such use is maintained when performance requirements change (Gutwin et al. 2015).

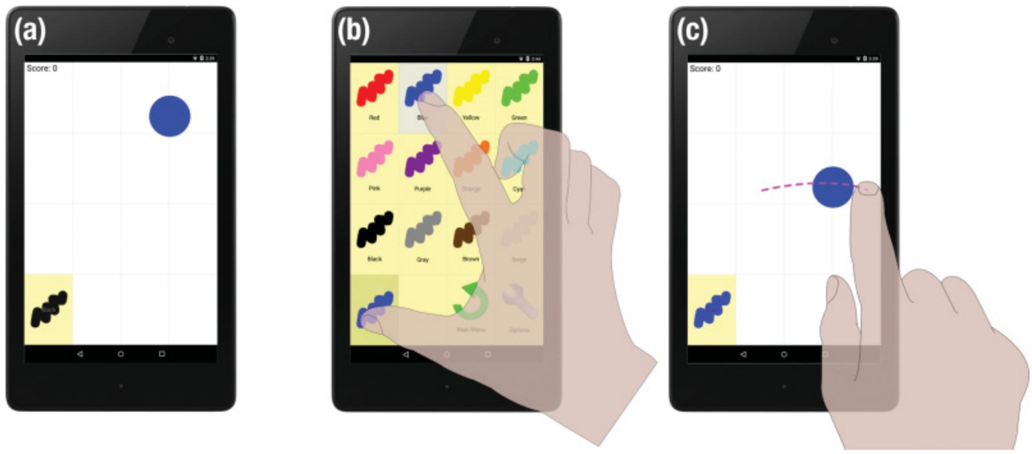

Fig. 3. Trials for Study 1. In each trial, a colored circle falls from the top of the screen (a). The participant selects the matching color (using one of the selection methods) (b), and then draws a slice gesture through the circle (c).

In Study 1, participants played a "Shape Slicer" game (Figure 3) with a performance requirement that changed over the course of the game. In a previous work (Gutwin et al. 2015), this type of game was shown to allow participants to quickly adopt use of expert selection in FastTap menus. However, the performance requirement only increased through the course of the game, so this previous work provided no insights into what happens when the performance requirement drops. As we will describe next, we modified the procedure from Gutwin et al. (2015) to add an additional stage with very low performance requirements at the end of the game.

4.1 Tasks and Stimulus

In each trial, a circle of a randomly assigned color falls from the top of the screen (Figure 3(a)). The participant's goal is to select the matching color from a FastTap menu (Figure 3(b)), and draw a slice gesture through the circle (Figure 3(c)). If the participant successfully slices the circle before it exits the bottom of the screen, the trial is coded as a success, otherwise it is coded as failed. Either way, the experiment proceeds to the next trial.

4.2 Study Design

In total, each participant experienced 288 trials, consisting of 24 blocks of 12 trials each. Each block included one trial for each of the 12 items in the menu. The order of the 12 items was randomly shuffled for each block, so the colors did not follow a predictable pattern from block to block.

The 24 blocks were arranged into three stages. In the first stage (blocks 0–3), circles fell at a *Slow* rate (4.6 seconds on screen, 0.5 seconds between trials). This was followed by a second stage (blocks 4–11) at a *Medium* rate (3.0 seconds on screen, 0.7 seconds between trials), and a third stage (blocks 12–19) at a *Fast* rate (1.9 seconds on screen, 0.8 seconds between trials). In the final stage (blocks 20–23), enemies fell at the *Slow* rate again. The final slow stage was intended to test whether participants would maintain their use of one-step selection when the performance requirement was reduced to a point where one-step selection was no longer required.

For each trial, we measured the time to select the matching color, and the participants' interactions with the FastTap menu, including the selection method used (novice two-step selection, or expert one-step selection). We also recorded selection errors (selecting an item other than the matching color).

The method described earlier closely follows the game study used by Gutwin et al. (2015) in terms of the nature of the game, the menu and items tested, and the speed of the first three stages of the experiment. The main difference is the inclusion of a final slow stage, to test whether participants would persist with using the expert technique, or return to using the novice technique.

4.3 Procedure

At the start of the study session, participants completed a consent form and a pre-study questionnaire that included demographic information. Next, the experimenter described the study task, including the objective and mechanics of the game. The experimenter then explained and demonstrated the two selection methods. These were intentionally presented in a neutral manner, identifying them as "two-step selection" and "one-step selection," respectively, to avoid suggesting that one method was superior to another. Participants were told that the game will include different stages, where the speed of the falling circles will change, and that they should try to be as quick and accurate as possible. Following the experiment, each participant completed a post-study questionnaire. In total, the experiment took ~30 minutes to complete, of which ~18 minutes were spent playing the game.

4.4 Participants and Recruitment

In this study, we were interested in how participants would perform in the final (Slow) stage of the game *after* they had transitioned to using the one-step selection method. However, pilot studies revealed that not all participants put equal effort into adopting the one-step selection method, even though failing to do so makes it difficult to do well in the game. This lack of adoption may have been aggravated because we intentionally did not instruct participants to use the one-step selection method, to avoid biasing them toward using one selection mode over another. With the preceding as motivation, we adopted a recruitment strategy in which we continued to recruit participants until we had reached 12 participants who passed an exclusion criterion of at least 25% usage of one-step selection in the Fast stage of the experiment.

In all, we recruited 14 right-handed participants from a university campus, of which 12 passed our exclusion criteria (9 male, 3 female, ages 19–34 (mean 25, SD 5)). None of the participants reported having a color vision deficiency. All participants were given a $10 honorarium for participating.

4.5 Results

Figure 4 shows participants' average use of one-step selection over the course of the four study stages. Comparing the final Slow stage with the Fast stage preceding it, we can see that use of one-step selection does not decrease in the final stage, and in fact continued to increase. The average per-participant rate of one-step selection in the final Slow stage was 92.9%, as compared to 83.5% in the final four blocks of the Fast stage. An RM-ANOVA did not show a significant main effect ($F_{1,11} = 4.2$, $p = 0.07$, $\eta^2_G = 0.06$) between these two periods.

Examining each participant's data individually, we found that only one participant used one-step selection less frequently in the final stage than in the final four blocks of the preceding stage, and only by a small amount (a decline from 100% to 97.5%). All other participants maintained or increased their use of one-step selection in the final stage.

These results indicate that participants continued to use one-step selection, even after the performance requirement was reduced. This provides evidence that the use of the expert technique for FastTap was not just a transient effect of the high performance requirement.

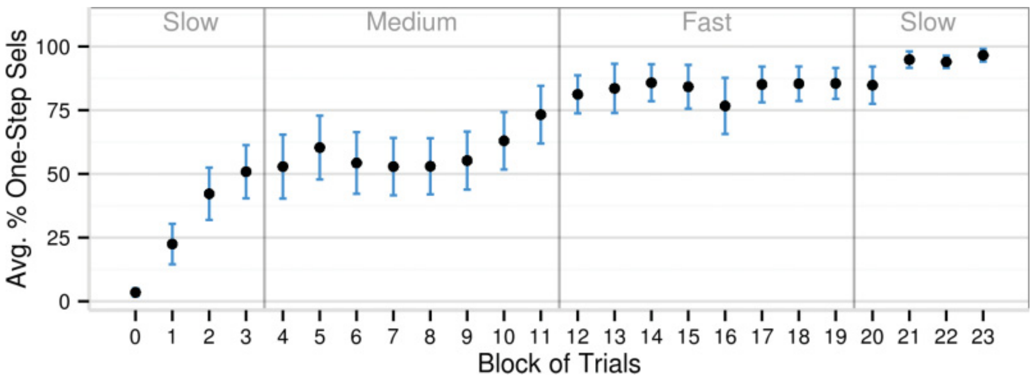

Fig. 4. Average per-participant use of one-step selection for Study 1, computed as the mean of per-participant means. Error bars indicate standard error.

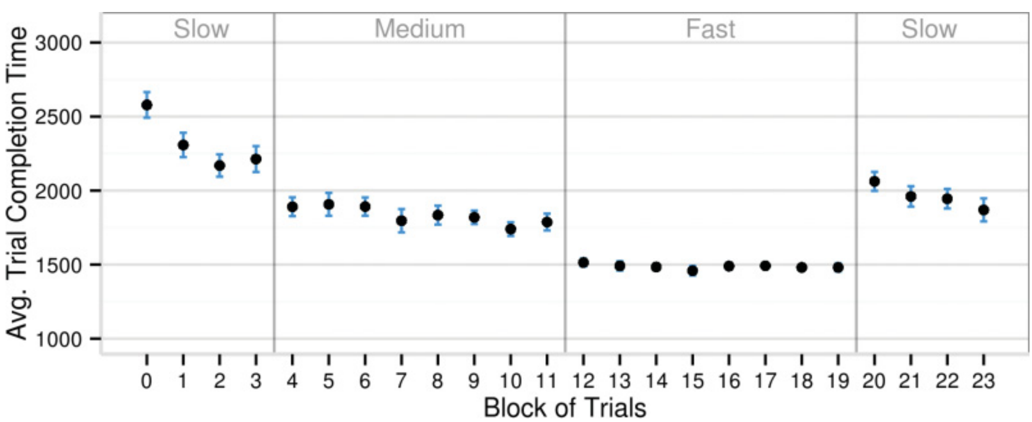

Fig. 5. Average per-participant trial completion times for Study 1, computed as the mean of per-participant means. Error bars indicate standard error.

4.5.1 Time to Complete Trials. Figure 5 shows the average trial completion time, which we define as the time taken to select the correct item during a trial. This metric is useful because it gives an idea of how quickly participants were acting during the trials, independent of the selection method they were using. We can see that, despite maintaining their use of one-step selection in the final Slow stage, participants selected items at a more relaxed pace as compared to the Fast stage preceding it.

The average per-participant trial time for the final Slow stage was 1959ms, as compared to 1486ms for the final four blocks of the Fast stage. An RM-ANOVA showed a significant main effect for trial time between the Final Slow stage and the last four blocks of the Fast stage ($F_{1,11} = 100.0$, $p < 0.001$, $\eta^2_G = 0.75$).

This result provides further confirmation that performance requirement is not inducing a transient use of the expert selection technique—participants took significantly more time to select items, demonstrating an awareness that performance requirement had dropped, even while they maintained or increased their use of one-step selection.

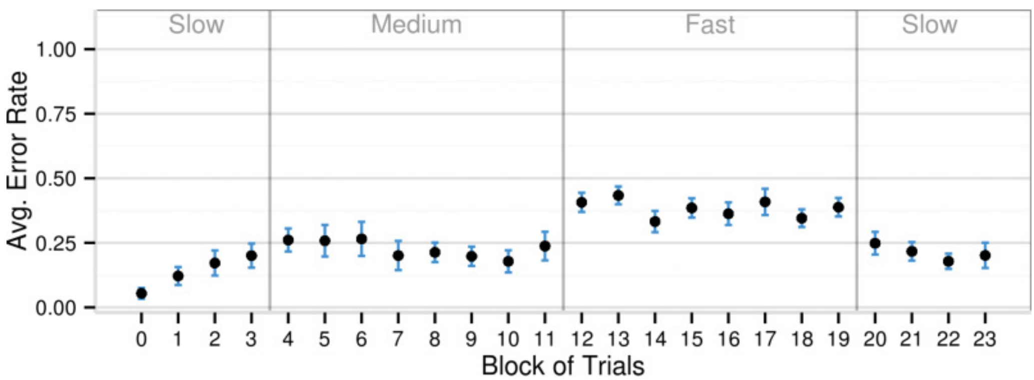

Fig. 6. Average error rates for Study 1, computed as the mean of per-participant means. Error bars indicate standard error.

4.5.2 Error Rate. Finally, we examined the average error rates across the study blocks (Figure 6). We see that the highest error rate was during the Fast stage, with the error rate dropping during the final Slow stage.

The lower error rate in the final Slow stage, coupled with the longer trial completion time measure, discussed earlier, suggests that participants were able to take advantage of the lower performance requirement in the final Slow stage to perform the one-step selection more carefully, perhaps taking more time to recall each item's location, or targeting their selection more carefully. However, the mean error rates in the final Slow stage of greater than 15% suggest that participants were still in the process of learning the one-step selection method.

4.6 Discussion

The findings from our first study suggest that expert interaction techniques can persist and continue to be used when performance requirement drops. This is significant when considering the alternative—what if we had observed that participants did *not* maintain use of the expert technique when performance requirement dropped? If this were the case, it would suggest that a high performance requirement must be maintained in order for use of the expert technique to continue, which would be discouraging for dedicated training systems. As it stands, our findings suggest that performance requirement can effectively be used to motivate a user to acquire the necessary skill to use an expert technique, and that use of that technique will be maintained once the performance requirement is removed.

Though our results are encouraging, we acknowledge that our study has several limitations. While we have presented evidence that use of the expert technique is maintained even when it is no longer strictly necessary to complete the task, it is possible that a number of other pressures are working on participants to create this effect. Our study results show some evidence for this, as there is a clear trend of participants beginning to adopt the expert technique even during the initial Slow stage, where they could easily complete the task using two-step selections alone. There are several possible reasons for this behavior.

First, it is possible that even the low performance requirement of the Slow stage imposed some sense of urgency, and this was sufficient to encourage participants to begin switching to the expert technique in the initial Slow stage, and maintain use of it in the final Slow stage. Second, participants may have anticipated that the speed of circles would increase, and preemptively attempted to transition during the initial Slow stage. This could explain their behavior in the final

Slow stage as well—participants may have viewed this stage as an opportunity to increase their knowledge of the expert technique in anticipation of another Fast stage. Third, it is possible that the overall context of "playing a game" affected participants' behavior by creating an expectation to try and perform well. Finally, we instructed participants at the start of the study to "Try to be as quick and accurate as you can"—a standard instruction for studies of interaction techniques, but an instruction that may have implied a performance requirement.

Our next study was designed to test some of these limitations by removing the performance requirement and any expectation of further fast stages.

5 STUDY 2–REMOVING THE PERFORMANCE REQUIREMENT

The results of Study 1 suggest that performance requirement was not the main factor driving participants to use the one-step selection method (at least in our experimental setting). However, the final testing stage in Study 1 did still impose some performance requirement, and we did not explicitly tell participants that they would not face a further fast stage. Thus, participants may have continued to use one-step selection out of an expectation of a future performance requirement.

To rule this out, we conducted a study that was much the same as Study 1, but before the final stage, the study system stopped and prompted the participant to listen to the experimenter. The experimenter then instructed the participant: "OK, you're almost done. In the next stage, the enemies will appear, but they won't move. Once you swipe an enemy, there will be a short delay, and then the next enemy will appear. There are 48 enemies in total, and you can work at your own pace to get through them."

This study design also introduces a limited test of whether use of an expert technique transfers when the task changes—the task in the final stage is no longer "swipe the circle before it reaches the bottom of the screen" but rather "swipe the unmoving shape at your own pace." Although the task mechanism (swiping) was the same, and the final stage was still carried out in the game environment, the complete lack of movement made the task in the final stage qualitatively different from the rest of the game.

5.1 Tasks and Stimulus

Trials were carried out in a similar fashion to Study 1 (see Section 4.1). To accommodate the changes to the final stage, we re-developed the study system, and in the process made a few additional changes. We made the circles larger, increasing their radius to 1/4 the screen width, with the idea that this could make it easier for participants to slice them. We also made the game slightly easier, decreasing the circle speeds for each of the stages, and regularized the time between trials to be 1 second for all trials. Finally, we made a few minor visual changes to the menu (see Figure 7).

5.2 Study Design

As in Study 1, each participant experienced 288 trials, consisting of 24 blocks of 12 randomly-ordered trials each. With our modifications to the study system, the enemy speeds for the stages of Study 2 were *Slow*—5.1 seconds on screen, *Medium*—3.3 seconds on screen, *Fast*—2.1 seconds on screen, with a consistent delay of 1.0 seconds between trials. Finally, as described earlier, in the final *Static* stage, circles were stationary and remained on the screen until sliced. There was a 1.0 second delay between trials in the final stage.

5.3 Procedure

The overall study procedure was the same as that of Study 1, with the exception of the final Static stage in place of a final Slow stage, and the minor modifications to trial timing described earlier.

Fig. 7. The game application used in Study 2. Apart from some small changes in visuals and the speed of the circles, the system is the same as that for Study 1.

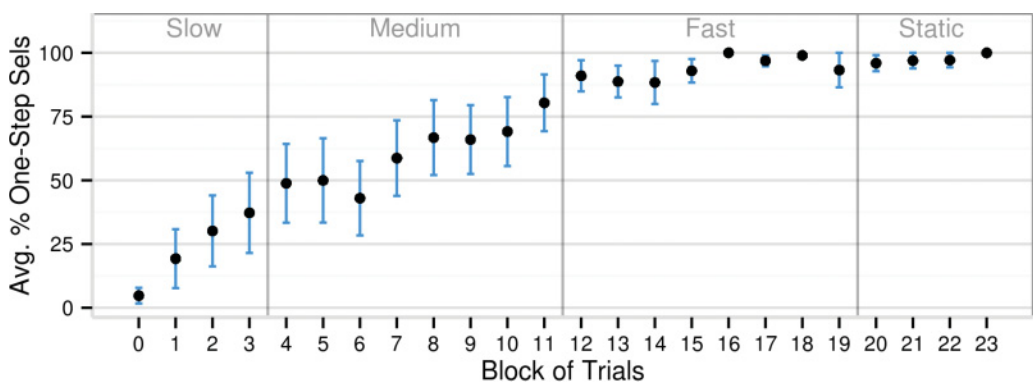

Fig. 8. Average use of one-step selection for Study 2, computed as the mean of per-participant means. Error bars indicate standard error.

5.4 Participants

For Study 2, we recruited nine participants, of which two were excluded from our analysis. (One excluded participant did not meet our exclusion criteria of at least 25% use of one-step in the Fast stage. The other was excluded because, after starting the game, it became clear that the participant did not understand how to perform one-step selection). Demographics for the remaining seven participants were consistent with those of Study 1 (3 male, 4 female, ages 19–35, mean 25, SD 5). All participants were right handed, and none reported color vision deficiencies. The study took about 30 minutes to complete, and participants were given a $5 honorarium for participating.

5.5 Results

The average use of one-step selection across the study blocks is shown in Figure 8. As in Study 1, we do not see a drop in one-step usage in the final stage of the study. In fact, the average per-participant rate of one-step selection in the Static stage was 97.4%, as compared to 96.9% in the final four blocks of the Fast stage. A RM-ANOVA did not show a significant main effect ($F_{1,6} = 0.8$, $p = 0.41$, $\eta_G^2 = 0.002$).

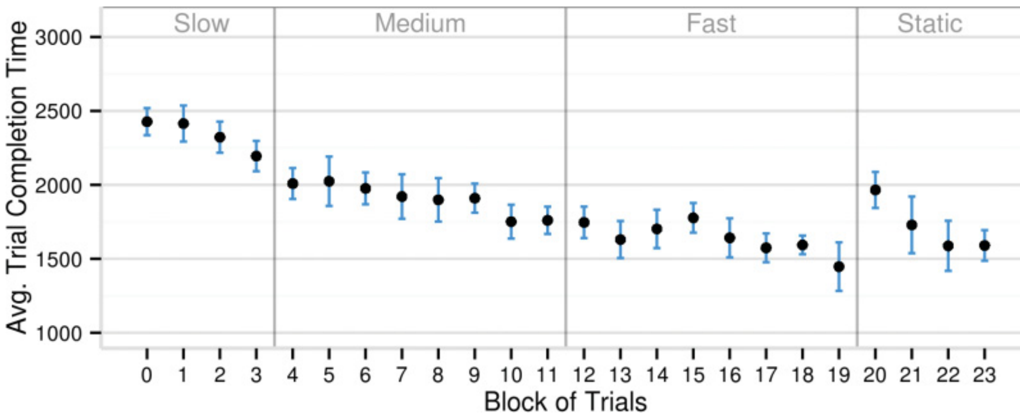

Fig. 9. Average trial completion times for Study 2, computed as the mean of per-participant means. Error bars indicate standard error.

Looking at participants' data individually, only one of the seven participants used one-step selection less in the final stage than in the previous four blocks (a decline from 100% to 98.5%). Of the six remaining participants, four maintained 100% use of one-step selection through both the final four blocks of Fast and the four blocks of Static, and two increased their usage in the Static stage (one from 98% to 100%, and the other from 80% to 83%).

The preceding results suggest that prompting participants to complete trials "at their own pace" and removing any performance requirement imposed by moving circles did not reduce the use of the expert technique in the final stage.

5.5.1 Time to Complete Trials. The average per-participant trial completion times for Study 2 can be seen in Figure 9. The results mirror those of Study 1—we see a rise in the average time to complete trials at the start of the final stage, when the performance requirement is removed, followed by a gradual decrease over the four blocks of the final stage. As in Study 1, this indicates that participants are aware that they no longer need to perform selections quickly, even while they maintained their use of the expert selection technique.

5.5.2 Error Rates. Average error rates for Study 2 are shown in Figure 10. In general, the error rates appear to be lower than those of Study 1, in particular for the Fast stage. This suggests that our modifications to ease the difficulty of the game were effective.

5.6 Discussion

Consistent with the results of the first study, use of the expert selection technique was maintained or increased by the nearly all participants in the final stage, even despite the fact that trials in this stage had no time limit for successful completion, participants were instructed to work at their own pace, and it was clearly communicated that this was the final trial block (i.e., there would not be additional fast trials). This provides further evidence that a high performance requirement imposed in a training context (to motivate participants to adopt an expert interaction technique), does not need to be maintained for the use of the expert technique to persist, at least in our experimental setting.

It is possible that additional factors, such as extrinsic motivators (e.g., the experimental setting), or the high-level task (i.e., playing a game) are behind participants maintaining use of the expert

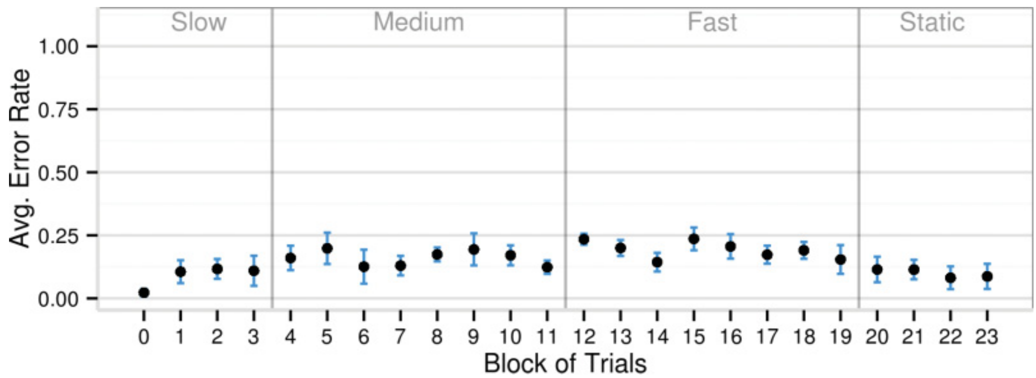

Fig. 10. Average error rates for Study 2, computed as the mean of per-participant means. Error bars indicate standard error.

selection technique. To investigate this possibility, we conducted a third study, described in the next section.

6 STUDY 3–PERSISTENCE IN DEDICATED TRAINING

Our first two studies suggest that performance requirement and the expectation of a future performance requirement alone are not the drivers of participants using the expert technique once they have learned it. This is encouraging, because it suggests that a high performance requirement can be imposed to motivate users to learn, and that, other things being equal, the performance requirement does not need to be maintained for users to continue using the expert technique. However, in real world dedicated training scenarios, other factors are not equal—in moving to a usage context, performance requirement will change along with additional extrinsic factors, such as the high-level task being performed, the user interface, extrinsic motivators, and the physical environment in which the technique is being used.

Our goal with Study 3 was to test post-training persistence in a more realistic scenario, in which a range of different factors are changing at once. We were particularly interested in examining the effect of different high-level tasks on maintenance of the expert technique. The specific research questions were the following:

- Once consistent use of an expert technique has been established in a training context, will it persist when performance requirement, high-level task, extrinsic motivators, and physical environment are changed?
- Will some types of high-level tasks (e.g., repetitive, creative, and complex) cause users to revert to the novice technique?
- How will an increasing time gap between training and usage contexts affect use of the expert technique?

To answer these questions, we first trained participants to use the menu of a touch-based drawing application through a game similar to that used in Studies 1 and 2 (Figure 11(a)). We then asked participants to complete a series of drawings over a week's time using the drawing application itself (Figure 11(b)). These two applications are intentionally similar to those used in previous work, in which it was shown that participants did *not* adopt one-step selection, even after spending 10 weeks of completing drawings on a regular basis (Gutwin et al. 2015). This previous finding suggests that it will be difficult to achieve transfer of the expert technique from the training game to

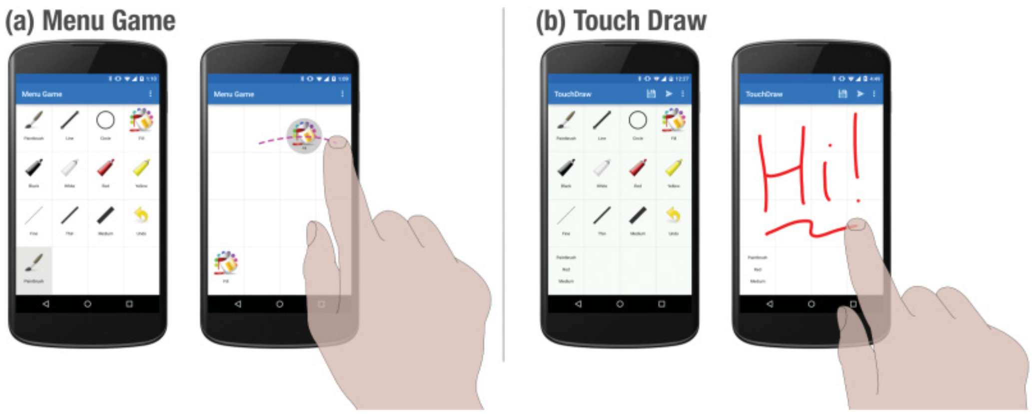

Fig. 11. The two applications used in Study 3.

the drawing tasks. However, in contrast to this previous work, our training procedure for this study was designed to ensure participants have learned and are able to successfully perform the one-step technique before they begin performing drawing tasks. In terms of previous work on transfer, our approach is for participants to *overlearn* the spatial locations of the commands (Driskell et al. 1992; Ebbinghaus 1913).

6.1 Procedure

In an initial session, conducted under an experimenter's supervision, participants installed the menu game and drawing application on their personal Android devices. They were then introduced to the drawing application and the game interfaces, and told "Today, we'd like you to play a game that will train you to use the menu for the drawing application. We want you to try and learn the locations of all of the menu items." They then played the game, which took approximately 18 minutes.

After the initial session, participants returned to their everyday activities. In the following days, participants were emailed sets of drawings (two per day) to complete on their own time, and were asked to send the drawings back to the experimenter. The first set of drawings was emailed out less than an hour after the initial session. Subsequent sets of drawings were sent out 1, 3, 5, and 7 days after the initial session. In each case, participants were given a drawing to reproduce, and asked to email the drawing and logs back to us within 12 hours of receiving the email.

At the end of the study, participants met with the experimenter to complete a demographics questionnaire and a short post-study questionnaire.

6.2 Tasks and Stimulus

The training game in the initial session was similar to that used in Studies 1 and 2—circles showing menu items fell from the top of the screen, and the participant needed to select the matching item and draw a slice gesture through the circle before it reached the bottom of the screen. As before, participants could select items using either two-step or one-step selection. However, the game followed a slightly different pattern of stages than in Studies 1 and 2—after an initial *Slow* stage, stages alternated *Medium, Fast, Medium, Fast*, and then ended on *Medium*. Our rationale was that this would encourage participants to adopt one-step during the game by allowing them to experience some *Fast* blocks early on, then provide some *Medium* blocks where they could increase their skill. Otherwise, the enemy speeds for Study 3 were consistent with Study 2: *Slow*—5.1 seconds on

Fig. 12. Drawing tasks assigned to participants during Study 3. Five sets of two drawing tasks each were assigned over the course of a week after the initial session.

screen, *Medium*—3.3 seconds on screen, *Fast*—2.1 seconds on screen, with a consistent delay of 1.0 second between trials. Each stage consisted of four blocks of 12 items each (one for each menu item, presented in randomized order).

As mentioned earlier, the drawing application's menu contained the same items as the game (Figure 11). Using this menu in the drawing application, the user can set the current tool (*Paintbrush, Line, Circle*, or *Fill*), color (*Black, White, Red*, and *Yellow*), and line thickness (*Fine, Thin*, and *Medium*). The menu button displays the current settings for these three attributes. With the menu closed, the user can draw using the current tool and settings by touching the screen area outside the menu button. An *Undo* command is also included in the drawing application.

The assigned drawings for each day of this study are shown in Figure 12. The *Cat* drawing was repeated on each trial day, to test whether repetition of a known drawing task would affect persistence of the expert technique. On Day 1, the *Shapes* drawing was included to provide a

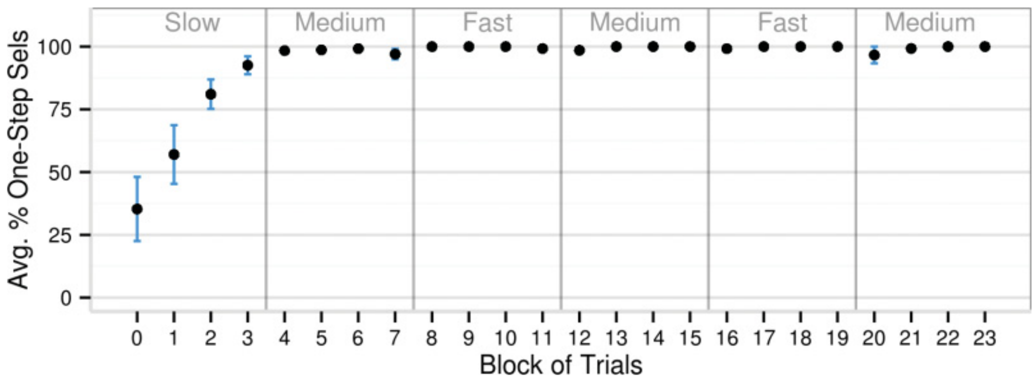

Fig. 13. Learning curve for the game portion of Study 3, showing percentage use of one-step selections, computed as the mean of per-participant means. Error bars indicate standard error.

drawing that required little planning, but forced the user to make frequent selections in the menu. The *Beaker* drawing on Day 5 was designed to be more complex than other drawings, and therefore to require more planning of which tools to use. Finally, on Day 7 participants were given an open-ended *Creative* drawing task, where they could draw what they wished, to test whether creative planning would influence the persistence of the expert technique.

6.3 Participants

We recruited 10 participants from a university campus (8 male, 2 female, ages 19–34 (mean 25, SD 5)). One participant was left handed, while the rest were right handed. Participants were given a $30 honorarium for participating in the study. We imposed an exclusion criterion that participants must exhibit at least 80% usage of the expert technique by the final four blocks of the initial game session, which was met by all participants.

6.4 Results

6.4.1 Game Session–Transitioning to One-Step. Figure 13 shows participants' average use of one-step selection for the game portion of Study 3. We can see that participants quickly adopted one-step selection in the initial *Slow* stage, and then used it nearly exclusively through the remaining stages.

In comparison to Studies 1 and 2, participants adopted one-step selection much more quickly during the game portion of this study. There are a number of potential explanations for this. First, the item positions in the menu may have been easier to remember, because they were grouped into semantic categories by row (tools, colors, and line thicknesses), and in the case of line thicknesses, followed a predictable thin-to-thick progression within the row. Second, in this study, we explicitly instructed participants to try and learn the locations of the items in the menu. Finally, the approach of alternating *Medium*/*Fast* stages may have encouraged participants to more quickly adopt one-step selection.

6.4.2 Game Session–Error Rates. Figure 14 shows participants' average error rates for the game portion of Study 3. The error rates are lower than those in Study 2, providing further evidence to suggest that the items in the menu for this study were easier to learn than the color menus used in Studies 1 and 2.

6.4.3 Drawing Tasks. In general, we had good compliance with study instructions, with all participants completing all of the assigned drawing tasks, and the majority completing them within

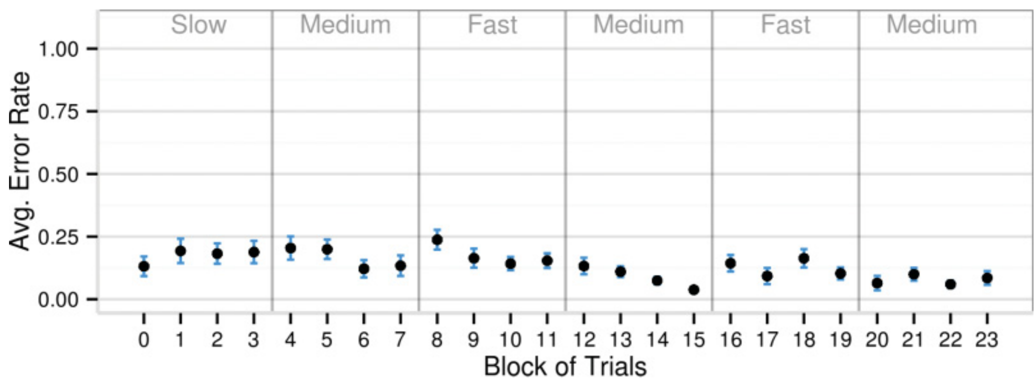

Fig. 14. Average error rates for the game portion of Study 3, computed as the mean of per-participant means. Error bars indicate standard error.

the 12-hour timeframe—only two participants, P5 and P7, required additional reminders and completed a number of the drawings on the day following that where they were assigned. However, all participants completed each set of drawings before the next set was assigned.

Figure 15 shows each participant's use of one-step selection for the 10 drawing tasks assigned during the study. Overall, more than 50% of selections were made with the one-step method—however, it was not the case that participants all used the expert technique half of the time. Instead, we observed a drastic difference between participants, with P1, P3, P5, P8, and P10 using one-step selection almost exclusively (with the exception of the *Ladybug* drawing for P8), and P2, P4, P6, and P9 almost exclusively using two-step selection. The only participant that appeared to use a mix of one-step and two-step selection is P7. Moreover, P7 commented in the final wrap-up session that he realized partway through the study that an advantage of two-step selection was the ability to open the menu and set more than one parameter (e.g., a tool, color, and line thickness). This may suggest that P7 spent more time experimenting with different methods for making selections than other participants, which could explain why his data differs from those of the other participants.

The finding that the expert technique did robustly transfer to a very different usage context for five of the ten participants suggests that, for some people, expert skills learned during a brief dedicated training session (in this case an ~18 minute game) *can* persist even when transferring to a usage context with very different high-level tasks, extrinsic motivators, physical environments, and performance requirements. In addition, for these participants the post-training persistence of the expert technique was consistent across a range of task types, including those that they had performed multiple times before (the *Cat* task), tasks designed to require numerous selections (the *Shapes* task), tasks that required more planning (the *Beaker* task), and tasks requiring creative thought (the *Creative* task).

However, the aforementioned result must be considered in light of the opposite result for four other participants. For these participants, there was almost no use of the expert method, even for commands that were carried out very frequently, and regardless of the type of task being carried out.

This pattern of all-or-nothing persistence is surprising. In the next section, we present the results of our post-study questionnaire, which was conducted to gain insights into this phenomenon. We also followed up on P7's comment that an advantage of two-step selection was the ability to perform multiple two-step selections for a given menu opening (i.e., to amortize the cost of opening the menu by making multiple selections, such as a tool and a color). To check whether

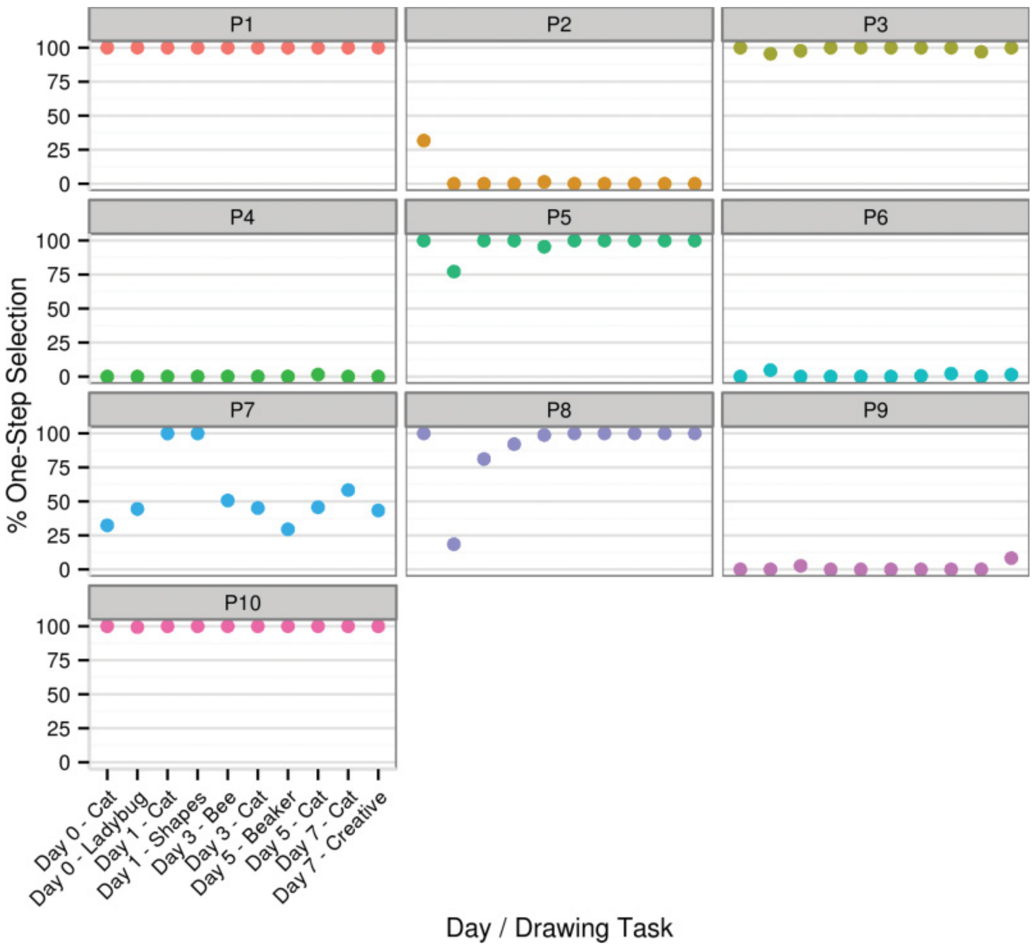

Fig. 15. Percentage of one-step selections for drawing tasks in Study 3.

this practice was widespread, we examined the two-step selections made by the four non-one-step participants (P2, P4, P6, and P9). In all, these participants opened the menu 2493 times over the course of the drawing tasks, and 17% of these menu openings were associated with two or more item selections. On an individual basis, the percentage of menu opens with two or more selections was between 11% and 20% (P2 20%; P4 11%; P6 16%; P9 19%). That this practice was performed by all participants may suggest that it contributed to their preference for the two-step technique.

Finally, we were interested in whether the non-one-step participants became faster with their two-step selections over the course of the 5 days of drawing tasks. To examine this, we analyzed the median time between when the menu button was pressed and a first item was selected, across the 5 days of drawings. This analysis did reveal a downward trend, with the median time for the 5 days being 939ms, 895ms, 836ms, 809ms, and 775ms, respectively. This suggests that these participants were becoming more confident in item locations and relying less on visual search for items over time. Further, this might suggest that these participants were gradually moving toward a point where they would switch to the expert selection technique (i.e., following the rehearsal hypothesis), though we cannot say this conclusively because we did not observe participants making the transition.

6.4.4 Post-study Questionnaire. In the post-study questionnaire, we tested participants' memory of the menu by presenting them a blank menu grid on paper, and then asking them to fill in the items from the study system. All but one participant filled in all item locations correctly. P2, who used two-step selection almost exclusively for the drawing tasks, placed four items incorrectly, all in the colors row (i.e., they ordered the items in this row incorrectly). These results suggest that most participants' memory of the actual item locations—regardless of whether they used one-step or two-step selections for the drawing tasks—was perfect or nearly so, suggesting that simple memory accuracy was not the main reason why some people reverted to using one-step selection.

We also asked participants to rate their confidence in their answers about the locations of items. P1, P3, P5, P7, P8, and P10—the one-step users—rated their confidence as 5/5 for all items (in response to the question "How confident are you that you have recalled the menu item locations correctly? 1 = Not at all confident, 5 = Very confident)." The two-step users showed mixed results: one participant (P6) rated their confidence as 5/5 for all items, and three more (P2, P4, and P9) rated their confidence as 4/5 for all items. Although the difference is not large, and a confidence of 4/5 is still reasonably high, this result may suggest that the non-one-step users chose not to use the expert method because they were less confident in their memory of the item locations (even if they had in fact accurately learned the locations). However, it is also possible that the higher confidence ratings for one-step users are the result of their continued use of one-step selection during the week of drawing tasks (which may have increased their confidence in the item locations).

To better understand the reasons why some participants used one-step selection for the drawing tasks, while others did not, we presented participants with a set of possible reasons and asked them to mark the reasons that they agreed with. Specifically, we asked the following question:

> Think about situations where you used the *two-step* method to select items while working on the drawing tasks. We want to understand the main factors that contributed to you *not* using the one-step method in these situations.
> Read through the following list of reasons, and then place a "1" next to your main reason for not using the one-step method (or write in a new reason). If there are multiple reasons, mark them in order of importance with "1", "2", "3", and so on.

Table 1 shows the options we presented to participants, with the answers selected or written in by the two-step users (and P7).

Each of the statements "I couldn't remember the positions of some of the items" and "I didn't want to accidentally choose the wrong item" were selected by three participants, with P6 and P9 agreeing to both statements. This may provide further evidence that confidence in item locations contributed to the choice to use the two-step selection technique.

Another possibility is that the posture in which participants used their devices contributed to their choice of selection. Two participants selected the statement "It was awkward to hit the menu button and an item at the same time," and P9 wrote in "The size of the screen was large, hence it was easier to use two fingers, one after holding the menu." P9's device was a tablet with a 7-inch-diagonal screen, which was the largest device used by a participant in the study. Moreover, based on observations of this participant during the initial session, and his write-in response, we suspect that he was using both hands to perform selections, with the thumb of the hand holding the device used to hit the menu button, and the fingers of the opposite hand used to make selections. In our own tests, this is a more awkward configuration for performing one-step selections, which may explain his preference for the two-step selection method.

Table 1. Rationales indicated by non-one-step participants (P2, P4, P6, P7, and P9) for using two-step selection

I'm not sure why, I just preferred the two-step method.	P4
It was awkward to hit the menu button and an item at the same time.	P4, P6
The two-step method was fast enough.	P4, P7
I couldn't remember the positions of some of the items.	P2, P6, P9
Opening the menu helped me plan out the drawing I was working on.	P7
I wasn't in any hurry.	P6
I didn't want to accidentally choose the wrong item.	P6, P9, P7
Seeing all the items in the menu was useful.	P2, P6
Remembering the locations of items takes time/effort.	(none)
I just didn't think to use one-step method while working on the drawings.	(none)
I always tried to use one-step, but may have used two-step by accident in some cases.	P2
Other (please specify)	P9—"The size of the screen was large, hence it was easier to use two fingers, one after holding the menu." P7—"Selecting multiple options in one go."

7 DISCUSSION

Our three studies examined the transfer of a memory-based expert selection technique when changing several factors between a training context and a usage context (including performance requirement, retrieval trigger, task type, extrinsic motivation, and physical environment). The main findings from the studies are as follows:

- In the game-based training context, participants adopted the expert selection method quickly (within 24 blocks and ~18 minutes), and used the expert method consistently during the game (Studies 1–3).
- Once participants were trained on the expert selection method, reducing the performance requirement (Study 1), or removing the performance requirement entirely (Study 2), did not reduce use of the expert method (within the game-based training context).
- Changing several elements between the game training context and the usage context (similar to what would change in a real-world training scenario) led to "all-or-nothing" use of the expert method—five of ten participants used the expert method almost exclusively, and four of ten almost never used it. Only one participant showed partial use of the expert method.

7.1 Explanations for Results

Here, we consider potential explanations for our main results, and several questions raised by these results.

Why did use of the expert method persist after reducing or removing performance requirements (Studies 1 and 2)?

Our conceptual framework for post-training persistence (Section 3) identified performance requirement as an important potential difference between training and usage contexts—a high performance requirement can be a valuable tool during training, as it can motivate people to adopt an expert technique. In Studies 1 and 2, participants continued to use the expert method even after the performance requirement was dramatically reduced or removed entirely. This means that even though the increasing performance requirement may have prompted the switch to the expert technique, decreasing the requirement did not have the opposite effect—people continued to use the expert method even when there was no longer any need to perform quickly. There are several potential reasons for the persistence of the expert method in these studies.

Our framework suggests that the differences between training and usage contexts (e.g., differences in task type, motivation, or performance requirement) can lead people to revert to using a novice technique instead of a newly learned expert method. In addition, prior work on skill transfer has shown that a variety of factors in the training program, in the individual, and in the usage environment, can have positive or negative effects on transfer and retention. The results of our first two studies suggest that a combination of factors result in the behavior we observed in the usage context, with some factors increasing the likelihood of persistence and some reducing the likelihood. In Studies 1 and 2, we modified only one factor (the performance requirement). This was done in order to test whether this aspect had a dominating effect on transfer, and our results show that it does not—instead, it appears as though the factors that promoted transfer in our study outweighed the effect of reducing the performance requirement. For example, our studies promoted overlearning during the training phase, and we used similar retrieval triggers, tasks, and physical environments for both the training and "usage" contexts. Therefore, it seems likely that if participants underwent an unconscious calculation when deciding what method to use for the final low-performance blocks, there were simply more terms in the equation that promoted the use of the expert method (e.g., "I remember the expert method well" and "this new setting is similar to the one I just experienced"), than terms that detracted from the expert method (e.g., "I don't need to go fast for this task, so I can use the old method").

However, the results of the first two studies clearly show that performance requirement is not as critical a "slippage factor" as we first expected. This is good news for designers who would like to use performance requirements as a way to increase the efficacy of training regimes.

Why did use of the expert method persist for some people in Study 3?

The continued use of the expert method observed in Study 3 can be explained along similar lines to the earlier discussion—the factors promoting the use of the expert technique outweighed the factors detracting from its use. What is interesting about Study 3, however, is that there was substantially more distance between the training context and the usage context (in terms of time since training, task environment and type, risk of making errors, physical environment, experimenter pressure to use the expert method, and performance requirement). All of these factors have been suggested in prior work as reasons that skill transfer can fail—but despite this substantial combined difference, the majority of participants in Study 3 continued using the expert method. We see six potential factors that may have contributed to the overall persistence of the expert method for these five participants.

The effectiveness of overlearning. Once a user achieves a high level of mastery in the use of a technique, it can become the normal way to carry out an action (e.g., many users use Ctrl + S to save a document without a great deal of conscious thought). One way in which mastery is achieved is though overlearning (Ebbinghaus 1913)—repeated practice after an initial level of facility with the technique has been reached. In our study, all participants learned the expert selection method quickly, and then carried out several further blocks of trials. By the end of the training session, participants demonstrated a high degree of mastery—that is, near-complete use of the expert method. Overlearning has been shown to have a positive effect on skill transfer (e.g. (Driskell et al. 1992)), and so it is possible that the amount of practice that we provided in the expert technique was enough to move some participants to a level where one-step selection was their default method. The error rates seen in Study 3, however, provide conflicting evidence about mastery—participants' error rates were still above 10% in the final block of training, suggesting that they had not yet memorized the command locations.

Recognition of the benefits of the expert method. The goal of most training programs is to provide users with knowledge and skills that will actually help them in their work, and despite the documented challenges of achieving transfer of training to usage contexts, there are cases where users recognize the benefits of switching to the expert technique. Therefore, one possible reason why five people continued to use the one-step selection method may simply be that they saw the value in that technique (e.g., recognizing that it was faster) and made a conscious decision to continue using it.

Limited pre-existing inertia for the novice method. Previous research in HCI has suggested that when one kind of interaction is well-practiced, there is inertia for this method that can prevent users from switching to a more efficient technique (e.g. (Carroll and Rosson 1987; Fu and Gray 2004)). In our third study, even though participants first learned the two-step selection method, they quickly made a transition to the one-step expert method, and therefore they may not have built up as much inertia with the novice technique as would be present in other settings.

Intrinsic (and possibly extrinsic) motivation. Although we did not ask our participants questions about their motivations, it is possible that some participants continued with the expert technique because of motivational factors, either intrinsic or extrinsic. For example, some users like achieving expertise with a software system, either for their own personal satisfaction or because they want to be seen as an expert user by others (either peers or the experimenters). That said, our instructions clearly indicated to participants that the drawing tasks would be carried out individually, and that people could carry out the tasks in any way they wished, so we believe that social standing or experimenter approval is less likely to have motivated people than personal interest or satisfaction.

Recency and momentum. Previous research has shown that the length of time between training and application of the learned skills or knowledge negatively affects transfer (e.g., (Rohrer et al. 2005)). In Study 3, participants were assigned their first drawing the same day as the training session, so the continued use of one-step selection may have been encouraged by the recency of training—that is, there may have been some momentum for the expert method that carried over from the game to the at-home tasks. However, if recency is a strong factor, we might expect to see decay in the use of one-step selection over time, and our results show that use of the expert method was as strong at the end of the at-home part of the study as it was at the start. Another possibility is that recency is a strong factor, but that the individual drawing tasks acted as additional training and kept the momentum going through the remaining study tasks, despite the two-day gaps between the later tasks.

Training-to-usage distance may have been smaller than expected. It is possible that our analysis may have overestimated the distances between the training and usage contexts. Although it is obvious that the differences were large enough that four participants did not continue using the expert method, it is worth re-assessing elements that could have mitigated the distances between training and use. It is true that several aspects of the setting were the same or similar (e.g., participants used the same devices for the game and the drawing), and the design of the drawing system may have reduced some distances (e.g., selections in the drawing menu were non-destructive, so the perceived risk of making a selection error was perhaps reduced).

Despite the possibility that some differences between the training and usage contexts were smaller than expected, there were still substantial differences in the task type, the physical setting, the instructions, and the performance requirements. It is therefore also worth noting that the positive influences of the aforementioned factors on the five users' decision were not outweighed by the negative influence of other distance-creating factors. That is, the differences between the overall task of "drawing" versus "playing a game" were not enough to cause a slip back to the novice method, and nor were the reduction in performance requirement, the change in physical setting, or the lack of experimental oversight on the completion of the tasks. This result provides encouraging evidence that training in expert methods is not likely to be derailed by these kinds of changes, at least for some users.

The preceding factors, whether individually or in combination, may have played a role in the five one-step users' decision to continue using the expert method. It is obvious, however, that these factors do not have an overwhelming effect, since they did not change the behavior of the four users who used two-step selection for the drawing tasks. We consider these users next.

Why did use of the expert method not persist for four users in Study 3?

Four of ten participants showed almost no use of the expert method during the at-home phase of the study, even though they were successful and consistent in using it during the game-based training. Our conceptual framework suggests that there are factors that either promote or detract from a user's decision to continue with an expert technique—and so if an expert technique fails to persist, the factors that detract from its continued use have outweighed the factors that promote it. Moreover, in the case of these four participants, this imbalance must have been immediate and substantial, since they abandoned the one-step technique as soon as they started the drawing tasks. Here, we consider specific possibilities why this group of four non-one-step users may have arrived at their (conscious or unconscious) conclusion to stop using the expert technique.

Error aversion. Participants' responses to our post-study survey show that two of the four non-one-step users (P6 and P9), as well as our "in-between" participant (P7), were worried about selecting the wrong item when using the expert method. Incorrect selections are more likely with one-step selection because it is a memory-based technique, and there is no visual feedforward information to confirm that the user is about to make the correct selection. However, these participants' hesitation to use one-step because of the possibility of making an error is somewhat puzzling. First, the FastTap menu was designed to have a very low error cost—none of the commands in the menu were destructive (with the possible exception of Undo), and any incorrect selection could be fixed simply by making an additional selection for the correct item. Second, all but one of the participants—including the non-one-step participants—demonstrated perfect knowledge of the menu's spatial layout during the memory test held at the end of the study. This meant that their hesitation was not based on an actual lack of knowledge about the item locations. It may be that despite participants' strong spatial knowledge and the low cost of making an error, some people are particularly error-averse—and because there are a few examples of user interfaces that are designed to invite guessing as a user selects command functionality, many users may have had prior experiences where making an error was costly. A psychological phenomenon called "negativity bias" may be able to partially explain this behavior (Baumeister et al. 2001; Kahneman 2013)—people often over-estimate the magnitude of a future negative event, such that "losses loom larger than gains" (Kahneman 2013). When faced with the prospect of a potential error, people may shy away from that perceived negative, even if the actual consequences are negligible.

Level of confidence in spatial memory. A related issue is that a user's level of confidence is also strongly related to their willingness to use a memory-based technique. Our post-study survey asked participants to rate how confident they were in their knowledge of each item's location in the menu. All participants scored their confidence at 4/5 or higher, but there was a small (but potentially important) difference between the one-step users and the non-one-step users. People who continued to use the technique all rated themselves at a confidence of 5/5, whereas three of the four non-one-step users rated their confidence as 4/5. This is a small difference, but may suggest that if some users are anything less than completely confident they will be reluctant to use a memory-based method because of the non-zero possibility of making an error (particularly if coupled with negativity bias as described earlier). These users may adopt a stance of "I'm not going to depend on my memory unless I'm certain"—even if their performance shows essentially perfect recollection of the locations. This possibility again points to the potential importance of overlearning as a training strategy (Driskell et al. 1992). The value of continued practice in a memory-based technique, even after mastery has been demonstrated, could have more to do with the benefits to people's confidence in their memory, than it does on the actual strength of the association.

Ergonomics and the physical usage setting. Two of the non-one-step users (P4 and P6) agreed with the statement "It was awkward to hit the menu button and an item at the same time." Although none of the participants appeared to have any difficulty with the ergonomics of the technique during the training session, it is possible that the physical environment where these participants carried out the drawing tasks may have made the one-step method more awkward. For example, as stated in earlier research, the one-step selection technique requires both hands if the device is being held (one to hold the device, the other to perform the selection); similarly, it can be difficult to position the fingers for a multi-touch (one-step) tap when the user holds the device in a two-handed grip; and the thumb-and-finger posture needed for FastTap can be difficult for users with long fingernails (Gutwin et al. 2014). In the training session, users consistently held the device in their left hand, and made selections with the thumb and forefinger of their right hand (all but one participant was right-handed). Although we did not track the specific postures used by our participants during the drawing trials, it is likely that the difference between the physical training environment and the physical usage context was a factor for some participants that detracted from the use of one-step selection. We note, however, that physical environment and device posture are likely to vary considerably across 7 days and 20 different tasks—and since we did not see variation in use of one-step selection for these four participants, it seems unlikely that physical context is the dominant factor in reducing use of the expert method (although in the following we consider the possibility that the first drawing experience was a "tipping point" for some participants).

Satisficing and perception of the distance between contexts. Previous research (reviewed in Section 2) suggests several general reasons why users fail to adopt faster expert techniques. One issue that may have occurred in our study is satisficing (the perception that the novice method is good enough for the tasks at hand). That is, part of the process of using an expert method in the usage context is deciding (consciously or unconsciously) that the expert method is appropriate (e.g., it provides value over some threshold, and has a cost under some other threshold). It is possible that non-one-step users had a different perception of the distances and thresholds in the usage scenarios, and felt that the two-step method was good enough for the drawing tasks. Some evidence for this possibility is seen in the post-study survey, in which two participants agreed with the statement "The two-step method was fast enough" and another agreed with "I wasn't in any hurry." The small sample size for these results, however, means that this issue should be followed up in further studies. Finally, other previously studied reasons for not adopting an expert method seem less likely—for example, the "performance dip" that can occur when users switch to an expert technique (Scarr et al. 2011) is unlikely to have occurred in our study, since users were already able to perform at a high level with the one-step method.

Local optimality and unexpected value of the novice method. Previous work has also introduced the idea that users may continue with a novice technique because it has advantages in a given situation, even if it is slower overall compared to the expert method. Some features of the two-step method may have acted as local optimizers for some of our participants. For example, the two-step method allows exploration of the entire menu, which can help people think about the tools that they will need for the drawing task (a possibility first considered in an earlier study of FastTap menus (Gutwin et al. 2015)). Two of the non-one-step users (P2 and P6) agreed with the statement "Seeing all the items in the menu was useful," which fits with this explanation. However, the fact that four users showed such consistency of two-step use across several types of drawing with different planning requirements suggests that these factors are unlikely to be the overarching reason for the behavior we observed. In particular, the two-step method was unlikely to be optimal for the frequently repeated Cat drawing, because increasing familiarity with the drawing meant that users did not have to carefully plan out the tools that they were going to use.

Another benefit of the two-step selection method was that the menu could be opened once, and then used to select multiple items (because the menu does not close until the thumb is lifted from the menu button). This capability was mentioned by our one variable-use participant (P7), and our analysis showed that it was used by all of the non-one-step participants for between 11% and 19% of menu openings. This may explain use of two-step selection for these instances, but it does not fully explain its use for the other 80–90% of selections.

Differences in intrinsic motivation. Finally, the non-one-step group may simply have been less interested in doing things the expert way, compared to the one-step group. There is some evidence for this possibility in the post-study questionnaire results, where one of the two-step participants agreed with the statement "The two-step method was fast enough" and another agreed with the statement "I wasn't in any hurry." The design of FastTap may have contributed to these attitudes, in that the novice method is also fairly fast, so returning to the novice method did not incur a substantial performance penalty, unlike shortcuts in other interfaces that can be dramatically faster than their mouse-based counterparts. That is, the similarity between the novice and expert forms of FastTap (which was intended to ease the transition from novice to expert) may have also made it easier to revert to using the novice method. This would be an unintended consequence of the design that should be studied further.

Why did we see "all-or-nothing" use of the expert method in Study 3?

The dramatic differences in usage during the at-home drawing tasks were surprising—nine of the ten participants either used the expert method all of the time (five participants) or none of the time (four participants)—and behavior was consistent across several types of tasks and a week of elapsed time. In the earlier sections, we have considered factors that may have promoted or detracted from the use of the expert method overall; here we look specifically at possible reasons why people's behavior was so consistent for all but one of the participants.

It may be that people make an initial decision about which technique to use when they first need to carry out a task in a new usage situation, and once they have made this decision, they continue to follow the chosen path rather than re-evaluating their choice for every task or command invocation. In the at-home setting where people carried out the drawing tasks, either the two-step or the one-step techniques were valid choices that could be used in all task situations (aside from the strategy of making multiple selections for a given two-step menu opening, discussed earlier); therefore, it is possible that our participants simply decided which technique they wished to use, and once this decision had been made it was not re-evaluated. Note that it may be that this decision is made without conscious consideration on the part of the user—they may just take the action that feels the most natural at the time. If this is the case, it may be valuable for integrated training systems to actively prompt the user when they first enter the usage context to "try out" the expert technique, to influence this initial decision.

While the preceding theory could explain our findings, general experience with real-world use of expert methods (such as keyboard shortcuts) suggests that people are not always absolute in their interaction choices, and that they do sometimes use different techniques in different situations—for example, sometimes printing a document using the menu item and sometimes using the "Ctrl + P" hotkey. This suggests that there are situation-specific factors that could influence an individual decision about a technique. For example, if the user's hand is already on the mouse, then there may be a higher likelihood of them choosing to print using the menu rather than the keyboard shortcut. Individual users may have a kind of implicit decision tree of factors, with several situational variables combining to determine whether a user chooses one or another method.

If this was the case for our participants in the drawing phase of the study, the consistent use of the two-step method for four people implies that there were enough negative factors in each usage situation such that the two-step method was always chosen. These factors may all arise from the specific demands of the current task—e.g., our four non-one-step users may have always been slightly unconfident about the item locations, and thus decided in each case to use the two-step method. However, it is also likely that there are factors that operate across multiple tasks—in particular, the technique that was used for the previous selection could exert various kinds of inertia (e.g., by being foremost in memory) that weighs in favor of continued use of that technique. If users do employ a type of implicit decision tree as they choose which technique to use for each task, it is interesting to note that the initial parts of the decision tree for two-step and one-step selection may be very similar. The mechanism for selecting a command (e.g., hand posture and movement) is nearly the same for both methods, which means that any inertia or established procedural memory could reinforce the better-known path.

This reasoning may also provide an explanation for the one participant who did not exhibit the all-or-nothing usage in the drawing tasks. In discovering that he could select several features in one menu opening, this participant may have consciously added an additional path to their decision tree—that is, depending on the number of selections to be made, he may have decided to use either the one-step or the two-step method.

Can existing frameworks of learning and transfer explain our results?

As described earlier (Section 2.2.6), there are several frameworks and theories that have been used to explain various aspects of learning, skill development, and skill transfer. When we consider the two groups of participants in our study (those who persisted with one-step selection and those who did not), it appears that the existing theories can explain the behavior of each of these groups individually, but face more of a challenge when asked to explain why the split occurred within the participants.

In the cases where participants did persist with one-step selection, a variety of theories could be called upon. For example, the ACT-R framework posits that with repeated practice, certain production rules become stronger and thus will be chosen when a user needs to complete a task. In our study, the production rule that was strongly practiced during training was one in which a stimulus (i.e., a particular command) was followed by the response of a one-step selection action. ACT-R would suggest that those participants who persisted with one-step had practiced this production rule sufficiently that it was the prominent production when the at-home tasks were carried out. However, the same reasoning can be used to explain why the two-step users did *not* persist with one-step—that is, they did not practice the one-step response often enough to make this production rule prominent after training. But all participants had very similar experiences during training, and very similar amounts of practice using the one-step technique. The discrepancy in persistence across participants suggests two possibilities—first, that different people require very different amounts of practice before a production rule has been sufficiently rehearsed; second, that there is an additional variable (or variables) in the users' experience, other than the amount of practice, that is influencing persistence.

We can see a similar pattern if we attempt to explain participant performance with the other theories discussed in Section 2.2.6. For example, the microstrategies theory can explain why some participants persisted with one-step selection. This theory states that when there are several ways to carry out an interactive task, users are sensitive to small differences in the efficiency of the different strategies, and will adopt those that are most efficient. In our study, one-step selection was faster than two-step selection (primarily because it requires only one physical action instead of two), and it is reasonable to conclude that participants could perceive this difference and use that in an internal calculation about which strategy to choose during the drawing tasks. However, the microstrategies theory does not explain why some participants would choose to stay with the two-step technique, even though it was substantially slower.

Fu and Gray's work on the "paradox of the active user" provides a contrasting explanation to that of microstrategies. Fu and Gray provide reasons why users may *not* switch to an optimal strategy, even if they know about it, and instead persist with a "stable but suboptimal" strategy. The main factors that Fu and Gray suggest can lead to this situation are a higher generality of the less-efficient technique across other usage situations, and the presence of incremental visual feedback that can reduce the user's cognitive load. In our study, the second of these two factors was present—the two-step method, although less efficient, did provide incremental visual feedback (i.e., the grid menu) that could reduce cognitive load (i.e., people can use recognition rather than recall with the visual menu). This reasonable explanation, however, does not provide an understanding of why the behavior was observed for some participants but not for others.

The strong division in our participants therefore presents a challenge for existing frameworks and models of transfer and strategy choice—although the theories can provide explanations for either the one-step behavior or the two-step behavior, there is little guidance on why some users chose the faster method while others did not. Some of the frameworks recognize that these divisions exist—e.g., Fu and Gray clearly recognize that the "paradox of the active user" only applies to some users, not all—but the frameworks do not give much detail on what leads people to choose different strategies under similar conditions (i.e., after the same training and the same experience with the system, as was the case in our study).

As suggested earlier, there must be additional factors that lead to this division. Several frameworks propose parameters that can account for a change in strategy choice (e.g., frequency of use in ACT-R), but do not specify the values for factors (or combinations of factors) at which a "tipping point" occurs and the user switches to a different method. This is interesting in the context of our study, because half of our users tipped one way, and (almost) half tipped the other way. The details of the cognitive, individual, or environmental factors that lead to these differences do not appear to be well modeled by existing formalisms.

What other factors could be at play here? Some have been identified in previous literature on skills transfer—in particular, individual differences in the specific characteristics of the trainee (see Section 2.2.3) including ability or motivation, and personality factors such as error aversion or negativity bias (Kahneman 2013). It would be interesting from a research perspective if individual differences really did lead to such a strong difference in our participant group. In particular, if strongly divergent results in persistence of expert methods can be caused by factors such as personality type or risk aversion, then it is clear that these factors deserve more attention.

It is possible, of course, that individual differences are not as large a factor as the earlier discussion may suggest—that is, it is possible that the split we observed in our participants is simply an anomaly, and that in most cases people will consistently go either one way or the other. Further studies are required to investigate this possibility, but it is worth mentioning that our study was not extreme in any way—we followed typical practices that could easily be found in other user-interface training programs (e.g., a brief training period, a simple application domain, and a user population that is often used in HCI studies). This suggests that the clear division in our results was less likely to be caused by something odd in our design—and if the division was so easily and so obviously found in our study, it may be common enough to warrant further investigation. Overall, further exploration into the effects of individual and environmental factors on skill transfer is warranted, and may allow existing theories and frameworks to be further generalized, and provide additional explanatory power.

Why did the game training context work well to teach the expert selection method?

The training context used in Study 3 enabled all participants to quickly acquire the required memory and skill and transition to using one-step selection, within ~18 minutes of playing an engaging game. Moreover, for five of the ten participants, this training stuck when transitioning to the usage context of the drawing tasks. We take these results as positive evidence that for some users (potentially even a majority of users), a small amount of game-based training can lead to strong retention in realistic usage contexts. This result also suggests that even fairly large differences between training and usage contexts can be bridged—and with only a small amount of training.

In terms of what exactly made the game training context work well, the results of our studies suggest that the performance requirement it imposed was effective at motivating adoption of the expert technique, and that this adoption persisted once the performance requirement was dropped and other aspects of the context changed. This is encouraging, because performance requirement is relatively easy for designers of training systems to manipulate (e.g., by simply imposing a time restriction on a task), but is difficult to manipulate in usage contexts. The game training context was also engaging and fun to play, sharing many similarities to quick-reaction games available on mobile devices. While participants in our studies were not given a choice of whether or not to use the training system, providing an enjoyable training experience is important in real-world training scenarios, where users would be free to abandon a training system if they found it unpleasant to use. In the sections that follow, we discuss in greater detail the potential for this type of game-based training.

7.2 Generalization to Other Techniques, Applications, and Training Systems

In this article, we have investigated post-training persistence for a particular expert interaction technique, application, and training system. Given this, it is important to consider how our results might generalize to other interaction techniques, software systems, and training contexts.

FastTap menus share many similarities with other memory-based interaction techniques, such as keyboard shortcuts and gestures. In particular, the process of learning all three includes a semantic component (the spatial location, shortcut key, or shape of the gesture) and a motor component (the action of performing a one-step selection, keyboard shortcut, or gesture). Given these similarities, we suspect that our results will generalize to these other types of expert interaction techniques as well. That being said, a factor that might impact the persistence of some of these techniques differently from others is the similarity or difference between the actions to perform novice versus expert selections—in FastTap menus (and marking menus (Kurtenbach 1993)), the novice and expert selections are performed with nearly the same motor action. In contrast, for techniques such as keyboard shortcuts, the motor action of performing a shortcut is very different from that of opening a menu and selecting an item. This difference might increase the likelihood that keyboard shortcuts will exhibit post-training persistence, because backsliding to the novice technique requires a very different motor action. Conversely, if the user spends much of their time in the usage context with one hand on the mouse, this might have an opposite effect, acting as an additional barrier to using the expert technique.

In terms of generalizability to other training systems, a key question is how important the game context is, and whether performance requirement imposed in other ways (e.g., a simple timer) will have the same effect as observed in our studies. This is an important question to answer because some methods of imposing a performance requirement may be easier to implement in a given training context than others. For example, it may be easy to implement training within an application through adding a simple timer, but may be difficult to integrate a more complicated game experience.

Finally, we believe that our results will generalize to other software systems, but also that differences between applications have the potentially affect post-training persistence. For example, the complexity of the application could play a role. In our studies participants were able to train and learn all of the commands available in the drawing application—something that will simply not be possible for feature-rich applications with hundreds or thousands of commands. As a result, people who train on an expert technique will necessarily be faced with using a mix of the novice and expert technique in the usage context. This could impact post-training persistence if the user adopts a "default" action of, for example, opening a menu when they need to perform a command. That said, many users of desktop software currently use a mixture of keyboard shortcuts and menus to invoke commands, and prior research has shown that users are able to employ subtly different "microstrategies" in how they interact with an interface, even within a fixed task context (Gray and Boehm-Davis 2000). Overall, investigating post-training persistence in the scenario where participants only learn the expert technique for some commands is an interesting area for future work. In the next section, we discuss additional possibilities for future work on training systems for more complex software systems.

7.3 Applications to Real-World Training

The overall goal of this research is to improve users' performance and expertise with interactive systems—and there are several positive results from our studies that suggest a game-based training approach can be successful for this. Although additional studies are clearly needed to expand on our results and explore the bimodal behavior seen in Study 3, we believe that the game-based training approach investigated in this article could be generalized to real-world training contexts relatively easily. In this section, we describe one potential application of this idea.

While spatial menus of the type studied in this article are not common in commercial applications, keyboard shortcuts are nearly universally offered in feature-rich software. For example, photo-editing applications such as Photoshop and the GNU Image Manipulation Program (GIMP) both offer extensive sets of keyboard shortcuts (see Figure 16 for a summary of those offered in GIMP), and the ability to effectively use these shortcuts is a hallmark of experienced users of the application. Given that keyboard shortcuts are also a memory-based expert technique, and share many similarities with spatial menus, we believe that it would be straightforward to adapt our game-based training approach to train GIMP or Photoshop users on keyboard shortcuts.

We foresee two main challenges in adapting the game-based training approach used in this article to this new application domain. First, there are many more keyboard shortcuts in applications like GIMP or Photoshop than the 12 commands of our drawing tool. There are two ways that we could adapt our approach to accommodate this difference. First, training on a subset of the keyboard shortcuts is a worthy initial goal. Prior studies of feature-rich applications have shown that individual users typically only use a small subset of the available functionality (Lafreniere et al. 2010; Linton et al. 2000), and "top ten shortcuts" lists are widely available for popular applications (e.g., http://digital-photography-school.com/10-best-photoshop-shortcuts/). This suggests that even limiting the training vocabulary to 12 items could make a substantial difference to people's performance. Second, we believe that the training approach used in this article could easily be expanded to train users on more items. The primary limitation on the number of items for a FastTap menu is the size of the grid—but this is not a constraint when learning keyboard shortcuts. Therefore, users could learn one subset of commands at first (e.g., a "top ten most useful" set) and then learn new subsets based on either their personal tasks or their own interests.

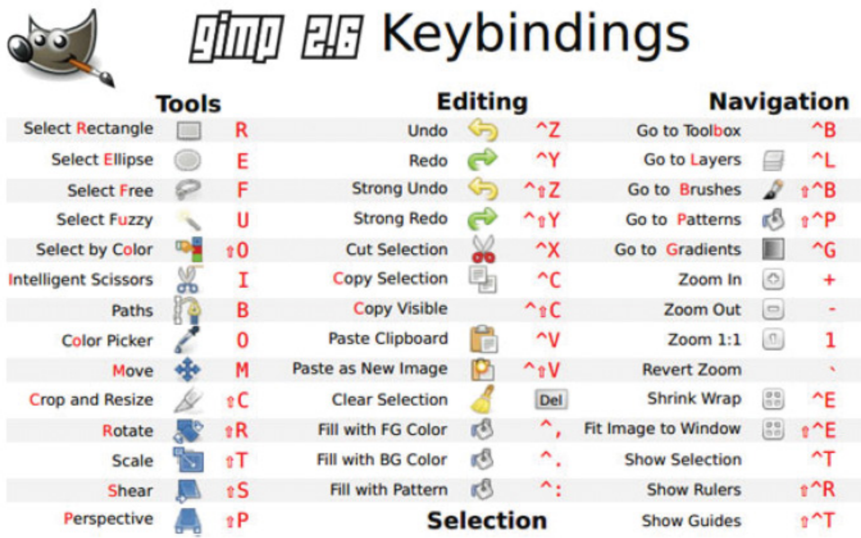

Fig. 16. Summary of GIMP keyboard shortcuts (from https://jeffkayser.com/projects/cheatsheet-gimp/, copyright Jeff Kayser).

A second challenge with adapting the game-based approach concerns how to present the keyboard shortcuts to the user during the initial stages of the game. While it was easy to open the FastTap menu in our studies and see all available commands, the commands in an application such as Photoshop are spread over a large number of toolbars and hierarchical menus, and it could be frustrating for a user to search through the interface to learn a command's keyboard shortcut. We see two possibilities for improving presentation. First, we could use the ExposeHotKey (EHK) strategy proposed by Malacria et al. (2013), which displays all hotkeys in an interface when the user presses a control key. Second, we could present a subset of hotkeys in a "cheat-sheet" style, where the shortcuts that are currently being learned are shown on the screen with their command names (e.g., a simplified version of the sheet shown in Figure 16). This latter approach would reduce the amount of search needed to find the correct shortcut. Researchers have also suggested hybrid approaches that highlight subsets of commands within the actual interface (Scarr et al. 2015).

We are currently building this type of training system for the open-source Inkscape vector-drawing software. Research questions that we will consider with these systems include whether keyboard shortcuts are learned as quickly as FastTap spatial locations, whether learning with the abstract cue of the game (i.e., retrieval of the hotkey is not set within a realistic task context) translates to real-world use, and whether the "all-or-nothing" phenomenon occurs—particularly because factors such as the users' hand location at the start of the task (keyboard or mouse) may have a large effect on which method users choose to use, and are more variable in applications that use the keyboard and mouse as input, as compared to the touchscreen setting focused on in this article.

7.4 Limitations

Our study results provide new insights into the post-training persistence of expert selection techniques, and reveal just how complex this phenomenon is. In particular, our findings suggest that a range of different factors may influence whether or not training in an expert technique will transfer to a usage context. Given this complexity, there are a number of potential limitations to our studies that are important to acknowledge.

First, the experimental setting may have imposed a different set of motivations on participants as compared to a real-world setting. For example, in Studies 1 and 2, participants may have inferred that we were studying adoption of the expert technique, and this may have affected their decision of whether or not to continue to use it once the performance requirement dropped. Likewise, in Study 3, participants may have felt that they were expected to maintain use of the expert technique, or may have simply thought more about their choice of which technique to use than they would in a real-world setting. While these extrinsic motivations imposed by the experimental setting are important to consider, we believe that a similar set of intrinsic motivations are likely to be present in real-world settings—people who choose to undergo training are probably doing so because they want to learn and use the expert technique.

Second, while we attempted to design the drawing tasks for Study 3 to cover a range of potential types of tasks (including routine tasks, tasks that require planning, and creative tasks), additional factors that have the potential to affect persistence may exist in real-world training scenarios. As one example, we did not test the scenario in which a user is already familiar and well-practiced with using a novice technique in an application, and then undergoes training for an expert technique. In this scenario, it may be more difficult for the expert technique to persist because the novice technique is associated with well-practiced or routine tasks the user performs in the application. Identifying and testing additional factors that may affect post-training persistence is an interesting area for future work.

7.5 Future Work

Drawing together the discussion points raised earlier, we see a number of directions for future work investigating the post-training persistence of expert interaction techniques. In the next few paragraphs, we outline some of these directions, as well as specific studies that could be conducted.

First, follow-up studies with a similar design to Study 3 could be used to test some of the hypotheses raised in our preceding discussion. For example, it would be valuable to test whether participants' confidence in their memory of menu items affects their use of the expert technique in a usage context—this could be achieved through a short memory test and subjective confidence rating before the first drawing task. Along similar lines, it would be interesting to test whether longer or shorter gaps between the training game and the start of drawing tasks leads to more or less decay in the use of the expert technique, or whether occasional "refresher" rounds of the training game, interleaved with drawing tasks, can supplement an intensive initial training session. Finally, we could probe more deeply into participants' attitudes toward expert techniques, and the physical setting or postures with which participants held the device when completing the drawing tasks, to better understand what is leading to the expert technique to persist or not.

Second, it would be valuable to test whether an initial choice to use a technique has an inertia or momentum that carries forward over multiple tasks or sessions. This could be achieved by periodically forcing the user to use either the novice or expert technique, then observing whether the forced technique continues to be used for subsequent drawing tasks or sessions.

Third, it would be valuable to investigate how strongly our findings generalize to different interaction techniques, training methods, and application types. In particular, it would be valuable to study the post-training persistence of keyboard shortcuts, and the effect that the user's hand position (e.g., resting on the keyboard, and one hand on the mouse) has on their choice of interaction technique.

Finally, perhaps the most important area for future work is to attempt to replicate our study in a full-featured real-world application, such as the proposed study of Inkscape described earlier. This would reveal the potential for this type of training for commercial applications, and help us to understand additional factors affecting post-training persistence for the types of expert interaction techniques that occur in real-world applications.

8 CONCLUSIONS

In this article, we have introduced the concept of *post-training persistence* for expert interaction techniques—that is, whether or not training in these techniques effectively transfers to a usage context and displaces available novice interaction techniques. As an initial step to investigate this phenomenon, we presented a framework that defines a range of potential factors that may affect whether an expert technique will persist, and conducted three studies to gain insights into a selection of these factors. Our findings suggest that performance requirement is a factor that can be applied in a training context to encourage adoption of an expert technique, but does not need to be maintained in the usage context for use of the expert technique to continue. We also observed that, upon entering a usage context after training, users quickly settle on either using the expert technique (i.e., the expert technique persists) or return to the novice technique, and that this choice is stable over subsequent days of usage sessions and a range of qualitatively different kinds of tasks. Finally, based on the findings from our three studies, we have identified a rich set of open research questions surrounding the post-training persistence of expert interaction techniques, as well as design implications for training systems, and ideas for how the game-based training approach we explored could be applied to feature-rich software applications used every day by millions of users.

REFERENCES

Jack A. Adams. 1987. Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychol. Bull.* 101, 1 (1987), 41–74. DOI: https://doi.org/10.1037/0033-2909.101.1.41

Fraser Anderson and Walter F. Bischof. 2013. Learning and performance with gesture guides. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. New York, NY: ACM, 1109–1118. DOI: https://doi.org/10.1145/2470654.2466143

John R. Anderson. 1983. *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Michelle Annett and Walter F. Bischof. 2013. Your left hand can do it too! Investigating intermanual, symmetric gesture transfer on touchscreens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. New York, NY: ACM, 1119–1128. DOI: https://doi.org/10.1145/2470654.2466144

Caroline Appert and Shumin Zhai. 2009. Using strokes as command shortcuts: Cognitive benefits and toolkit support. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*. New York, NY: ACM, 2289–2298. DOI: https://doi.org/10.1145/1518701.1519052

Winfred Arthur, Winston Bennett, Pamela L. Stanush, and Theresa L. McNelly. 1998. Factors that influence skill decay and retention: A quantitative review and analysis. *Hum. Perform.* 11, 1 (March 1998), 57–101. DOI: https://doi.org/10.1207/s15327043hup1101_3

Timothy T. Baldwin and J. Kevin Ford. 1988. Transfer of training: A review and directions for future research. *Pers. Psychol.* 41, 1 (March 1988), 63–105. DOI: https://doi.org/10.1111/j.1744-6570.1988.tb00632.x

Scott Bateman, Jaime Teevan, and Ryen W. White. 2012. The search dashboard: How reflection and comparison impact search behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*. New York, NY: ACM, 1785–1794. DOI: https://doi.org/10.1145/2207676.2208311

Roy F. Baumeister, Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs. 2001. Bad is stronger than good. *Rev. Gen. Psychol.* 5, 4 (2001), 323–370. DOI: https://doi.org/10.1037/1089-2680.5.4.323

Howard J. Baumgartel, Jill I. Reynolds, and Rukhsana Z. Pathan. 1984. How personality and organisational climate variables moderate the effectiveness of management development programmes: A review and some recent research findings. *Manag. Labour Stud.* 9, 1 (1984), 1–16.

Suresh K. Bhavnani and Bonnie E. John. 2000. The strategic use of complex computer systems. *Hum.-Comput. Interact.* 15, 2 (September 2000), 107–137. DOI: https://doi.org/10.1207/S15327051HCI1523_3

Mary L. Broad and John W. Newstrom. 1992. *Transfer of Training: Action-Packed Strategies to Ensure High Payoff from Training Investments*. Addison-Wesley Publishing Company.

William A. S. Buxton. 1995. Chunking and phrasing and the design of human-computer dialogues. In. *Human-Computer Interaction*. Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg (Eds.), San Francisco, CA: Morgan Kaufmann Publishers, Inc. 494–499.

Stuart K. Card, Allen Newell, and Thomas P. Moran. 1983. *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: L. Erlbaum Associates Inc.

John M. Carroll and Mary Beth Rosson. 1987. Paradox of the active user. In *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. MIT Press, 80–111.

Nicholas J. Cepeda, Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. 2006. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychol. Bull.* 132, 3 (2006), 354–380. DOI: https://doi.org/10.1037/0033-2909.132.3.354

Andy Cockburn, Carl Gutwin, and Saul Greenberg. 2007a. A predictive model of menu performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. New York, NY: ACM, 627–636. DOI: https://doi.org/10.1145/1240624.1240723

Andy Cockburn, Per Ola Kristensson, Jason Alexander, and Shumin Zhai. 2007b. Hard lessons: Effort-inducing interfaces benefit spatial learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. New York, NY: ACM, 1571–1580. DOI: https://doi.org/10.1145/1240624.1240863

Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. Supporting novice to expert transitions in user interfaces. *ACM Comput. Surv.* 47, 2 (November 2014), 31:1–31:36. DOI: https://doi.org/10.1145/2659796

Deborah Compeau, Lorne Olfman, Maung Sei, and Jane Webster. 1995. End-user training and learning. *Commun. ACM* 38, 7 (July 1995), 24–26. DOI: https://doi.org/10.1145/213859.214791

Fergus I. M. Craik and Robert S. Lockhart. 1972. Levels of processing: A framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 6 (December 1972), 671–684. DOI: https://doi.org/10.1016/S0022-5371(72)80001-X

E. R. F. W. Crossman. 1959. A theory of the acquisition of speed-skill. *Ergonomics* 2, 2 (February 1959), 153–166. DOI: https://doi.org/10.1080/00140135908930419

James E. Driskell, Ruth P. Willis, and Carolyn Copper. 1992. Effect of overlearning on retention. *J. Appl. Psychol.* 77, 5 (1992), 615–622. DOI: https://doi.org/10.1037/0021-9010.77.5.615

Hermann Ebbinghaus. 1913. *Memory: A Contribution to Experimental Psychology*. University Microfilms.

K. Anders Ericsson, Ralf Th. Krampe, and Clemens Tesch-Romer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* (1993), 363–406.

Paul Morris Fitts and Michael I. Posner. 1967. *Human Performance*. Brooks/Cole Publishing Company.

J. Kevin Ford, Miguel A. Quiñones, Douglas J. Sego, and Joann Speer Sorra. 1992. Factors affecting the opportunity to perform trained tasks on the job. *Pers. Psychol.* 45, 3 (September 1992), 511–527. DOI: https://doi.org/10.1111/j.1744-6570.1992.tb00858.x

J. Kevin Ford and Daniel A. Weissbein. 1997. Transfer of training: An updated review and analysis. *Perform. Improv. Q.* 10, 2 (June 1997), 22–41. DOI: https://doi.org/10.1111/j.1937-8327.1997.tb00047.x

Marguerite Foxon. 1993. A process approach to the transfer of training. *Australas. J. Educ. Technol.* 9, 2 (December 1993). DOI: https://doi.org/10.14742/ajet.2104

Wai-Tat Fu and Wayne D. Gray. 2004. Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cogn. Sci.* 28, 6 (November 2004), 901–935. DOI: https://doi.org/10.1016/j.cogsci.2004.03.005

A. M. Gentile. 1972. A working model of skill acquisition with application to teaching. *Quest* 17, 1 (January 1972), 3–23. DOI: https://doi.org/10.1080/00336297.1972.10519717

Marilyn E. Gist, Anna G. Bavetta, and Cynthia Kay Stevens. 1990. Transfer training method: Its influence on skill generalization, skill repetition, and performance level. *Pers. Psychol.* 43, 3 (September 1990), 501–523. DOI: https://doi.org/10.1111/j.1744-6570.1990.tb02394.x

Wayne D. Gray and Deborah A. Boehm-Davis. 2000. Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *J. Exp. Psychol. Appl.* 6, 4 (2000), 322–335. DOI: https://doi.org/10.1037/1076-898X.6.4.322

Wayne D. Gray and John K. Lindstedt. 2016. Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance. *Cogn. Sci.* (October 2016). DOI: https://doi.org/10.1111/cogs.12412

Wayne D. Gray, Chris R. Sims, Wai-Tat Fu, and Michael J. Schoelles. 2006. The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychol. Rev.* 113, 3 (July 2006), 461–482. DOI: https://doi.org/10.1037/0033-295X.113.3.461

Nathan Green, Jan Kruger, Chirag Faldu, and Robert St. Amant. 2004. A reduced QWERTY keyboard for mobile text entry. In *Proceedings of the CHI'04 Extended Abstracts on Human Factors in Computing Systems (CHI EA'04)*. New York, NY: ACM, 1429–1432. DOI: https://doi.org/10.1145/985921.986082

Tovi Grossman, Pierre Dragicevic, and Ravin Balakrishnan. 2007. Strategies for accelerating on-line learning of hotkeys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. New York, NY: ACM, 1591–1600. DOI: https://doi.org/10.1145/1240624.1240865

Carl Gutwin, Andy Cockburn, and Benjamin Lafreniere. 2015. Testing the rehearsal hypothesis with two FastTap interfaces. In *Proceedings of Graphics Interface (GI'15)*. Canadian Information Processing Society.

Carl Gutwin, Andy Cockburn, Joey Scarr, Sylvain Malacria, and Scott C. Olson. 2014. Faster command selection on tablets with FastTap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'14)*. New York, NY: ACM, 2617–2626. DOI: https://doi.org/10.1145/2556288.2557136

Jeff Hendy, Kellogg S. Booth, and Joanna McGrenere. 2010. Graphically enhanced keyboard accelerators for GUIs. In *Proceedings of Graphics Interface (GI'10). Toronto, ON, Canada: Canadian Information Processing Society*, 3–10.

R. J. Herrnstein, George F. Loewenstein, Drazen Prelec, and William Vaughan. 1993. Utility maximization and melioration: Internalities in individual choice. *J. Behav. Decis. Mak.* 6, 3 (September 1993), 149–185. DOI: https://doi.org/10.1002/bdm.3960060302

Poika Isokoski and Roope Raisamo. 2000. Device independent text input: A rationale and an example. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI'00)*. New York, NY: ACM, 76–83. DOI: https://doi.org/10.1145/345513.345262

Daniel Kahneman. 2013. *Thinking, Fast and Slow*. Anchor Canada.

Brian Krisler and Richard Alterman. 2008. Training towards mastery: Overcoming the active user paradox. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges (NordiCHI'08)*. New York, NY: ACM, 239–248. DOI: https://doi.org/10.1145/1463160.1463186

Per Ola Kristensson and Shumin Zhai. 2007. Learning shape writing by game playing. In *Proceedings of the CHI'07 Extended Abstracts on Human Factors in Computing Systems (CHI EA'07)*. New York, NY: ACM, 1971–1976. DOI: https://doi.org/10.1145/1240866.1240934

Gordon Kurtenbach and William Buxton. 1991. Issues in combining marking and direct manipulation techniques. In *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology (UIST'91)*. New York, NY: ACM, 137–144. DOI: https://doi.org/10.1145/120782.120797

Gordon Kurtenbach and William Buxton. 1994. User learning and performance with marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'94)*. New York, NY: ACM, 258–264. DOI: https://doi.org/10.1145/191666.191759

Gordon Paul Kurtenbach. 1993. *The Design and Evaluation of Marking Menus*. Ph.D. Dissertation. University of Toronto, Toronto, ON, Canada.

Benjamin Lafreniere, Andrea Bunt, John Whissell, Charles L. A. Clarke, and Michael Terry. 2010. Characterizing large-scale use of a direct manipulation application in the wild. In *Proceedings of Graphics Interface (GI'10)*. Canadian Information Processing Society, 11–18.

Richard N. Landers and Michael B. Armstrong. 2017. Enhancing instructional outcomes with gamification: An empirical test of the technology-enhanced training effectiveness model. *Comput. Hum. Behav.* 71, C (2017), 499–507. DOI: https://doi.org/10.1016/j.chb.2015.07.031

David M. Lane, H. Albert Napier, S. Camille Peres, and Aniko Sandor. 2005. Hidden costs of graphical user interfaces: Failure to make the transition from menus and icon toolbars to keyboard shortcuts. *Int. J. Hum.-Comput. Interact.* 18, 2 (May 2005), 133–144. DOI: https://doi.org/10.1207/s15327590ijhc1802_1

Frank Linton, Deborah Joy, Hans-Peter Schaefer, and Andrew Charron. 2000. OWL: A recommender system for organization-wide learning. *Educ. Technol. Soc.* 3, 1 (2000), 62–76.

A. M. Machin. 2002. Planning, managing, and optimizing transfer of training. In *Creating, Implementing, and Managing Effective Training and Development*. San Francisco, CA: John Wiley & Sons, Inc., 263–301.

Sylvain Malacria, Gilles Bailly, Joel Harrison, Andy Cockburn, and Carl Gutwin. 2013a. Promoting hotkey use through rehearsal with exposeHK. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. New York, NY: ACM, 573–582. DOI: https://doi.org/10.1145/2470654.2470735

Sylvain Malacria, Joey Scarr, Andy Cockburn, Carl Gutwin, and Tovi Grossman. 2013b. Skillometers: Reflective widgets that motivate and help users to improve performance. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST'13)*. New York, NY: ACM, 321–330. DOI: https://doi.org/10.1145/2501988.2501996

Edgar Matias, I. Scott MacKenzie, and William Buxton. 1996. One-handed touch typing on a QWERTY keyboard. *Hum.-Comput. Interact.* 11, 1 (March 1996), 1–27. DOI: https://doi.org/10.1207/s15327051hci1101_1

Jane McGonigal. 2011. *Reality is Broken: Why Games Make Us Better and How They can Change the World* (Reprint ed.). New York: Penguin Books.

Allen Newell. 1990. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Daniel L. Odell, Richard C. Davis, Andrew Smith, and Paul K. Wright. 2004. Toolglasses, marking menus, and hotkeys: A comparison of one and two-handed command selection techniques. In *Proceedings of Graphics Interface (GI'04)*. Waterloo, ON, Canada: Canadian Human-Computer Communications Society, 17–24.

Thomas J. Peters and Robert H. Waterman Jr. 1982. *In Search of Excellence: Lessons from America's Best-Run Companies*, New York: HarperCollins Canada/Non-Fiction.

Philip Quinn. 2016. *Economic Behaviour and Psychological Biases in Human–Computer Interaction*. Ph.D. Dissertation. University of Canterbury.

Doug Rohrer, Kelli Taylor, Harold Pashler, John T. Wixted, and Nicholas J. Cepeda. 2005. The effect of overlearning on long-term retention. *Appl. Cogn. Psychol.* 19, 3 (April 2005), 361–374. DOI: https://doi.org/10.1002/acp.1083

Janice Z. Rouiller and Irwin L. Goldstein. 1993. The relationship between organizational transfer climate and positive transfer of training. *Hum. Resour. Dev. Q.* 4, 4 (December 1993), 377–390. DOI: https://doi.org/10.1002/hrdq.3920040408

Eduardo Salas and Janis A. Cannon-Bowers. 2001. The science of training: A decade of progress. *Annu. Rev. Psychol.* 52, 1 (February 2001), 471–499. DOI: https://doi.org/10.1146/annurev.psych.52.1.471

Joey Scarr, Andy Cockburn, Carl Gutwin, and Andrea Bunt. 2012. Improving command selection with CommandMaps. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*. New York, NY: ACM, 257–266. DOI: https://doi.org/10.1145/2207676.2207713

Joey Scarr, Andy Cockburn, Carl Gutwin, and Philip Quinn. 2011. Dips and ceilings: Understanding and supporting transitions to expertise in user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. New York, NY: ACM, 2741–2750. DOI: https://doi.org/10.1145/1978942.1979348

Joey Scarr, Carl Gutwin, Andy Cockburn, and Andrea Bunt. 2015. StencilMaps and EphemeralMaps: Spatially stable interfaces that highlight command subsets. *Behav. Inf. Technol.* 34, 11 (November 2015), 1092–1106. DOI: https://doi.org/10.1080/0144929X.2015.1046927

Richard A. Schmidt. 1991. Frequent augmented feedback can degrade learning: Evidence and interpretations. In *Tutorials in Motor Neuroscience*. Jean Requin and George E. Stelmach (Eds.), NATO ASI Series, Springer Netherlands, 59–75.

Richard Schmidt and Tim Lee. 2011. *Motor Control and Learning: A Behavioral Emphasis* (5th ed.). Champaign, IL: Human Kinetics.

Barry Schwartz, Ed Wasserman, and Steven J. Robbins. 2001. *Psychology of Learning and Behaviour* (5th ed.). New York: W. W. Norton & Company.

Chia Shen, Kathy Ryall, Clifton Forlines, Alan Esenther, Frederic D. Vernier, Katherine Everitt, Mike Wu, Daniel Wigdor, Meredith Ringel Morris, Mark Hancock, and Edward Tse. 2006. Informing the design of direct-touch tabletops. *IEEE Comput. Graph. Appl.* 26, 5 (September 2006), 36–46. DOI: https://doi.org/10.1109/MCG.2006.109

Mark K. Singley and John R. Anderson. 1989. *The Transfer of Cognitive Skill.* Cambridge, MA: Harvard University Press.

Josef H. Tubiana and Gershon Ben-Shakhar. 1982. An objective group questionnaire as a substitute for a personal interview in the prediction of success in military training in Israel. *Pers. Psychol.* 35, 2 (1982), 349–357. DOI: https://doi.org/10.1111/j.1744-6570.1982.tb02200.x

Endel Tulving and Donald M. Thomson. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* 80, 5 (1973), 352–373. DOI: https://doi.org/10.1037/h0020071

William Vosburg. 2000. *Factors Affecting the Transfer of Training from the Classroom to the Worksite.* Ph.D. Dissertation. Vanderbilt University.

Tsung juang Wang. 2009. Educational benefits of multimedia skills training. *TechTrends* 54, 1 (December 2009), 47–57. DOI: https://doi.org/10.1007/s11528-009-0363-x

Xiaoqian Wang. 2010. *Factors Influencing Knowledge and Skill Decay in Organizational Training: A Meta-Analysis.* Ph.D. Dissertation. University of Oklahoma.

Kenneth N. Wexley and Timothy T. Baldwin. 1986. Posttraining strategies for facilitating positive transfer: An empirical exploration. *Acad. Manage. J.* 29, 3 (1986), 503–520. DOI: https://doi.org/10.2307/256221

Kenneth N. Wexley and Gary P. Latham. 2001. *Developing and Training Human Resources in Organizations* (3rd ed.). Upper Saddle River, NJ: Pearson.

R. S. Woodworth and E. L. Thorndike. 1901. The influence of improvement in one mental function upon the efficiency of other functions (I). *Psychol. Rev.* 8, 3 (1901), 247–261. DOI: https://doi.org/http://psycnet.apa.org/doi/10.1037/h0074898

Mike Wu, Chia Shen, Kathy Ryall, Clifton Forlines, and Ravin Balakrishnan. 2006. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *Proceedings of the 1st IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'06)*. Washington, DC: IEEE Computer Society, 185–192. DOI: https://doi.org/10.1109/TABLETOP.2006.19

Eldad Yechiam, Ido Erev, Vered Yehene, and Daniel Gopher. 2003. Melioration and the transition from touch-typing training to everyday use. *Hum. Factors* 45, 4 (Winter 2003), 671–684.

Shumin Zhai and Per-Ola Kristensson. 2003. Shorthand writing on stylus keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'03)*. New York, NY: ACM, 97–104. DOI: https://doi.org/10.1145/642611.642630

Shumin Zhai, Alison Sue, and Johnny Accot. 2002. Movement model, hits distribution and learning in virtual keyboarding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'02)*. New York, NY: ACM, 17–24. DOI: https://doi.org/10.1145/503376.503381

Received December 2016; revised June 2017; accepted June 2017