STRATEGIES IN SKILL ACQUISITION: RECONCILING CONTINUOUS MODELS OF THE LEARNING CURVE WITH ABRUPT STRATEGY SHIFTS

JEROMY KYRAM ANGLIM

Submitted in total fulfilment of the
requirements of the degree of Doctor of Philosophy

March 2011

PSYCHOLOGICAL SCIENCES
THE UNIVERSITY OF MELBOURNE

Abstract

How does task completion time change with practice and what processes underlie this change? Despite over 100 years of scientific research (e.g., Bryan & Harter, 1899) no wholly satisfactory answer has yet emerged. After analysing many skill acquisition datasets Newell and Rosenbloom (1981) influentially declared that the relationship between practice and task completion time was best represented by a power function labelling the relationship the *Power Law of Practice*. Use of the term 'law' might suggest that the case was closed, yet several recent findings have challenged the Power Law's 'legal' status. First, Heathcote, Brown, and Mewhort (2000) concluded that the Power Law of Practice was an artefact of aggregation. Across a large number of skill acquisition datasets they showed that when analysed at the individual-level the exponential function tended to provide superior fit. Second, several researchers have suggested that strategy shifts may even cause discontinuities in the learning curve (e.g., Delaney, Reder, Staszewski, & Ritter, 1998; Haider & Frensch, 2002; Rickard, 1997). In summary, findings suggest that the power function is an analytical artefact and that learning may in some instances involve discontinuities.

In response to these challenges, this thesis had three aims. The first aim was to develop and test mathematical models of the relationship between practice, strategy use and performance. The second aim was to assess the role of individual differences, including prior experience, ability,

and personality in predicting strategy use and performance. The third aim was to model the differential effects of instructed versus self-initiated strategy shift on strategy use and performance. Collectively, the aims were designed to provide a multifaceted explanation of the relationship between practice, strategy use, and performance.

To achieve these aims three studies were conducted. In each study participants completed a set of trials on a text editing task. On each trial strategy sophistication and task completion time were measured. Final sample sizes in the three studies were $n_1 = 63$, $n_2 = 154$, and $n_3 = 154$. Each study also measured a selection of individual difference variables including prior experience, demographics, personality, and ability. Text editing was chosen as the criterion task because strategy use is important to task performance and strategy use could readily be measured. The text editing task was developed to enable trial-level measurement of strategy use. Strategy sophistication was operationalised as the proportion of key presses used that were classified as sophisticated (e.g., using control and right cursor keys to move between words) as opposed to simple (e.g., using just the right cursor to move between characters). In Studies 1 and 2 all participants received the same instructions. In Study 3 participants were randomly assigned to one of three conditions with varying instructions. In a No Training condition practice preceded without interruption, in a Training condition additional instructions were presented halfway through practice, and in a Control condition a filler task was presented halfway through practice. Aims 1 and 2 were assessed by Study 1 and 2 and the No Training condition of Study 3. Aim 3 was assessed by comparing the conditions in Study 3.

With regard to Aim 1 results from the three studies told a consistent story. Results reiterated the importance of analysing data at the individual-level. While at the group-level, a three parameter power function provided superior fit, at the individual-level a three parameter exponential function was significantly better in two out of three studies. Similarly, at the group-level, strategy sophistication was a continuously increasing, monotonically decelerating function of practice, well modelled by a three parameter Michaelis–Menten function. In contrast, at the individual-level, the functional form of the relationship varied dramatically between individuals with a variety of often discontinuous functions providing good fit.

Although abrupt strategy shifts did occur, meaningful discontinuities in the relationship between practice and task completion time were rare. Findings supported a model that explained how abrupt strategy shifts can co-occur with continuous learning curves. These findings were that: (a) strategy shifts were more likely to occur early in practice when other learning was occurring; (b) trial-to-trial variance in task completion time was often large relative to the benefits of the strategy shift; and (c) strategy shifts often took several trials to be fully realised. These and other factors combined to generally smooth out the discontinuous effects of strategy shift on performance.

In relation to the second aim, concerning individual differences, ability and prior experience consistently emerged as moderate to strong predictors of task performance, whereas self-reported Big 5 personality was unrelated to task performance. Similar but generally weaker relationships were found between individual differences and strategy sophistication. A model that proposed that the effect of ability and prior experience on task performance was mediated by strategy sophistication was not supported. Findings were broadly consistent with cognitive correlates and skill transfer models of individual differences.

In relation to the third aim, looking at differences between instructed strategy shift and self-initiated strategy shift, hypotheses were partially supported. In summary, relative to self-initiated strategy shifts, instructed strategy shifts were more abrupt. Performance also tended to decline sharply immediately following the instructed strategy shift. After additional practice, performance was similar to groups that had not received instructed strategy shift. The study highlighted how the dynamics of instructed strategy shift differ from self-initiated strategy shift with regards to discontinuities.

Taken together the results tell an interconnected story regarding the relationship between practice, strategy use, and performance. This thesis contributes to skill acquisition research through a unique combination of features including trial-level measurement of strategy use, individual-level modelling, and the use of nonlinear and discontinuous functions. It is hoped that future research will build on this approach using other samples, tasks, and contexts.

Declaration

I, Jeromy Anglim, certify that

(i) this thesis comprises only my original work towards the PhD except where indicated in the Preface*,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) this thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signature: ______________________________

Date: ______________________________

Preface

The dataset presented as Study 1 of this thesis comes from a collaboration between Niloufar Mahdavi, Janice Langan-Fox (my own principal supervisor), and myself. The study formed part of Niloufar Mahdavi's honours thesis who was supervised by Janice Langan-Fox. The overall design in terms of measurement of abilities and use of a criterion text editing task was based on a series of studies that Janice Langan-Fox had completed with various research students in the preceding years. Niloufar Mahdavi's use of the dataset was distinct from my use. I was involved in some design decisions, most notably the measurement of strategy use, the choice of text editing keys, and the passage of text to be edited. I programmed the experimental task, designing how strategy use and performance data were recorded and setting out the algorithms by which strategy was extracted from key logs. All analyses presented in this thesis were conducted from raw data and represent my own work.

Initial analyses of Study 1 were presented in the following conference proceedings

- Anglim, J., Langan-Fox, J., & Mahdavi, N. (2005). Modeling the Relationship between Strategies, Abilities and Skilled Performance. *CogSci 2005, 27th Annual Meeting of the Cognitive Science Society*, July 21–23 Stresa, Italy.

Acknowledgements

I would like to thank:

- Janice Langan-Fox, my principal supervisor, for inspiring me to pursue a PhD, keeping me on track, setting high standards, and guiding me during the critical moments.
- Alex Wearing for his wisdom, support, and encouragement.
- Yoshi Kashima for his collegiality, suggestions, and encouragement.
- Richard Bell for his helpful suggestions and encouragement.
- Niloufar Mahdavi for kindly and professionally collaborating with me.
- Philip Smith, Paul Dudgeon, Murray Aitkin, and Garry Robins for their instruction and support.
- Phillip Ackerman and Ruth Kanfer for sharing some of their raw data with me. Although this data was not incorporated into the final thesis, it helped me develop ideas about modelling learning curves at the individual-level.
- Richard Moulding for reading an early draft and providing useful suggestions.
- The many individuals who developed the various open source tools—R, LaTeX, Sweave, git, make, bibtex, and more—used in the creation of this thesis.
- The participants for giving up their time to be involved in my research.
- Corrine for her love and support, the way that she listened, her gentle encouragement, and her practical advice.

- Jasmine for her love and support over the course of my life, helping me get over the hurdles and continue my journey in psychology, and reminding me from time to time of John C. King's prophetic words of encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Contents

1.1 Research Problems and Aims

The functional form of the learning curve and the mechanisms that cause it have been the subject of over 100 years of scientific research (e.g., Bahrick, Fitts, & Briggs, 1957; Bryan & Harter, 1899; Heathcote et al., 2000; Lane, 1987; Newell & Rosenbloom, 1981; Snoddy, 1926). Despite this long history, several fundamental gaps in understanding remain. This thesis aims to contribute empirical and theoretical knowledge with the intent of closing gaps in understanding regarding three interrelated topics: (a) the relationships between practice, strategy use, and performance, (b) individual differences in strategy use and performance, and

(c) the effect of instructed versus self-initiated strategy shifts on strategy use and performance.

This chapter first introduces the motivating problem behind each of these topics culminating in three main aims. This is followed by an outline of the scope and method used to achieve these aims. Finally, the structure of the thesis is set out.

1.1.1 Practice, Strategy, and Performance

In recent years researchers modelling the learning curve in psychology have focused on three types of functions: *power*, *exponential*, and *discontinuous*. Newell and Rosenbloom (1981) analysed the relationship between practice and task completion time in existing datasets across a range of tasks, often at the group-level. They concluded that the power function provided such a consistently good fit that the relationship should be labelled the *Power Law of Practice*. Many influential researchers subsequently accepted the Power Law (Blessing & Anderson, 1996; F. J. Lee & Anderson, 2001; Logan, 1992, 1988) often considering it a requirement for any formal model of learning.

More recently, Heathcote et al. (2000) suggested that the conclusion drawn by Newell and Rosenbloom (1981) was an artefact of the long-known (e.g., Estes, 1956; Lane, 1987), but often ignored, biasing effects of group-level analysis. When Heathcote et al. (2000) modelled data at the individual-level (i.e., the psychologically meaningful level) the exponential function tended to provide superior fit. A further challenge came from Haider and Frensch (2002) who used simulation, theoretical arguments, and data, albeit from only one study, to suggest that group-level analysis can and does smooth over discontinuities in individual-level learning curves. Haider and Frensch (2002) argued that discontinuities would be expected when individuals abruptly shift to a better strategy.

Interestingly, the tasks analysed by Heathcote et al. (2000) did not appear to allow for such strategy shifts.

A parallel line of research has examined the relationship between practice and strategy use, where strategy is typically defined as an optional method of task completion. Much of this research has concerned either children (e.g., Crowley & Siegler, 1999; Lemaire & Siegler, 1995; Robinson, 2001; Siegler, 1987, 1988a, 1996, 2006) or tasks involving algorithm–retrieval strategy shift (e.g., Delaney et al., 1998; Logan, 1992, 1988; Palmeri, 1999; Rickard, 1997, 1999). In contrast, only a few studies have looked at the shift from simple to sophisticated strategy use in an adult population. Fewer again have measured and modelled individual-level strategy use with practice (for an exception, see John & Lallement, 1997). Nonetheless, a body of research has emerged showing that: (a) most tasks have a wide range of possible strategies (e.g., Siegler & Crowley, 1996; Siegler & Shipley, 1995), (b) some strategies are more effective than others (e.g., Delaney et al., 1998; F. J. Lee, Anderson, & Matessa, 1995; John & Lallement, 1997), (c) more effective strategies tend to be used with practice (e.g., Delaney et al., 1998; John & Lallement, 1997), and (d) strategy shift, although commonly gradual, is sometimes abrupt (e.g., Delaney et al., 1998; John & Lallement, 1997; Yechiam, Erev, & Parush, 2004).

An accurate model of the relationship between practice and performance is essential to both basic science and applied fields. However, existing research has not yet determined the functional form of the relationship between practice and performance. Similarly more research is required to characterise the functional form of the relationship between practice and strategy use at the individual-level. Finally, existing models have inadequately reconciled how abrupt strategy shifts can coexist with continuous models of the learning curve. This led to the first aim of this thesis:

Aim 1 *Model the relationship between practice, strategy use, and performance at the individual-level.*

1.1.2 Individual Differences

Another way of understanding the relationship between practice, strategy use, and performance is from an individual differences perspective. Several researchers (e.g., Ackerman, 1987, 1988; Fleishman, 1972; Taatgen, 2001) have attempted to understand learning processes from an individual differences perspective. However, existing research has rarely looked at individual differences in strategy use and performance simultaneously. More data is needed on the relationship between predictors such as ability, personality, prior experience, age, and gender and criterion measures, such as strategy use and task performance. This led to the second aim of this thesis:

Aim 2 *Assess the role of individual differences in predicting strategy sophistication and performance.*

1.1.3 Instructed Strategy Shift

One way of understanding the effect of strategy shift on performance involves examining differences between self-initiated and instructed strategy shifts. Taken together several previous studies (e.g., Delaney et al., 1998; Yechiam et al., 2004) suggest that instructed strategy shifts may be more abrupt and are more likely to result in a drop in performance than self-initiated strategy shifts. More research is needed to clarify these findings. This led to the third and final aim of this thesis:

Aim 3 *Model differences between self-initiated and instructed strategy shifts on the relationship between practice, strategy use, and performance.*

1.2 Scope and Study Design

The above aims are interconnected. Measures of individual differences help to explain the variability that results when practice, strategy use, and performance are studied at the individual-level. The distinction between self-initiated and instructed strategy shifts helps to explain the conditions under which strategy shifts lead to discontinuities in the learning curve. While Aim 1 is given the greatest attention in this thesis, Aims 2 and 3 complement and extend Aim 1.

The scope of the aims in this thesis is narrowed in terms of sample, task, and context. While generalisation is a relative concept, this thesis aimed to be most relevant to understanding skill acquisition in relation to (a) adults, (b) tasks of moderate complexity, i.e., more complex than many tasks used in cognitive psychology but less complex than many real-world jobs or tasks studied in the expertise literature, (c) tasks with both psychomotor and cognitive components, (d) tasks where task completion time is the main measure of performance, and (e) tasks that allow for a shift from simple to sophisticated strategy use (i.e., not algorithm-retrieval shift). This focus captures many tasks common to educational, work, and personal settings, and is particularly representative of tasks involving interaction with computers.

To achieve these three aims within the defined scope, three studies were conducted. Each study involved participants completing a battery of individual difference measures, followed by a period of practice on a criterion text editing task. A particular advantage of the text editing task was that it enabled trial-level measurement of strategy use and task completion time. The three studies involved moderately large sample sizes in order to model variation in the relationships between practice, strategy use, and performance at the individual-level, and obtain robust estimates of predictor–criterion correlations. Data were analysed at the group- and individual-levels. Graphical and model fitting approaches were applied in order to compare the relative fit of various candidate models of the relationship between practice and task completion time and practice and strategy sophistication.

1.3 Thesis Structure

The remainder of this thesis is composed of three topic chapters and a general discussion chapter. Chapters 2 to 4 are devoted to Aims 1 to 3 respectively. Each of these three topic chapters contain a literature review, hypotheses, results, and a discussion related to the respective aim of the chapter. The topic chapters are based on the same three studies with the exception of Chapter 4 (Aim 3) which is based only on Study 3. Thus, to minimise redundancy, most of the method is described in Chapter 2 with methodological details specific to Aims 2 and 3 described in their respective chapters. Chapter 5 provides a general discussion where links between the three topics are provided, and limitations and possibilities for future research are discussed.

Chapter 2

Practice, Strategy, and Performance

Contents

2.1 Introduction

2.1.1 Overview

Research on the relationship between practice and performance has a long history (e.g., Bryan & Harter, 1899; McGeoch, 1927). Throughout this long history, a diverse range of tasks have been studied, including Morse Code transmission (Bryan & Harter, 1899), cigar rolling (Crossman, 1959), inverted reading (Kolers, 1975), alphabet arithmetic (Logan, 1988), numerosity judgements (Palmeri, 1997), playing computer games (Donchin, 1995), and controlling air traffic (F. J. Lee & Anderson, 2001). This history has also been recorded in various quantitative (e.g., Ackerman, 1987; Heathcote et al., 2000; Newell & Rosenbloom, 1981), and narrative reviews (e.g., Fitts & Posner, 1967; Lane, 1987; Langan-Fox, Armstrong, Balvin, & Anglim, 2002; Proctor & Dutta, 1995; Rosenbaum, Carlson, & Gilmore, 2000; Speelman & Kirsner, 2005) as well as more specific reviews of cognitive skill acquisition (VanLehn, 1996), motor skill acquisition (Adams, 1987), and the differences between cognitive and psychomotor skill acquisition (Rosenbaum et al., 2000).

The following section reviews relevant literature. The review starts by examining models of the relationship between practice and task completion time along with empirical evaluations. It then summarises research on the relationship between practice and strategy use. Finally, the theory and measurement of strategy relative effectiveness is discussed, particularly as it relates to practice. Hypotheses are proposed and justified throughout.

2.1.2 Practice and Performance

Overview

Influential quantitative syntheses of learning curve datasets have been conducted by Mazur and Hastie (1978), Newell and Rosenbloom (1981), and Heathcote et al. (2000). A book length treatment of research on the learning curve can be found in Lane (1987). Before discussing models of the relationship between practice and performance, the terms 'practice' and 'performance' are first defined.

Practice can refer to many activities. Within the skill acquisition literature, practice tends to refer to the effect of repeated task performance. Such a definition is distinct from the everyday use of the term which often includes adaptive training and external instruction. In skill acquisition studies, practice is commonly operationalised as either the amount of time spent practicing the task or the number of practice trials, where little has been made in the literature of this distinction. The focus on repetition leads to research designs where initial instructions are provided but typically no additional instructions are given once practice is initiated.

Task performance has been defined and measured in many different ways. At an abstract level task performance is any evaluative attribute of task execution. More typically, task performance is operationalised as completion time, accuracy, quality, or attainment. Attainment measures are merely a multiple of the inverse of completion time. As mentioned in the introduction, this thesis focuses on tasks where performance is operationalised as task completion time. This includes both reaction time when responding to a stimuli, and also task completion time on tasks involving multiple steps. Task completion time is applicable to a large number of tasks where it is speed of performance which is the main attribute that distinguishes levels of performance. Focus on task completion time also provides a link with the general chronometric aims of cognitive science.

Together, the relationship between practice and performance is known as the *learning curve*, reflecting the seemingly lawful improvement in performance that occurs with practice. Such a relationship can and has been modelled at both the *individual-level* and the *group-level*. Individual-level modelling occurs when the raw data is the performance over practice for an individual. Individual-level parameter estimation may be done separately for each individual (for a good example, see Heathcote et al., 2000) or parameters can be estimated simultaneously using approaches such as nonlinear multilevel modelling (for a review see Cudeck & Harring, 2007). Group-level modelling occurs when raw data is averaged over participants, typically using the mean, for a given trial or block of trials.

The subsequent section focuses on models of the learning curve where, in line with the scope of this thesis, practice is operationalised as amount of task repetition, and performance is operationalised as task completion time. After summarising proposed models of the learning curve, the empirical evidence for these models is critically reviewed.

Quantitative Models

Learning curves have been studied from multiple perspectives. In addition to the extensive literature in psychology, researchers in industrial engineering have frequently modelled learning curves in industrial settings (for a review see Lane, 1987). While researchers in industrial engineering have developed various models of the learning curve, such models have little relevance to understanding psychological mechanisms of learning. Such research is analysed at the group-level and typically confounds employee learning with other sources of improvement such as technologogical innovation. Thus, while group-level modelling has many important applications, such as when predictions are desired about the future productivity of a group, it is the individual-level which is most relevant for understanding psychological processes.

Also, some researchers have used polynomial models of the learning curve. However, such models have little relevance to the present discussion. A polynomial of a sufficiently high degree can be used to model any empirical function. Such models are sometimes adopted for convenience of estimation within latent growth curve frameworks (e.g., Lang & Bliese, 2009; Morrison & Brantner, 1992; Voelkle, Wittmann, & Ackerman, 2006). However, such models use linear approximations for processes that can be more parsimoniously represented by a nonlinear modelling framework. Polynomial models do not capture the asymptotic behaviour of learning curves and make inaccurate predictions outside the range of data (for discussion of the benefits of nonlinear modellings, see Bates & Watts, 1988; Cudeck & Harring, 2007; Kelley, 2009). They also require more parameters to obtain equivalent levels of fit relative to truly nonlinear functions.

As described by Lane (1987) and others, many quantitative models of the learning curve have been proposed. However, as stated in the introduction, this thesis focuses mainly on power, exponential, and discontinuous functions.

When presenting functions in this thesis, notation is used that is consistent with Huet, Bouvier, Poursat, and Jolivet (2004). The power function (first proposed by Snoddy (1926), and rediscovered by De Jong (1957)) in its three parameter form is

$$f_{P3}(x, \theta) = \theta_1 x^{-\theta_2} + \theta_3, \tag{2.1}$$

and the three parameter exponential function is

$$f_{E3}(x, \theta) = \theta_1 \exp(-\theta_2(x - 1)) + \theta_3, \tag{2.2}$$

where $f(x, \theta)$ is the expected task completion time for trial x and the subscripts $P3$ and $E3$ denote the three parameter power and exponential functions respectively. The amount of practice is denoted by x and is typically represented by an integer starting from 1 for the power function and 0 for the exponential function. This thesis uses $x - 1$ in the exponential function (i.e., Equation 2.2) in order to model all learning curves with trial number starting from 1. The asymptotic level of expected performance is captured by θ_3, which is predicted to be approached as amount of practice approaches infinity. Amount of learning is the difference between expected performance on the first trial (i.e., $x = 1$) and asymptotic performance (i.e., θ_3), and is captured by θ_1. The shape of the learning curve is determined by θ_2.

A two parameter power function can also be obtained by fixing the asymptote θ_3 to zero. While having fewer parameters is desirable, it makes the implausible prediction that after infinite practice, task completion time would be zero. Thus, while the function often predicts the observed data well, it makes unreasonable predictions outside the range of the data, which is one of the key motivations for nonlinear modelling.

The power function sometimes includes a fourth parameter,

$$f_{P4}(x, \theta) = \theta_1(x + \theta_4)^{-\theta_2} + \theta_3, \tag{2.3}$$

to represent prior learning (introduced by Seibel, 1963). This parameter has typically been fixed to zero by subsequent researchers (e.g., F. J. Lee & Anderson, 2001). Also, Heathcote et al. (2000) have proposed a four parameter function called the APEX function which incorporates aspects of both the three parameter power and exponential functions,

$$f_{AP}(x, \theta) = \theta_1 \exp(-\theta_2 x)x^{-\theta_4} + \theta_3. \tag{2.4}$$

Further discussion of mathematical properties of the above functions can be found in Newell and Rosenbloom (1981) and Heathcote et al. (2000).

The preceding functions are all monotonically decreasing and monotonically decelerating. This contrasts with a set of functions proposed by several researchers that involve discontinuities. Discontinuous functions can capture temporary plateaus in learning (e.g., Bryan & Harter, 1899) and abrupt improvements in performance (e.g., Haider & Frensch, 2002; Kolers & Duchnicky, 1985). Discontinuities can occur in performance itself or in the rate of learning (i.e., the first derivative). While learning curve researchers in psychology have rarely formalised discontinuous models, such models are often discussed in statistics and econometrics (e.g., Zeileis, Leisch, Hornik, & Kleiber, 2002; Zeileis, Kleiber, Krmer, & Hornik, 2003). Discontinuous models can be seen as piecewise functions that have one or more breakpoints and where the number of segments is equal to one more than the number of breakpoints. Each segment has its own function. The functions for each segment may be of the same or different functional type. The timing of the breakpoints is also typically a parameter.

It would appear that psychological researchers have rarely formally analysed discontinuities. One exception, Beem (1995), critical of informal methods (e.g., Kyllonen, Lohman, & Woltz, 1984), developed software for a limited form of breakpoint analysis and applied it to several psychological datasets (Ippel & Beem, 1987; Luwel, Beem, Onghena, & Verschaffel, 2001; Luwel, Verschaffel, Onghena, & De Corte, 2003; Verschaffel, De Corte, Lamote, & Dherdt, 1998). Seber and Wild (2003) discuss modelling discontinuities in nonlinear functions. Reviews of software (Andersen, Carstensen, Hernndez-Garca, & Duarte, 2009), an outline of the statistical theory (Chapter 9 in Seber & Wild, 2003), and a general review (Hansen, 2001; Khodadadi & Asgharian, 2008), are also available, along with several influential papers on the topic (Andrews, 1993; Brown, Durbin, & Evans, 1975; Chu, Stinchcombe, & White, 1996; W. Kramer, Ploberger, & Alt, 1988; Ploberger & Kramer, 1992). In the absence of formalised discontinuous models of the learning curve, the following paragraphs attempt to extract plausible discontinuous models for model testing purposes.

The most basic form of a discontinuous learning model involves two constants where the constant changes following a breakpoint,

$$f_{\mathrm{CnCn}}(x,\theta) = \begin{cases} \theta_1 & \text{if } x \le \theta_3 \\ \theta_2 & \text{if } x > \theta_3 \end{cases}. \tag{2.5}$$

Such a model could represent an idealised version of insight learning. Haider and Frensch (2002) discuss it implicitly as a hypothetical function potentially relevant to performance following abrupt strategy shift at the individual-level. Given that existing research strongly suggests that improvements in performance with practice are more gradual, the model is unlikely to provide a good fit to data. However, it does provide a baseline comparison for evaluating more sophisticated discontinuous models.

More plausible discontinuous models combine a breakpoint with monotonic improvement and deceleration typical of most learning curves. Such models are suggested, but not formalised, in the work of Delaney et al. (1998). In an attempt to formalise such discontinuous models whilst also keeping the number of parameters to a reasonable level, the following two models are proposed. The first model combines two two-parameter power functions separated by a breakpoint

$$f_{\mathrm{P2BP2}}(x,\theta) = \begin{cases} \theta_1 x^{-\theta_2} & \text{if } x \le \theta_5 \\ \theta_3 x^{-\theta_4} & \text{if } x > \theta_5 \end{cases}. \tag{2.6}$$

The second such model is a three parameter exponential model with a different asymptotic parameter based on the breakpoint

$$f_{\mathrm{E3B}}(x,\theta) = \begin{cases} \theta_1 \exp(-\theta_2 x) + \theta_3 & \text{if } x \le \theta_5 \\ \theta_1 \exp(-\theta_2 x) + \theta_4 & \text{if } x > \theta_5 \end{cases}. \tag{2.7}$$

All three discontinuous models proposed include a single breakpoint as a parameter. This is a simplification, given that multiple discontinuities are possible, but arguably a useful one for facilitating parameter estimation. All three discontinuous functions permit both abrupt improvements and abrupt declines in performance. The *P2BP2* and *E3B* functions can both potentially capture, for example, the temporary drop in performance, that might follow a strategy shift.

Empirical Evaluation

Power Functions Several early researchers proposed that the power function (De Jong, 1957; Snoddy, 1926) was a good model of the relationship between practice and task completion time. Newell and Rosenbloom (1981) influentially analysed many existing skill acquisition datasets across various types of tasks (e.g., Crossman, 1959; Hirsch, 1952; Kolers, 1975; Neisser, 1963; Neves & Anderson, 1981; Seibel, 1963; Snoddy, 1926). The datasets varied in whether they were individual-level, group-level, or cross-sectional. Newell and Rosenbloom found that the two parameter power function provided a reasonable fit to the obtained data and that a four parameter power function was superior to a three parameter exponential function. They found the evidence in support of the power function so compelling that they labelled the relationship the Power Law of Practice.

Many subsequent researchers have accepted the Power Law of Practice and incorporated it into cognitive architectures and other models of the learning process (for a review, see Ritter & Schooler, 2002). These include Logan's Instance Theory of Automaticity (Logan, 1992), Anderson and colleagues' ACT-R (Anderson et al., 2004), Lee and Anderson's (2001) model of subtasks, and Delaney et al.'s (1998) model of learning within strategies. Subsequent researchers have typically used the three parameter function to model the learning curve, often at the group-level.

Several reasons for not including the fourth parameter exist (see Heathcote et al., 2000). First, the fourth parameter is meant to represent prior practice. Thus, it arguably should not be a parameter that is fit to the data; it should be measured. However, measurement of prior practice is often not possible. Second, many studies are designed with novel tasks, such that prior practice should be zero. Third, a four parameter model may sometimes suffer from issues of estimation particularly in noisy individual-level data. Fourth, the three parameter power function often performs quite well at the group-level.

Exponential Function More recently several researchers (i.e., Heathcote et al., 2000; Haider & Frensch, 2002; Lacey, 2007; Myung, Kim, & Pitt, 2000) have challenged the validity of the Power Law of Practice. These researchers argued that group-level analyses provide a biased model of the individual-level, and that it is the individual-level which is psychologically meaningful. The general bias of group-level analyses has long been known in the learning curve literature (see Estes, 1956), but it is often not mentioned by researchers performing group-level analyses. The critique of group-level analysis however is consistent with the increasing interest in psychology in quantitative and theoretical models of within-person change (e.g., Cudeck & Harring, 2007; Raudenbush, 2001; Singer & Willett, 2003).

A major critique of group-level analyses was presented by Heathcote et al. (2000) who analysed a larger number of studies than Newell and Rosenbloom (1981). In contrast to Newell and Rosenbloom (1981), Heathcote and colleagues performed all analyses both at the group-level and at the individual-level. Consistent with actual use in subsequent research, they compared the three parameter power function with the three parameter exponential function. Consistent with Newell and Rosenbloom (1981), Heathcote et al. found that at the group-level the power function provided superior fit. However, at the individual-level, the exponential function provided a better fit for a majority of individuals in a majority of studies. Also, when four parameter models were included, the APEX function provided better fit than the four parameter power function.

Heathcote et al. (2000) showed that the potential biasing effects of group-level analyses applied to actual skill acquisition datasets. The results suggest that the power law should be repealed and possibly replaced with an exponential law. Alternatively, the functional form of the learning curve may vary between individuals with the exponential function more often being superior to the power function. It should also be noted that while Heathcote et al.'s analyses included a wide range of task types (i.e., memory search, counting, mental arithmetic, alphabet arithmetic, visual search, mental rotation, and simple motor learning), they did not include more complex psychomotor tasks such as text editing or air traffic control. However, based on the research of Newell and Rosenbloom (1981) and Heathcote et al. (2000), the following hypotheses are proposed:

Hypothesis 2.1 *The relationship between practice and task completion time at the group-level is better explained by a three parameter power function than a three parameter exponential function.*

Hypothesis 2.2 *The relationship between practice and task completion time at the individual-level is better explained by a three parameter exponential function than a three parameter power function.*

If both Hypothesis 2.1 and Hypothesis 2.2 are correct, then by implication, group-level models provide biased evidence of the relationship between practice and performance at the individual-level.

Discontinuous Functions A second critique of the Power Law of Practice comes from Haider and Frensch (2002). They argued that group-level analyses often disguised discontinuities in the learning curve at the individual-level. If this were true, neither the power nor the exponential function would be an appropriate model. Haider and Frensch showed mathematically and by simulation that discontinuities in the learning curve at the individual-level could completely disappear at the group-level if, as would be expected, the timing of the discontinuities varied between individuals.

Haider and Frensch (2002) also presented empirical data from an alphabet string verification task where discontinuities consistently emerged in individual-level learning curves. The task required participants to read a string of letters such as "C D E F G [4] L" (p. 393) and indicate whether after substituting the number for alphabetical letters the string was in alphabetical order. The task was structured such that the initial letters were always in alphabetical order. Thus, an effective strategy was for participants to ignore the initial letters and only process the letter–number–letter triple (e.g., "G[4]L"). Graphs of individual-level data suggested that in addition to participants getting faster with practice, most participants had a discontinuity whereby reaction time increased abruptly. The timing of this discontinuity varied between participants. Haider and Frensch argued that the discontinuity occurred as a result of participants abruptly ceasing to process the initial redundant letters in the string.

Generalising these findings, Haider and Frensch (2002) proposed that abrupt strategy shifts from less to more effective strategies would exist on many tasks. They suggested that if a shift to a more effective strategy was abrupt and particularly if it occurred later in practice, a discontinuity in the learning curve would occur. They reasoned that the absence of previously identified discontinuities was due to a reliance on group-level analyses.

The validity of Haider and Frensch's argument depends on the empirical frequency of discontinuities. Many researchers have analysed data at the individual-level and have not found discontinuities (e.g., Kolers, 1975; Seibel, 1963). Similarly, Bryan and Harter (1897, 1899) found evidence for plateaus in individual-level learning curves for learning Morse Code. However, subsequent researchers have generally failed to replicate this finding and have suggested that plateaus are rare, often the result of poor training, and not of fundamental theoretical interest (Keller, 1958; Reed & Zinszer, 1943).

In summary, discontinuities sometimes occur at the individual-level. However, prevalence rates of discontinuities, the conditions which increase prevalence, and the types of discontinuities that occur are less clear. The infrequency of empirical reports of discontinuities may be partially explained by group-level analyses and a lack of readily available breakpoint modelling techniques. However, it is more likely that discontinuities occur infrequently and possibly only under special conditions. In particular, tasks that permit an abrupt shift in strategy seem more likely to show discontinuities. However, even then, there are several reasons to suspect that discontinuities will be rare. This led to the following two hypotheses:

Hypothesis 2.3 *Discontinuities do not occur in the relationship between practice and task completion time at the group-level.*

Hypothesis 2.4 *Discontinuities occasionally occur in the relationship between practice and task completion time at the individual-level.*

If true, the above hypotheses in combination imply another biasing effect of group-level analyses.

Conclusion

In summary, individual-level learning curves are the psychologically meaningful level of analysis, and research suggests that the group-level provides a biased representation of the individual-level. When modelling the individual-level, the best evidence is that a three parameter exponential function provides on average a superior fit in comparison to the three parameter power function. This might lead us to call such a relationship the Exponential Law of Practice, but the best fitting function seems to vary sufficiently between individuals that the term *law* may be inappropriate. Also, more research is needed in order to ascertain the conditions under which discontinuities are likely to occur. In particular, this requires the application of quantitative modelling approaches that are designed to identify discontinuities. It may be that tasks that permit abrupt strategy shifts from poorer to better strategies provide likely candidates for discontinuities. Such tasks were absent from Heathcote et al.'s (2000) review and, as previously mentioned, constitute the focus of this thesis.

2.1.3 Practice and Strategy

Overview

The following section of the literature review focuses on the relationship between practice and strategy use. As well as being inherently interesting, changes in strategy use are one reason to expect discontinuities in the learning curve. Before discussing the effect of strategy shift on performance, research on the relationship between practice and strategy use is first critically reviewed.

Researchers have explored the relationship between practice and strategy use from a range of perspectives including childhood learning (Lemaire & Siegler, 1995; Siegler, 1988b, 2006; Siegler & Crowley, 1996, 1992; Siegler & Shipley, 1995; Siegler & Stern, 1998), expertise (e.g., Ericsson & Charness, 1994), time and motion (e.g., Gilbreth, 1911), cognitive architectures (e.g., productions in ACT-R, Anderson, Matessa, &

Legiere, 1997), algorithm–retrieval shifts (Delaney et al., 1998; Logan, 1988; Rickard, 2004), and human computer interaction (e.g., methods in the GOMS approach, (Card, Moran, & Newell, 1983)). Reviews of research examining strategy use can be be found in Siegler and Shipley (1995) and Cary and Reder (2002).

Before reviewing the literature on strategy use it is important to define the term strategy. The term has been used in several different ways even within cognitive psychology. In this thesis *strategy* refers to an optional means of task completion (e.g., Delaney et al., 1998; Lemaire & Siegler, 1995; Rickard, 1997). An equivalent but more elaborate definition is a "method used for obtaining solutions in skill-learning tasks in which different methods are possible" (Touron & Hertzog, 2004, p. 565). The usage of the term strategy in this thesis is consistent with the usage of many other researchers (e.g., Delaney et al., 1998; John & Lallement, 1997; Lemaire & Siegler, 1995; Haider & Frensch, 2002; Siegler & Lemaire, 1997; Siegler & Shipley, 1995; Yechiam et al., 2004). This definition allows for both behavioural and cognitive strategies. Such a definition does not require conscious awareness (see Reder & Ritter, 1992).

'Strategy' can be a nebulous term capturing any possible qualitative distinction in task execution. As stated in the introduction, this thesis is focused on tasks that permit a simple–sophisticated strategy shift. Such a shift commonly (a) is facilitated by explicit processes of instruction, deduction, or analogy, (b) is typically able to be executed accurately once known, and (c) typically requires practice to be executed rapidly. The following examples from the literature capture the above class of strategy shifts: (a) one-at-a-time (simple) versus aggregated (sophisticated) creation of computer graphics (Bhavnani & John, 2000; Charman & Howes, 2003), (b) simple versus sophisticated text editing commands (Cook, Kay, Ryan, & Thomas, 1995), (c) mouse (simple) versus script-based (sophisticated) spreadsheet manipulation (Yechiam et al., 2004), (d) lifting bricks from the ground (simple) versus using an adjustable level (sophisticated) when laying bricks (Gilbreth, 1911); (e) one at a time (simple) or sequential (simple) versus opportunistic (sophisticated) strategies for allocating planes to runways on an air-traffic control simulation (John & Lallement, 1997); (f) processing all information (simple) versus ignoring unnecessary information (sophisticated) on an alphabet verification task (Haider & Frensch, 1999, 2002).

This focus contrasts with algorithm–retrieval strategy shifts, which have been frequently researched in relation to learning curves. Examples of algorithm–retrieval strategy shifts include solving novel arithmetic problems by either computing or retrieving an answer (Blessing & Anderson, 1996; Delaney et al., 1998; Rickard, 1997), alphabet arithmetic (Logan, 1988), retrieving an answer or using a plausibility strategy for story recollection (Reder, 1988), dot counting strategies (Palmeri, 1997), and scan versus retrieval strategies on a noun–pair lookup task (Ackerman & Woltz, 1994).

Algorithm-retrial shift differs from simple-sophisticated shift in several respects. First, participants tend to adopt a retrieval strategy with practice in the absence of explicit instruction to retrieve, whereas many participants do not adopt a sophisticated strategy by mere practice alone (see work on insight, e.g., Luchins, 1942). In particular, explicit instruction can can dramatically increase the uptake of sophisticated strategies. Second, use of retrieval strategy requires that an association has formed in memory, whereas individuals are often able to perform a sophisticated strategy if they are specifically told to use it. Finally, retrieval shifts typically involve an immediate improvement in speed, whereas sophisticated strategies can involve a temporary reduction in performance as performance of the new strategy is refined.

Theoretical Models

Many theoretical models exist that either directly concern strategy use or have strategy use as a component. Some of these are general models of human cognition such as ACT-R (Anderson et al., 2004), and others concern particular tasks or domains. All such models have to deal with three things: (a) how a strategy enters the repertoire, (b) how a strategy is selected, and (c) how selection of strategies from a repertoire changes over time.

In order for a strategy to be selected it is required that the strategy be known to the individual in some sense. Mechanisms can be broadly distinguished based on whether strategies enter explicitly versus implicitly (for a theory and commentary, see Dienes & Perner, 1999). Crowley, Shrager, and Siegler (1997) also discuss differences in strategy shift processes using the terms *metacognitive* and *associative* for what roughly corresponds to explicit and implicit processes respectively. Explicit mechanisms include analogy, deduction, and receiving new declarative knowledge, such as through written or verbal instruction (Reber, 1989). Implicit mechanisms are based on detecting correlations in the task environment.

Strategy selection has been incorporated into several models of problem solving and skill acquisition (e.g., Lovett, 2005; Lovett & Anderson, 1996; J. W. Payne, Bettman, & Johnson, 1988; Schunn, Reder, & Nhouyvanisvong, 1997; Siegler & Shipley, 1995). While each model has its nuances, a general characterisation is as follows. An individual has a repertoire of strategies each of which has an associated utility in a context. The probability of a strategy being selected is monotonically related to its perceived relative utility for the individual on the task in the context. The utility is assumed to include both task-relevant goals such as efficiency and accuracy, as well as costs associated with cognitive resource requirements. Gray, Sims, Fu, and Schoelles (2006) suggested resource costs included time to access an option, motoric complexity, and memory demands. This framework is at the heart of the 'rational' perspective proposed by Anderson (1990) and reflected in the ACT-R theory (Anderson et al., 2004).

Models of strategy selection typically include a stochastic component. This is captured in Siegler's Overlapping Waves Theory (see Siegler, 2006) based on observations that children continue to use multiple strategies over long periods of time and that learning occurs as a gradual process of replacing simpler strategies with better and more sophisticated strategies. It is also captured in the results of many probability matching studies whereby instead of participants always selecting the most probable outcome, participants tend to choose options in proportion to their probability of success (e.g., Lovett & Anderson, 1996). Several explanations for this diversity include: (a) the broader adaptiveness of maintaining a diverse strategy repertoire, (b) forgetting processes, and (c) the complexity of the environment.

Strategy selection also changes with practice both in terms of strategies entering the repertoire and in the utilities associated with given strategies. Lemaire and Siegler (1995, p. 83) proposed that there were four sources of cognitive change: "(a) which strategies are used, (b) when each strategy is used, (c) how each strategy is executed, (d) how strategies are chosen." Standard three phase theories of skill acquisition (Anderson, 1982; Fitts & Posner, 1967; Schneider & Shiffrin, 1977) suggest that phase one involves a search and refinement of strategies on a task. In a related sense, inadequate performance, which is more likely to occur early in practice, may trigger a search for a better strategy. M. K. Singley and Anderson (1989) suggested that the search for a better strategy is more likely to occur when a strategy fails completely as opposed to when it is merely suboptimal. Bhavnani and John (2000) note that there can be cognitive costs in executing a strategy and costs in identifying a superior strategy. Despite these shortcomings, strategy selection is typically assumed to be relatively adaptive given the individual's knowledge and resources.

Strategy selection theories also assume that perceived utilities for a given strategy are reasonably accurate. In order for perceived utilities to be accurate, a process is posited whereby feedback resulting from strategy use is assumed to cause an update of perceived utility based on effectiveness of the strategy in terms of performance outcomes and the cognitive resource requirements. For example, Gray and Boehm-Davis (2000) showed that strategy use changed in adaptive ways in response to interface changes that made some strategies slightly slower. In another example, Erev, Ert, and Yechiam (2008) within the context of decision making under uncertainty propose a model whereby individuals first engage in a period of exploration where the utility of various choices are evaluated followed by a period of exploitation where choices with greater perceived utility tend to be sampled. The updating process also implies the existence of an attribution processes whereby results from an action are attributed to the strategy used.

Methodological Issues

Before discussing the empirical relationship between practice and strategy use it is worth discussing methodological issues related to measuring and modelling strategy use. Purely cognitive strategies often require probes to measure strategy use (e.g., Rickard, 1997; Touron, 2002). Two issues often discussed in the literature are validity (i.e., can participants accurately report their strategy use?) and reactance (i.e., does having participants report their strategy use change their behaviour?). In other cases, strategy use has been inferred from reaction time (e.g., Haider

& Frensch, 2002). One of the benefits of studying strategies with behavioural consequences is that reactance is unlikely to occur. In particular, computerised tasks can be designed that permit the silent recording of participant behaviour. An example of this can be seen in the use of key logs to extract strategy use on an air traffic control simulation (John & Lallement, 1997).

A third methodological issue relates to the level of analysis when modelling the relationship between practice and strategy use. As with the relationship between practice and performance, modelling at the group-level leads to a biased representation of the individual-level. If strategy shifts are abrupt and occur at time points that vary across individuals, a smooth function will tend to result at the group-level even though a discontinuous model is more reasonable. While some researchers have reported individual-level strategy use (e.g., Delaney et al., 1998; John & Lallement, 1997) many researchers have only reported results at the group-level (e.g. Touron & Hertzog, 2004). While the group-level can provide a convenient overview of the general transition of a strategy, it is the individual-level that indicates whether the shifts were abrupt or gradual, and partial or complete.

Models

Overview In order to add rigour to the modelling of the relationship between practice and strategy use, a set of models of the relationship are proposed. All the models assume that strategy use, i.e., that which is being modelled, has been defined on a zero to one scale. In the present thesis, strategy use was operationalised as the proportion of keys classified as sophisticated, but in other cases, it might the probability of using a given sophisticated strategy. While some researchers have discussed mathematical models of the relationship, this has not always been the case, and thus, potential models had to be developed based on qualitative descriptions in the literature.

Assumptions of strategy shift immediately rule out linear, quadratic, and higher-order polynomial functions. While a polynomial of sufficiently high-order will provide good fit to observed data, as mentioned in relation to practice and performance, such polynomial functions have specific problems. First, they make inappropriate predictions outside the range of the data. An inverted U-shaped quadratic function eventually predicts that strategy sophistication will decline with practice. Both linear and quadratic functions suggest that with sufficient practice, probability of use of sophisticated strategy will be either less than zero or greater than one. Second, they do not suggest plausible change mechanisms in the same way that some other functions do.

More plausible functions are those that increase monotonically to an asymptote. The remainder of this section presents several plausible functions, focusing on a *Constant* function, two monotonically increasing and decelerating functions (*Michaelis–Menten* and *Saturating Exponential*), a sigmoidal function (*Logistic*), and two functions with discontinuities (*Constant–Constant* and *Constant–Saturating Exponential*).

Constant A simple model that can be used to describe individuals who do not change strategy is the Constant model,

$$f_{\mathrm{Cn}}(x, \theta) = \theta_1. \tag{2.8}$$

Such a function can be used to describe a range of individuals. These include constant low, medium, and high strategy sophistication individuals, as well as individuals who are variable in their strategy sophistication but the general probability of strategy sophistication over time is constant.

Michaelis–Menten The Michaelis–Menten (Michaelis & Menten, 1913) function is a monotonically increasing and decelerating function (for a discussion, see Chapter 3 of Bolker, 2008). Bolker (2008) reviewed the history of the function noting how it was proposed by Michaelis and Menten (1913) in relation to enzyme kinetics. Bolker (2008) also commented on other similar functions such as the Monod function, Holling Type II, and the Beverton–Holt model. A three parameter version of the Michaelis–Menten function allows for a non-zero starting point and a non-zero asymptote,

$$f_{\mathrm{MM}}(x, \theta) = \frac{\theta_1(x-1)}{\theta_2 + (x-1)} + \theta_3, \tag{2.9}$$

where x is trial number starting from 1, θ_1 is the amount of change, θ_2 is the number of trials it takes for half the change to occur, and θ_3 is strategy sophistication on trial 1. The function is positive, negatively accelerated, and approaches an asymptote. In order for the values to be meaningful, the following constraints need to be satisfied: $\theta_1 \geq 0$; $\theta_2 > 0$; $\theta_3 \geq 0$; $\theta_1 + \theta_3 \leq 1$. Cudeck and Harring (2007) stated that the Michaelis–Menten function provided a clear set of interpretable parameters, using Frensch, Lindenberger, and Kray (1999) to illustrate this point. Frensch et al. (1999) used the function to model the relationship between rate of stimulus change to monitoring accuracy on a continuous monitoring task. However, following a literature search, no evidence was found of the function being applied to modelling the relationship between practice and strategy use.

Saturating Exponential Another monotonically increasing and accelerating function is the Saturating Exponential function. Bolker (2008) suggested that the differences between it and the Michaelis–Menten function are subtle. The function is

$$f_{\mathrm{SE}}(x, \theta) = \theta_1(1 - \exp(-\theta_2(x - 1))) + \theta_3, \tag{2.10}$$

where θ_1 is strategy sophistication on trial 1, $\theta_1 + \theta_3$ is asymptotic strategy sophistication, and θ_2 is a shape parameter.

Logistic A contrast to the above two functions is the logistic function, which suggests a sigmoidal relationship between practice and probability of strategy use. One parameterisation of the three parameter logistic function is

$$f_{\mathrm{Lg}}(x, \theta) = \frac{\theta_1}{1 + \exp(-(x - \theta_2)/\theta_3)}, \tag{2.11}$$

where θ_1 represents asymptotic proportion of strategy use, θ_2 represents the trial of the inflection point (i.e., half the asymptote) and θ_3 is a scaling value. θ_1 is often set to 1.0 where it is assumed that strategy use will approach 100% with infinite practice.

Constant–Constant The Constant–Constant model represents an abrupt onset and an abrupt shift in strategy use,

$$f_{\mathrm{CnCn}}(x, \theta) = \begin{cases} \theta_1 & \text{if } x \leq \theta_3 \\ \theta_2 & \text{if } x > \theta_3 \end{cases}, \tag{2.12}$$

where there is a breakpoint at θ_3 after which, strategy use changes from θ_1 to θ_2. This is consistent with an abrupt shift typically from not using a strategy to using a strategy.

Constant–Saturating Exponential A second model with a breakpoint is the Constant–Saturating Exponential,

$$f_{\mathrm{CnSE}}(x, \theta) = \begin{cases} \theta_1 & \text{if } x \leq \theta_5 \\ \theta_2(1 - \exp(-\theta_3(x-1))) + \theta_4 & \text{if } x > \theta_5 \end{cases}. \tag{2.13}$$

This function is consistent with an initial period of low or otherwise constant strategy use, followed by an abrupt onset of a strategy shift, but where the actual shift is gradual. A similar simpler model might also be explored where θ_4 is constrained to be equal to θ_1. In some respects it is similar to a logistic function, yet it is theoretically distinguished by its abrupt onset.

Empirical Evaluation

The empirical evidence regarding the relationship between practice and strategy use is diverse in terms of task, design, sample, and context. Also, a lot of existing research has the previously mentioned methodological issues related both to inadequate strategy measurement and not reporting individual-level analyses.

Before discussing simple–sophisticated strategy shifts, the algorithm–retrieval strategy shift literature is briefly described. Reflecting the fundamental importance of algorithm–retrieval shift, many studies have examined this shift (e.g., Delaney et al., 1998; Logan, 1992, 1988; Palmeri, 1999; Rickard, 1997, 1999, 2004). Single digit multiplication provides a prototypical example. With practice, individuals transition from solving problems, like four times four, using general algorithms, such as repeated summing, to directly retrieving answers from memory. Common findings from such studies include: (a) at the group-level, retrieval use tends to increase following either an exponential or logistic function, (b) retrieval shift is largely item specific, (c) within an item there is some, but not a lot, of shift back and forth between algorithm and retrieval, (d) strategy shift is less abrupt when operating over more naturalistic time frames, (e) often a small subset of individuals rarely if ever use retrieval, (f) individuals are relatively adaptive in their selection of algorithm or retrieval strategies.

Simple-sophisticated shift differs in several ways from algorithm–retrieval shift. Retrieval strategies require that the stimulus-response pair has been memorised which takes at least one trial, but typically many more to occur; sophisticated strategies often only require awareness of the strategy, and can often be used on the first trial. Transition to retrieval strategies typically occur naturally with practice; sophisticated strategies are often not suggested by the simpler strategy; for example, using the mouse does not suggest the existence of a script (Yechiam et al., 2004). Retrieval strategies are almost always faster than the algorithm strategy; sophisticated strategies can be slower than the simpler strategy when first adopted; for example, learning more sophisticated text editing short cut keys can be awkward at first.

While the research on individual-level strategy shifts from simple to sophisticated strategies is sparse, several relevant findings exist. John and Lallement (1997) examined strategy use on an air traffic control simulation. They found that around half the participants did not shift strategy, which is consistent with the Constant model. Other participants adopted a different strategy after several trials, although John and Lallement (1997) did not specify exactly when the transition occurred. They labelled 39 of the 50 observed strategy shifts as gradual and the remainder as abrupt. These two levels of abruptness are consistent with Constant–Constant and Constant–Saturated Exponential models. If the onset of the shift occurred in the first trial, then Michaelis–Menten or Saturating Exponential may provide better fit.

Another line of research has examined how people use computer software. A common observation is that users often plateau at strategies that are asymptotically suboptimal (for a summary of this literature, see Bhavnani & John, 2000; J. M. Carroll & Rosson, 1987; Charman & Howes, 2003). Classic studies in time-and-motion found that industrial workers rarely discovered the most efficient strategies for performing their tasks (Gilbreth, 1911). Charman and Howes (2003) found that about half the participants in their study shifted to a more sophisticated graphics drawing strategy, although results did not indicate how abrupt the shift was nor the timing of the shift. Charman and Howes (2003) suggested that strategy shift may be greater in the lab than in the real-world because participants are focused more on low-level task goals whereas real world workers are more concerned with high-level project goals. Furthermore, any discussion of optimality raises the question of what is being optimised, and the various costs and benefits of searching for optimal strategies.

Similarly, Yechiam et al. (2004) examined the use of a mouse versus a script-based strategy, along with a much slower keyboard-based strategy, on a spreadsheet manipulation task. Half of the participants were allowed to use the mouse strategy from the start, while the other half were required to use the script strategy initially. The script was difficult to learn, but ultimately quicker than the other strategies. They found that participants were generally unlikely to switch to the script strategy when it was introduced half-way through practice. Also, although Yechiam et al. (2004) did not report individual-level results, the pattern of results suggested that the shift to the using the script-based strategy tended to occur on the first or second trial of its availability or not at all. It would also seem that the shift to the script-based strategy was abrupt.

Individual-Level Predictions

In summary, the literature on simple–sophisticated strategy shift is sparse. Results are somewhat contingent on task, context, and participant properties. Studies have also often failed to report results in such a way to enable an understanding of individual-level patterns of shift in terms of timing and abruptness. As with the relationship between practice and task completion time, it is the individual-level that is relevant for understanding how individuals learn. In light of the above literature, the following hypotheses are proposed.

Hypothesis 2.5 *For individuals that shift strategy, some shifts are gradual and others are abrupt.*

Abrupt shifts will be indicated by support for the Constant–Constant model. Gradual shifts will be reflected by support for Michaelis–Menten, Saturating Exponential, Logistic, and Constant–Saturating Exponential models. In general strategy shift is likely to be more abrupt where explicit instructions are given, generalisation is straight forward, and the sophisticated strategy is immediately and always far more effective than the simpler strategy.

Hypothesis 2.6 *Many individuals will be characterised by only partial strategy sophistication.*

This will be indicated by participants who use a mix of both simple and sophisticated strategies.

In contrast to algorithm–retrieval shifts which often take time to develop and can follow various continuous functions, simple–sophisticated strategy shift is hypothesised to have the following pattern:

Hypothesis 2.7 *The probability of a strategy shift occurring decreases monotonically with practice.*

Given the tendencies for inertia, it is likely that some individuals never discover sophisticated strategies.

Hypothesis 2.8 *For some individuals, the relationship between practice and strategy sophistication involves near continuous use of a simple strategy.*

This will be indicated by support for the Constant model with a θ_1 (i.e., the mean proportion of strategy use) at a low level.

It is also clear that some individuals either know a sophisticated strategy from the start or learn it on the very first trial.

Hypothesis 2.9 *For some individuals, the relationship between practice and strategy sophistication involves near continuous use of a sophisticated strategy.*

This will be indicated by support for the Constant model with a θ_1 (i.e., the mean proportion of strategy use) at a high level.

Group-Level Predictions

While not the primary focus of this thesis, it is still interesting to consider the relationship between practice and strategy sophistication at the group-level. Several functions have been proposed in the literature. At other times, results have been presented graphically that suggest a functional form without defining one. The following discussion is limited to functions where the dependent variable is the proportion of instances of use of a sophisticated strategy, or for comparison purposes, a retrieval strategy.

It is generally assumed that use of a sophisticated strategy increases with practice based on the adaptive nature of strategy use. Thus, the following questions remain: First, is sophisticated strategy use zero on the first trial? Second, does sophisticated strategy use asymptote at 100 percent sophisticated strategy use or something less than that? Third, although the assumption of a monotonically increasing function seems almost certain, is the function monotonically decelerating, or does it follow a sigmoidal function? Fourth, related to the third question, what is a functional form of the relationship? Finally, how does this functional relationship vary as a function of task, sample, and context?

For several reasons, a limited set of the previously mentioned functions seem plausible for modelling the relationship between practice and strategy sophistication at the group-level. These are the Michaelis–Menten, Saturating Exponential, and Logistic. The two discontinuity functions seem implausible at the group-level because the timing of any discontinuity should vary between individuals, and such discontinuity models are assumed to hold only for some participants. The Constant model is inconsistent with ample evidence that strategy use does change from simple to sophisticated strategies at least for some participants.

At least three functions have been proposed to model the group-level relationship between practice and strategy use. First, Rickard (1997) applied a logistic function to model algorithm–retrieval shift. The logistic function assumes that the rate of adoption of a strategy initially accelerates before decelerating. This is a plausible model of algorithm–retrieval shift because it may take several trials for a memory trace to become established. However, in the case of simple–sophisticated strategy shift, monotonic deceleration seems more plausible. Search for a sophisticated strategy may be maximal when an individual is first orienting to a task at the start of practice. Initial task instructions may also trigger insights that affect early trials.

Touron and Hertzog (2004) used a one parameter version of the Saturating Exponential function to model the relationship between practice and probability of using a retrieval strategy on an algorithm–retrieval task at the group-level. The one-parameter version was achieved by implicitly constraining θ_1 to 1.0 and θ_3 to 0.0, based on the assumption that initial probability of retrieval would be zero, and after infinite practice would be one. While such constraints may be reasonable in the context of algorithm–retrieval shift, they seem inappropriate in relation to simple–sophisticated strategy shift. It is likely that initial strategy sophistication will often be above zero and asymptotic levels will be below one.

Based on the literature and hypotheses mentioned earlier, various expectations result regarding the relationship between practice and strategy sophistication at the group-level. These individual-level assumptions are that (a) some individuals use the sophisticated strategy from the first trial, (b) some individuals never use the sophisticated strategy even after extensive practice, (c) the probability of shift decreases with practice, and (d) individuals vary in whether the shift is gradual or abrupt. Based on these assumptions it would be expected that the group-level curve relating practice to strategy sophistication should (a) start at a non-zero level, (b) approach an asymptotic level less than one, (c) and be a negatively accelerated function, whereby the greatest change happens in the early period of practice.

Although they did not use a mathematical model, F. J. Lee et al. (1995) presented data broadly consistent with the above propositions. They graphed the relationship between practice and use of a more efficient runway landing strategy, called the Hold 1 strategy, on the Kanfer–Ackerman Air Traffic Control Task. On trial one, around 28% of participants used the sophisticated strategy. The proportion of participants using the Hold 1 strategy increased monotonically with practice with the rate of increase monotonically decelerating with practice. The proportion of participants using the Hold 1 strategy on the final trial was around 48 percent.

Based on the above theoretical considerations both the Michaelis–

Menten and Saturating Exponential functions should provide a good fit at the group-level. This should be superior to the Logistic function. It also seems worthwhile to compare the fits of the above three functions on a task that allows for simple–sophisticated strategy shift to compare relative fit and also check the assumptions made above regarding starting and asymptotic levels of strategy use. Specifically, the following hypothesis is proposed.

Hypothesis 2.10 *The relationship between practice and strategy sophistication at the group-level is effectively modelled by a three parameter Michaelis-Menten function.*

In combination, as with practice and performance, these hypotheses assert that the relationship between practice and strategy sophistication at the group-level is a biased representation of the individual-level. The group-level relationship is the sum of individual curves, and these individual curves are predicted to have different functional forms and varying parameters within curves. If this is the case, then even if the group-level relationship corresponds to the functional form for some individuals, it would be inappropriate for others.

2.1.4 Strategy and Performance

Overview

This section examines the relationship between strategy use and performance. First the concept of strategy relative effectiveness is defined and some of the challenges of measuring it are discussed. Then models and empirical data on the effect of strategy shift on task completion time are presented. This section aims to highlight reasons why discontinuities would or would not be expected in the learning curve given that strategy shifts are sometimes abrupt.

Relative Effectiveness

In general, strategy relative effectiveness is the degree to which one strategy is more or less effective than an alternative. In theory strategy relative effectiveness may be defined relative to many different criteria including accuracy, speed, quality, or minimisation of cognitive load. However, for the purpose of this thesis, strategy relative effectiveness will generally be operationalised as the difference in expected task completion time for two strategies.

Clearly, strategy relative effectiveness depends on many factors; some of the most prominent include: (a) general properties of the strategies, (b) individual differences in general, (c) individual differences in how well each strategy has been learnt, (d) features of the task and context, and (e) amount of practice on the task and on each strategy. Despite the effect of these factors, researchers have often sought to estimate the absolute relative effectiveness of two strategies. Theoretical and empirical approaches can be contrasted.

Theoretical estimates can be derived from various models of task performance. Such models are typically grounded in broader frameworks of human action such as GOMS (Card et al., 1983), CPM-GOMS (Gray, John, & Atwood, 1993; Gray & Boehm-Davis, 2000), cognitive architectures (for a review see John, 2003), and industrial engineering approaches (Konz, 1987). Task analytic approaches are based on assumptions of the nature of information processing and validated using empirical data.

While theoretical estimates may be useful, they are only as accurate as the theory that is used to generate them. Such theories are validated using empirical data, and they generate predictions that are capable of empirical testing. They are also less developed in areas of specific interest to this thesis around learning and individual differences. Strategy relative effectiveness at the time of a strategy shift and during the learning process is critical to understanding continuity in the learning curve. Individual differences also lead to variability in performance, which is often only incorporated in heuristic ways such as the way that GOMS uses typing speed as a measure of individual differences or the way that it assumes that the skill has been learnt.

Empirical Estimates

Empirical estimates are calculated by obtaining a sample of performance times using the strategies of interest. The difference in task completion time for the two samples constitutes a measure of strategy relative effectiveness. However, as discussed extensively by Siegler and Lemaire (1997), empirical estimates of strategy relative effectiveness are often biased. Biases arise because observations are typically not sampled randomly. For example, participants with great skill in general may be more likely to use better strategies, or sophisticated strategies may be more likely to be used with practice after other learning has already occurred. As such the estimate of strategy relative effectiveness is biased due to the correlation between strategy use and another important predictors of task performance.

To highlight the problems of bias in measurement of strategy relative effectiveness, three examples from the literature are provided. First, F. J. Lee et al. (1995) compared trial performance on the Kanfer–Ackerman Air Traffic Control Task where a strategy either was or was not used. They treated the data from the 18 trials and 58 participants as 1044 observations and reported the correlation between use of the superior plane landing strategy and task performance. Such an estimate is biased because some individuals used the strategy more than others and the use of the strategy increased with practice.

Second, John and Lallement (1997), also on the Kanfer–Ackerman Air

Traffic Control task, compared trial performance across several classified strategies on the final trial. This approach dealt with the confounding effect of practice. However, it did not deal with the confounding effect of individual differences. It is quite plausible that participants who used the superior strategy would also be more skilled in other ways. In addition, by only examining the final trial it was not possible to estimate relative effectiveness at the time of the strategy shifts.

Third, Touron, Hoyer, and Cerella (2004) presented a graph of average reaction time for algorithm and retrieval shift over practice. However, the individuals that adopt retrieval early are quite different to those that adopt it later. Each of these examples highlights the challenge of getting an estimate of the relative effectiveness of a strategy for a particular individual at a particular instance in time.

In order to get unbiased estimates of strategy relative effectiveness Siegler and Lemaire (1997) proposed the *Choice / No-Choice* design (for another exmaple, see, Walsh & Anderson, 2009). The Choice / No-Choice design involves measuring strategy use and performance in multiple conditions. In the *choice* condition participants are allowed to choose the strategy that they wish to use. In the *no-choice* conditions—one for each strategy—participants are required to use a designated strategy. Comparing performance between strategies in different no-choice conditions is meant to yield an unbiased estimate of strategy relative effectiveness.

While the Choice / No-Choice design overcomes many problems with estimation of strategy relative effectiveness, the design is not suited to certain situations. First, the design can not be used to estimate relative effectiveness at the time of a strategy shift. It assumes that learning effects are not dominating estimates. Second, it can not be used when it is strategy relative effectiveness at the time of self-initiated strategy shift that is of interest. In such cases, the no-choice condition both informs the participant of the existence of a given strategy and instructs the individual to use the strategy. If, as is to be expected, self-initiated strategy shifts have particular dynamics related to discovery and exploration, such timings of strategy shift would not be observed with the Choice / No-Choice design.

In summary, there is at present no standard method for estimating the relative effectiveness of an old and new strategy at the time of a self-initiated strategy shift. Looking at individual-level differences may be a good approach, particularly where the task is constant over the onset point.

Models of Performance following Strategy Shift

Despite these difficulties, several researchers have proposed models of, and empirically estimated, strategy relative effectiveness within the context of the learning curve. Such models should address both the relative effectiveness immediately following a strategy shift, and also how strategy use improves with practice.

Perhaps the earliest of such models was proposed by Crossman (1959). He proposed that the learning curve resulted from changes in probabilities of strategy use over time. Strategies were assumed to result in a fixed performance level. Probability of strategy use was assumed to move towards faster strategies and the speed of this transition in probabilities was greater when the relative effectiveness was also greater. The problem with Crossman's model, however, was that first, the model confuses strategy use, a process, with performance. Strategies can be performed at various degrees of effectiveness both by individuals and over time. Second, Crossman did not measure strategy use and thus did not validate the model assumptions about shifting probabilities of strategy use. Subsequent research (e.g., John & Lallement, 1997) measuring strategy use has shown that the model makes incorrect predictions about the relationship between practice and strategy use.

A second set of models are concerned with algorithm–retrieval shift (Delaney et al., 1998; Logan, 1988; Palmeri, 1999; Rickard, 1999). While this thesis focuses on simple–sophisticated strategy shifts, models of algorithm–retrieval shift still provide useful insights. In general, studies have shown that both algorithm and retrieval strategies get faster with practice. In these tasks the retrieval strategy is typically substantially quicker than the algorithm even from early trials of use. Because strategy shift is typically gradual over a set of items the learning curve is assumed to be, and typically is, continuous.

Delaney et al. (1998), in particular, proposed the idea that the learning curve is best represented by power functions within strategies. The model takes strategy use as given and makes a prediction. Although they did not express it as such, the model can be expressed as follows:

$$f_{s2}(x, z; \theta) = (\theta_1 + \theta_2 z_i) x_i^{-(\theta_3 + \theta_4 z_i)} \quad (2.14)$$

$$= \begin{cases} \theta_1 x_i^{-\theta_3} & \text{if } z_i = 0 \\ \theta_2 x_i^{-\theta_4} & \text{if } z_i = 1, \end{cases} \quad (2.15)$$

where $f_{s2}(x, z; \theta)$ is expected task completion time for the i^{th} trial, given strategy use z_i (0 indicating no strategy use and 1 indicating strategy use), trial number x_i and the parameter values θ_1, θ_2, θ_3, θ_4. Delaney et al. (1998) found that this model predicted reaction time over and above a standard two parameter power function.

There are several reasons to critique the model proposed by Delaney et al. (1998). First, in algorithm–retrieval studies the main skill that is acquired is the ability to retrieve. It is not possible to retrieve without practice. Thus, by incorporating strategy use as a variable in prediction, it is unfair to compare predictions with models that only incorporate amount of practice. Second, it is meaningless to talk about predicted performance time using the retrieval strategy on early trials when an individual is unable to retrieve. In such a situation, if the relevant stimulus–response pair is not represented in memory, an individual will not be able to retrieve the correct answer. Third, it is unclear why practice with retrieval would increase algorithmic reaction time as implied by the model. Fourth, the two parameter power function makes the implausible prediction that task completion time will approach zero with infinite practice both for the algorithm and the retrieval strategy.

While substantial research exists on modelling algorithm–retrieval shift, there are several reasons to assume that different models are required for tasks involving simple–sophisticated strategy shifts. In addition to the possibility of abrupt shifts, evidence suggests that performance may temporarily decline immediately after shifting to a sophisticated strategy (e.g., Study 3 in Delaney et al., 1998).

The general dynamics of simple–sophisticated strategy shift can be seen in many domains of life. The researcher shifts from using a word processor to LaTeX and from a menu driven statistics package to one based on writing code. The student decides whether to shift from hunt-and-peck to touch typing (see Yechiam, Erev, Yehene, & Gopher, 2003). These are everyday examples where a temporary drop in performance follows a strategy shift before ultimately superior performance is attained. Yechiam et al. (2004) uses the concept of *escalation of commitment* (Staw, 1981) to explain why an individual may be reluctant to temporarily drop performance for uncertain future benefits. A related idea is that of local minima, whereby strategy use is seen as an optimisation problem where individuals can get caught in local minima.

Several studies are relevant to understanding relative effectiveness at the time of simple–sophisticated strategy shift. However, each one tends to be missing an element to make the desired inferences. Haider and Frensch (1999, 2002) present data from an alphabet string verification task.

The individual-level learning curves show discontinuities. The timing of these discontinuities varies between individuals. If Haider and Frensch's interpretation is correct, the difference between pre- and post- shift reflects the relative effectiveness of learning to ignore the first few letters in the verification task. However, because there is no direct measure of strategy use, the results are open to alternative interpretations.

Other studies have an appropriate design, but lack the necessary measurements or analyses. John and Lallement (1997) obtained trial-level strategy use and performance data but only compared performance across strategies on the final trial. Even though the reaction time data made it fairly clear, Yechiam et al. (2004) did not measure strategy use on each trial. In summary, more research is needed with trial-level measurement of strategy use and performance on tasks that have a simple–sophisticated strategy shift in combination with individual-level analysis.

Reconciling Abrupt Strategy Shift with Continuous Learning Curves

The immediately preceding section has summarised models and empirical evidence regarding the influence of strategy shift on the learning curve. If it is true that discontinuities in learning curves are rare yet abrupt strategy shifts are rather common, then some combination of factors must prevent strategy shifts from causing discontinuities. Strategy use then becomes a partial mediator along with other learning processes of the effect of practice on task completion time. Potential factors include: (a) the time to learn the strategy, (b) the occurrence of gradual strategy shifts, (c) the process of generalising the use of a strategy to variations in task features and context, and (d) the greater probability of the shift occurring early in practice.

It was previously proposed that discontinuities in the learning curve at the individual-level are rare (i.e., Hypothesis 2.4). First, this can be reconciled with abrupt strategy shifts by considering that strategy shifts are only sometimes abrupt. Furthermore, abruptness can mean abrupt onset would be less likely to lead to an abrupt performance shift. Second, as asserted previously some individuals do not shift strategy either because they always use the sophisticated strategy (i.e., Hypothesis 2.8) they never use the sophisticated strategy (i.e., Hypothesis 2.9), or changes in strategy use from trial to trial do not reflect a systematic shift (i.e., Hypothesis 2.6). Third, as asserted previously strategy shift is often gradual (i.e., Hypothesis 2.5).

Even when strategy shifts do occur and are identified, there are reasons to expect a lack of discontinuity in task completion times. There are several reasons for this. First, based on general principles of summing random variables, trial-to-trial variance in performance should be greater at the individual-level than at the group-level. In many instances this variance will be large in comparison to strategy relative effectiveness both immediately following the shift and once the new strategy has been learnt. Second, as asserted previously strategy shifts are more likely to occur at the start of practice (i.e., Hypothesis 2.7) when the rate of learning potentially due to other factors is at its maximum (i.e., this is consistent with Hypothesis 2.2) thereby increasing the trial-to-trial variance at the time of shift. Third, although it is difficult to measure empirically, it is assumed that following a strategy shift, strategy specific learning typically takes place. This means that the full strategy relative effectiveness is not realised until a period of practice has passed.

2.1.5 The Current Studies

In order to test the above hypotheses, three studies were conducted. For the purposes of the analyses in this chapter each study examined a single group of participants performing a keyboard text editing task over a period of practice. Performance was measured as the time to complete text editing changes. Other researchers have used text editing to study other aspects of skill acquisition (e.g., Armstrong, 2000; Card, Moran, & Newell, 1980; Cook et al., 1995; Harvey & Rousseau, 1995; S. J. Payne, Squibb, & Howes, 1990; Robertson & Black, 1983; M. Singley & Anderson, 1985, 1987).

Keyboard text editing was chosen as the task in this thesis for several reasons. First, text editing permits a variety of strategies. Such strategies can readily be classified as simple or sophisticated. Second, sophisticated strategies should be more asymptotically efficient, but may take time to acquire. Third, key logs permit automated trial-level measurement of strategies.

The three studies were designed to test the hypotheses under varying conditions, including degree of initial text editing training and the structure of the text editing requirements. Studies 1 to 3 involved a progressive increase in the amount of instruction provided to participants on sophisticated text editing strategies. Study 1 involved a consistent set of editing requirements across trials, whereas Study 2 and 3 varied the editing requirements. More details of variation across the three studies are described in the method.

The main expectations for the three studies are reflected in the hypotheses presented in this chapter. However, in addition, Study 1 trial-to-trial variance in both performance and strategy sophistication at the individual-level was expected to be less than in the other two studies because Study 1 involved both more edits per trial and consistent edits across trials. Use of sophisticated strategies was expected to be greatest in Study 3 and least in Study 1 based on the corresponding levels of strategy instruction.

2.2 Method

2.2.1 Overview

This section describes the method used in the three studies reported in this thesis. The three studies share many common features, but also differ in several respects. In order to describe the method of the three studies in a clear and concise way, common features are first presented followed by a description of the unique aspects of each study. Additional details are presented in Appendices A, B, and C.

2.2.2 Common Features

Overview

Each study involved participants completing a selection of individual difference measures, then performing a series of practice trials on a keyboard based text editing task lasting, depending on the study, between 30 and 60 minutes. On each trial of the text editing task key presses were recorded. Key presses allowed for measurement of strategy use and task completion time.

This chapter examines the relationship between practice, strategy sophistication, and task completion time. Methodological aspects related to individual differences is reserved until Chapter 3. Also, Study 3 had three conditions, one of which involved a continuous block of practice on the text editing task (i.e., the No Training Condition) and was broadly equivalent to Study 1 and Study 2. In contrast, the other two conditions (i.e., Control and Training) in Study 3 involved various mid-Practice manipulations. Only the No Training condition is described and analysed in this chapter. The other two conditions pertain to the effect of instructed strategy shifts and are described and analysed in Chapter 4.

The main differences between the studies are summarised in Table 2.1.

Table 2.1: Main Design Differences between the Three Studies.

Feature	Study 1	Study 2	Study 3
Sample size	63	154	154 (61 in No Training Condition)
Sample source	Personal networks	Undergraduates	Undergraduates
Experiment Duration	150 minutes	120 minutes	100 minutes
Amount of Practice	54 Trials	30 Minutes	30 blocks each lasting at least 60 seconds
Experimental manipulation	None	None	Training, No Training, and Control
Trial variation	Constant	Varying	Varying
Edits per trial	Six	One	One
Edits shown	On paper	On screen	On screen
Edit types	Delete, Replace, Cut and Paste,	Delete, Replace, Cut and Paste, Insert	Delete
Initial strategy instructions	Given list of keys but not told to use	Given list, demonstration of how to use	Theoretical training and practice using each key
Key list	On paper	Permanently on screen	Available at end of block
Practice Trials	One Practice Trial (no time limit)	None	None
Typing Test	10 Thumbs	Anglim Typing Test	Anglim Typing Test
Prior Experience	17 items	Two items	14 items
Personality Test	None	100 Item IPIP	50 Item IPIP
Ability Tests	ERV, IT, CC, NC, NS, CS, RT1v1, RT2, RT4v1	ERV, IT, CC, NC, NS, RT1v2, RT4v2	None

Note. The following initials for ability tests were used ERV: Extended Range Vocabulary Test (Ekstrom, French, Harman, & Dermen, 1976), IT: Inference Test (Ekstrom et al., 1976), CC: Cube Comparison Test (Ekstrom et al., 1976). NC: Number Comparison Test (Ekstrom et al., 1976), NS: Number Sort (Ekstrom et al., 1976), CS: Clerical Speed and accuracy (Ekstrom et al., 1976); RT1: Simple Reaction Time, RT2: 2-Choice Reaction Time, RT4: 4-Choice Reaction Time.

Variation in individual difference measures, task instructions, and task features allowed for an exploration of the generality of the propositions put forward in this thesis.

While a text editing task was used in each study, several design elements varied between the studies, including (a) initial instructions, (b) accessibility of the text editing keys, (c) editing requirements on each trial, (d) duration of practice and number of trials, (e) aspects of the participant interface (performance feedback; position and size of the editing text box), and (f) method of ending trials (trial-time out; ended by accurate completion versus participant initiated ending). This section focuses on the features of the text editing tasks that were common across studies. These include: (a) the key press information that was recorded, (b) general aspects of performance measurement, and (c) measurement of strategy sophistication.

Key Press Measurement

At the key press level, the following variables were measured for every key press: (a) block number (where appropriate), (b) trial number, (c) raw key press number: reset to zero at the start of each trial, (d) raw delay: milliseconds since the previous key press or, in the case of the first key press, since the start of the trial, (e) the key pressed, and (f) whether modifiers `Ctrl`, `Shift`, and `Alt` were held down at the time of the key press. The raw key press data file contained two types of key presses: key presses that involved the participant striking the key with their finger and key presses that resulted from holding a key down and letting the keyboard repeatedly fire. When `Ctrl`, `Shift`, and `Alt` were pressed repeatedly, the second and subsequent presses were removed, because it is equivalent to a single key press. Raw key press numbers were adjusted so that after removal of repeated modifier keys, the new key press numbers were in the format $1, 2, 3, ...$ without integer gaps. Likewise, adjusted delay was constructed from raw delay by aggregating the total time associated with an instance of repeated firing of modifier keys and adding this time to the next physical key press in the key log.

Strategy Sophistication

A measure of *strategy sophistication* was extracted from the key logs for each participant on each trial. All key presses were classified as either *sophisticated*, *simple*, or *neutral*. The following keys were classified as sophisticated: (a) `Ctrl+Left`, (b) `Ctrl+Right`, (c) `Ctrl+Down`, (d) `Ctrl+End`, (e) `End`, (f) `Ctrl+Home`, (g) `Home`, (h) `Ctrl+Up`, (i) `Shift+Down`, (j) `Shift+End`, (k) `Shift+Home`, (l) `Ctrl+Shift+Left`, (m) `Ctrl+Shift+Right`, (n) `Shift+Up`, (o) `Ctrl+Backspace`, and (p) `Ctrl+Del`. The following keys were classified as simple: (a) `Left`, (b) `Right`, (c) `Shift+Left`, (d) `Shift+Right`, (e) `Backspace`, and (f) `Del`. Any other keys were classified as neutral. Strategy sophistication at the trial-level was measured as

$$\frac{SO}{SO + SI}, \tag{2.16}$$

where SO and SI are the respective trial counts of sophisticated and simple key presses. Block-level strategy sophistication was measured as mean trial-level strategy sophistication for accurate trials in the block.

2.2.3 Study 1

Participants

An initial sample of 116 adults participated in Study 1. However, 14 (12%) participants were excluded because of corrupt or missing data, and 39 (34%) were excluded because their data was deemed invalid. An analysis of reasons for invalid data are presented in Appendix A.4. The size of the final retained sample for analysis was 63. Of the retained sample 44 were female and 19 were male. Median age was 22 (mean = 23, min = 18, max = 39).

Participants were recruited through personal networks. All participants were initially screened to ensure that they had (a) competence in English as defined by (i) the ability to carry out a conversation in English, (ii) having lived in Australia for at least eight years, (iii) and had been in an English speaking educational system; (b) experience using a word processor on average once per week; (c) normal or corrected to normal vision; and (d) no severe physiological problems that would impair movement of the hands and wrists.

Text Editing Task

The Study 1 text editing task was programmed in Visual Basic 6. The task was programmed specifically for the study and adapted from Anglim (2000b). Calls to the Windows API were used to improve accuracy of temporal measurement. In contrast to earlier versions of the text editing task used in our research group, this version was the first to log key presses of participants. The timing and type of all key presses were recorded. Participants used standard 104 key IBM/Windows keyboards and 17 inch CRT monitors.

Screenshots of the user interface are shown in Figure 2.1 and Figure 2.2. The display included an area for editing text. Below this was feedback and instructions on the keys to press to start and stop a trial. A counter displayed the number of seconds that had passed on the trial. The current trial number, the response time for the previous trial, and the accuracy for the previous trial were all displayed. Additional dialogue boxes displayed at the end of blocks and the experiment are shown in Appendix A.3.1.

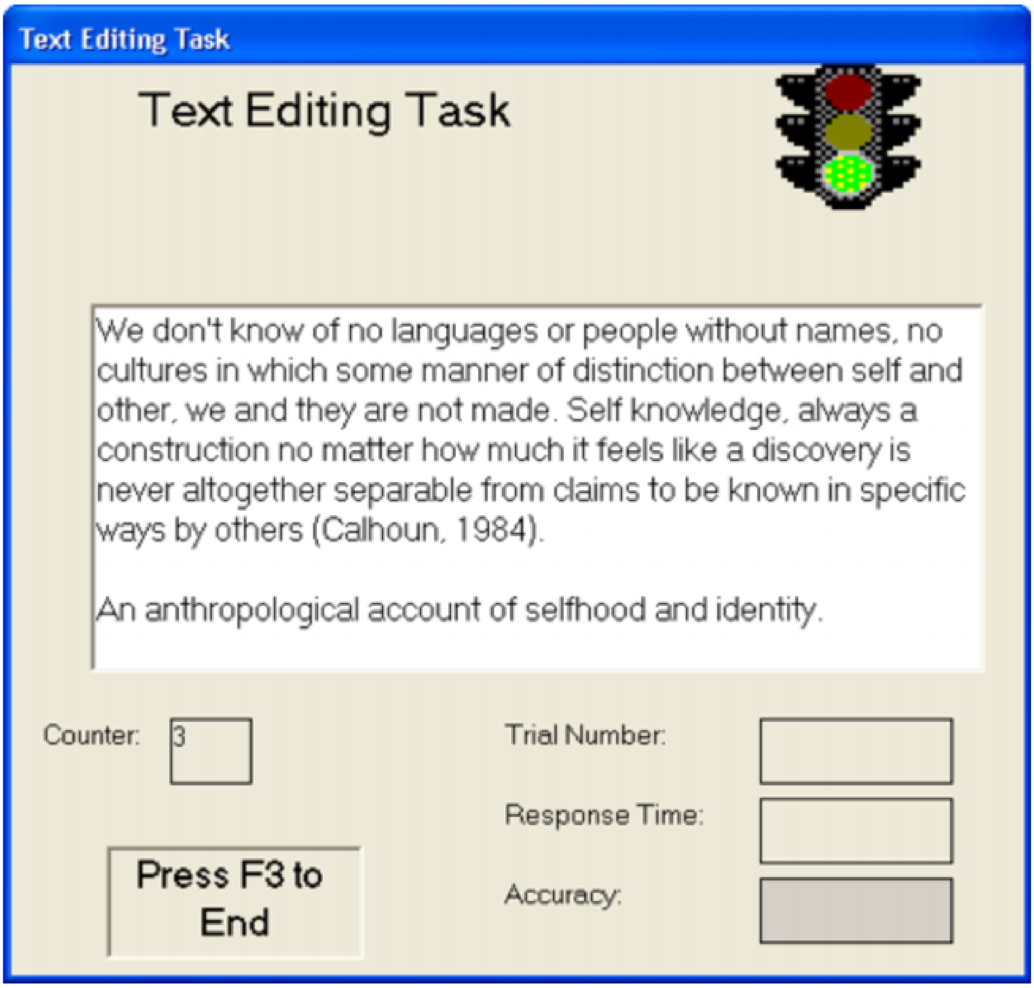

Figure 2.1: Study 1 text editing task screenshot in Active Mode.

Each trial of the task required participants to use the keyboard to make a set of text editing changes. The initial text included three long sentences. The task required six editing changes to be made. These involved deleting words, cutting and pasting words, selecting passages of text, and inserting characters. Table 2.2 shows the original and corrected text. Table 2.3 sets out the required changes. Participants received a piece of paper that indicated the required changes and a piece of paper with a list of text editing shortcut keys (see Table 2.4).

In Study 1 the set of text editing changes were identical for each trial.

Figure 2.2: Study 1 text editing task screenshot in Stop Mode.

Consistent editing changes meant that participants were more likely to achieve automaticity. It also minimised the time spent by participants reading the editing requirements on subsequent trials.

Each trial started when the participant pressed `F2`. This activated the text editor and started the timer. Trials ended either after the passage of one minute or when the participant pressed `F3` to indicate that the trial was complete. At the end of each trial, feedback was displayed on the time to complete the previous trial and the number of correct edits out of 12.

For each trial, task completion time and accuracy for each edit were

Table 2.2: Study 1 Original and Corrected Text for the Text Editing Task

Version	Text
Original	`We don't know of no languages or people without names, no cultures in which some manner of distinction between self and other, we and they are not made. Self knowledge, always a construction no matter how much it feels like a discovery is never altogether separable from claims to be known in specific ways by others (Calhoun, 1984).`  `An anthropological account of selfhood and identity.`
Corrected	`We know of no people without names, no languages or cultures in which some manner of distinction between self and other, we and they are not made. Self-knowledge, always a construction no matter how much it feels like a discovery is never altogether separable from claims to be known in specific ways by others (Calhoun, 1994).`  `An anthropological account of identity and selfhood.`

recorded. Information about the nature and timing of key presses was also recorded. Performance was measured as the time to complete the edits. Accuracy was out of 12 based of 12 textual checks of the final text as set out in Appendiex A.3.3. Appendix A.4 discusses how trials with imperfect accuracy were processed. Measurement of strategy sophistication was derived from the key logs as described in Section 2.2.2.

Procedure

Table 2.5 shows the sequence and duration of experimental tasks. Participants first completed the individual difference measures (analysed and described further in Chapter 3). Then after a break they completed the

Table 2.3: Study 1 List of Text Editing Task Requirements

Edit	Description
1.	Delete "don't"
2.	Cut "languages or" and Paste after the second "no"
3.	Insert "-" then backspace between "Self" and "knowledge"
4.	Delete "8" and insert "9"
5.	Cut "selfhood" and paste after "and"
6.	Cut "identity" and paste before "and"

text editing task.

Prior to starting the text editing task, instructions were read out verbatim as set out in Appendix A.3.2. In brief the instructions provided: (a) information on why text editing is a useful real-world skill; (b) an overview of the task; (c) an overview of the sequence of trials and breaks; (d) information about the list of short-cut keys; (e) reiteration not to use the mouse; and (f) a request to start the task. Participants were given a list of text editing short-cut keys as shown in Table 2.4. Participants were expected to read this list while they completed an initial practice trial.

The text editing task involved one practice trial and 54 performance trials. The practice trial had no time limit and was the same text as the main task. Performance was only recorded on the performance trials. Participants were encouraged to take short breaks after trials 18 and 36.

Table 2.4: Study 1 List of Shortcut Keys Provided to Participants

Key Combination	Action
Cursor keys	Moves cursor
Control-X	Cut
Control-C	Copy
Control-V	Paste
Control-Z	Undo
Shift	Select text
Shift & Control & Left or Right	Select word
Control & Left or Right	Move between words
Backspace	Delete letter to the left of cursor
Delete	Delete letter to the right of cursor
Control & Delete	Delete word after the cursor
Home	Move cursor to start of line
Shift & Home	Select text between where the cursor was and the start of the line
Control & Home	Move cursor to start of document
Shift & Control & Home	Select text between where the cursor was and the start of the document
End	Move cursor to end of line
Shift & End	Select text between where the cursor was and the end of the line
Control & End	Move cursor to end of document
Shift & Control & End	Select text between where the cursor was and the end of the document
F2	Start Trial
F3	End Trial

Table 2.5: Study 1 Experimental Protocol

Task	Duration
Individual Differences	
Introduction	2 min
Demographics	2 min
Study 1 Prior Knowledge Questionnaire	3 min
10 Thumbs Typing Test	1 min instructions; 2 min task
Inference Test	1 min instructions; 12 min task
Extended Range Vocabulary Test	1 min instructions; 12 min task
Cube Comparison Text	1 min instructions; 6 min task
Number Comparison Test	1 min instructions; 3 min task
Number Sort Test	1 min instructions; 3 min task
Clerical Speed and Accuracy	1 min instructions; 6 min task
Simple Reaction Time V1.0	1 min instructions; 4 min
2-Choice Reaction Time V1.0	1 min instructions; 5 min
4-Choice Reaction Time V1.0	1 min instructions; 5 min
BREAK	5 min
Text Editing Task	
Task Instructions	5 min
Practice Trial	2 min
Trials 1 to 18	approx 15 min
BREAK	5 min
Trials 19 to 36	approx 15 min
BREAK	5 min
Trials 37 to 54	approx 15 min
Debrief	

2.2.4 Study 2

Participants

Participants in Study 2 were drawn from students in a third year undergraduate psychology subject at an Australian university who consented to allow their involvement to be used for research purposes. The final retained sample of participants completing the text editing task was 154, drawn from an initial sample of 193 (i.e., 79.8% were retained). Reasons for removal of cases are presented in Appendix B.4. Of the retained sample 113 were female (73.9%) and 40 were male (26.1%) with one unknown. Median age was 21 (mean = 21.1, min = 18, max = 32).

Text Editing Task

The Study 2 text editing task was programmed using Visual Basic.Net. The design of the study and the nature of the task differed in several respects from Study 1.

Each trial involved making one of the following four types of edits: *Delete*, *Replace*, *Insert*, or *Cut and Paste*. Each trial had different editing requirements. Each trial was based on initial text taken from a passage from *Alice's Adventures in Wonderland* (L. Carroll, 1865). The corrected text is shown in Table 2.6. Table 2.7 shows the four trial types with information on the instructions and the details of the trial. A database of 400 items was created each with a different editing requirement. For each of the four trial types there were 100 items.

Table 2.6: Study 2 Corrected Text for Text Editing Task

Corrected Text
`Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversation?'`  `So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.`  `There was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually took a watch out of its waistcoat-pocket, and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop down a large rabbit-hole under the hedge.`  `In another moment down went Alice after it, never once considering how in the world she was to get out again. The rabbit-hole went straight on like a tunnel for some way, and then dipped suddenly down, so suddenly that Alice had not a moment to think about stopping herself before she found herself falling down a very deep well.`

Note. Text is taken from *Alice's Adventures in Wonderland* (L. Carroll, 1865).

Table 2.7: Study 2 Text Editing Trial Types

Trial Type	Instructions	Details of Task	Item Construction
Delete	Delete red text	Text appeared in red with a red strike through	One or more words was inserted (5 to 40 characters in length); these words needed to be deleted; position varied throughout base text.
Insert	After blue, type	Text was highlighted in blue; the word specified was to be typed after this point	One word was deleted from the base text and needed to be inserted
Replace	Delete red text and replace with	Text appeared in red with a red strike through; this was to be deleted and a single text was to be typed in its place	One or more words were inserted (3 to 20 characters) and these needed to be replaced typically with one word
Cut and Paste	Cut red text and paste after blue	Red text appeared on the screen as well as text that was highlighted in blue	A string of text was indicated to be cut and pasted into an insertion point. The text to be cut ranged from 12 to 344 characters.

The editing requirements were incorporated into the computer display. This removed the need for participants to look at a separate piece of paper to read the editing requirements. This modification made it easier to have different editing requirements on each trial, yet still allow participants to obtain a level of automaticity. It also made the editing task more similar to how individuals edit when making changes that they have mentally identified as opposed to responding to a commented draft with requested changes. Colour was used to indicate the type of editing change required.

A screenshot of the task interface is shown in Figure 2.3. At the top of the screen were instructions for the trial indicating the trial type. In the case of Insertion and Replace trials, the content of the insertion and replacement were displayed here. Below this was a tip about editing text which changed on each trial. Below this taking up the majority of the display and aligned to the bottom left of the screen was the text editing box. In the upper right was information on the trial number, the seconds remaining on the current trial, and the minutes of task time remaining in the experiment. Below this was a list of text editing keys. The mouse was disabled.

Each trial commenced with an initial three second delay where the instructions for the trial were displayed and the text editing box was visible but shaded in grey. After this delay the text editing box became editable, as was evident by its background turning white. This triggered the onset of trial completion time measurement. The cursor started in the top-left position of the text box. During the trial the time remaining on the trial counted down from 40 seconds. The trial ended automatically either when the participant successfully made the editing change or when 40 seconds had passed since the commencement of the trial.

The text editing tips were included to encourage participants to adopt new strategies. The 14 tips are presented in Table 2.8. The tip for each

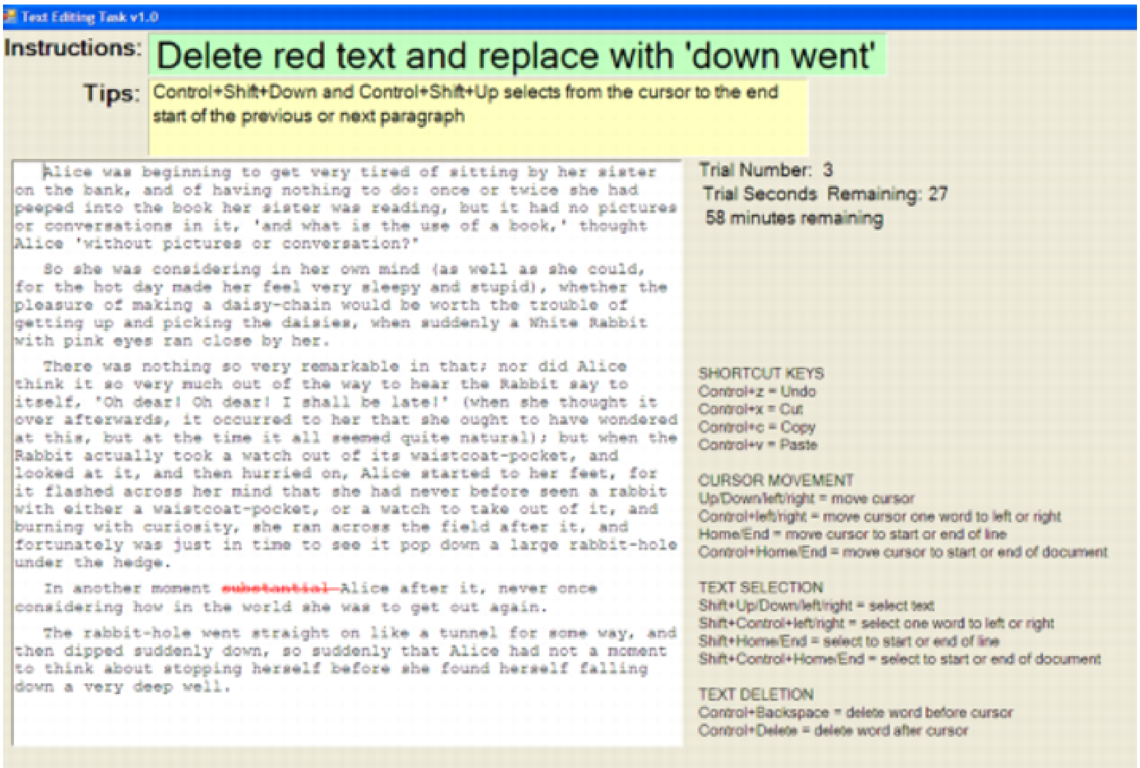

Figure 2.3: Study 2 text editing task screenshot

trial was randomly selected from the set of tips.

Trial-level performance was measured using a standardised form of task completion time. Before standardisation occurred several cases and trials were excluded for a range of reasons, details of which are presented in Appendix B.4. Adjusted task completion times were based only on retained data. Z-scores of task completion times for each trial type were generated. These standardised times were then rescaled to the grand mean and standard deviation of all retained trials. This is expressed mathematically in Appendix B.2.

Models at the individual-level used trial number as a predictor. Trial 1 was deemed to start on the second actual trial of the experiment. Trials that were not retained still contributed to the trial number. For example, if trial 4 was not retained for an individual, a participant's trial sequence would be 1, 2, 3, 5. That is to say, trial number 5 was not adjusted to

Table 2.8: Study 2 Text Editing Tips

	Text Editing Tip
1	Control+down moves the cursor to the start of the next paragraph; Control+up moves the cursor to the start of the previous paragraph
2	Control+z can be used to undo a change
3	Home Key moves the cursor to the start of the current line; End Key moves the cursor to the end of the current line
4	Control+Home moves the cursor to the beginning of the file; Control+End moves the cursor to the end of the file
5	Control+Shift+Left selects the word to the left; Control+Shift+Right selects the word to the right
6	Control+Left moves the cursor one word to the left; Control+Right moves the cursor one word to the right
7	Control+Delete deletes the word in front of the cursor, or the remainder of the word in front of the cursor if the cursor is half-way through a word
8	Control+backspace deletes the word behind the cursor
9	Shift+Down and shift+Up select whole lines of text
10	One way to delete text involves selecting text and then pressing delete
11	Shift+down selects the line below; Shift+up selects the line above
12	Control+Shift+Down and Control+Shift+Up selects from the cursor to the end start of the previous or next paragraph
13	When editing, place your right hand on the cursor keys. Index on Left and use to press Delete and Backsapace; middle on Up or Down and use to press Home and End; Ring finger on Right.
14	When editing place your left hand near left corner of keyboard. Little finger on left control; Ring finger on shift; Middle finger on z; and index finger use to press x, c or v.

be trial 4.

The study used a fixed duration for all participants. As such the number of trials varied between participants. In order to calculate group-level task completion time and strategy sophistication, trial numbers were rescaled within each individual as a percentage of the total trials for that individual that were completed. This percentage completion measure of trial number was then broken up into 30 blocks. Performance time and strategy sophistication at the block-level was the mean of the corresponding trial-level measures.

Procedure

Table 2.9 sets out the experimental protocol for Study 2. Participants were given a copy of the personality test to complete in their own time prior to the main experimental session. In the main experimental session participants completed the three general ability tests and the two perceptual speed tests. They were then given a 5 minute break.

Table 2.9: Study 2 Experimental Protocol

Task	Duration
One Week Prior to Main Experimental Session	
100 Item IPIP Personality Test	Approx 10 min
Main Experimental Session	
Ability Testing	
Inference Test	1 min instructions; 12 min task
Extended Range Vocabulary Test	1 min instructions; 12 min task
Cube Comparison Test	1 min instructions; 6 min task
Number Comparison Test	1 min instructions; 3 min task
Number Sort Test	1 min instructions; 3 min task
Text Editing Task	
Instructions and Training	10 min
Trials	30 min
One week after the Main Experimental Session	
Simple Reaction Time V2.0	1 min instructions; 4 min task
4-Choice Reaction Time V2.0	1 min instructions; 5 min task
Anglim Typing Test	3 min
Prior Experience Questionnaire	4 min

Participants were then given verbal text editing instruction made up of four parts: (a) motivating introduction, (b) overview of hand positioning, (c) follow-along demonstration of text editing keys, and (d) task specific instructions. The instructions aimed to motivate participants and give an initial exposure to each text editing key. Details of the instructions are provided in Appendix B.3. Participants then completed as many trials as they could in 30 minutes. Participants completed the psychomotor ability, prior experience, and typing test one week later. This chapter only analyses strategy sophistication and task completion time on the text editing task. Description and analysis of individual difference measures are presented in Chapter 3.

2.2.5 Study 3

Participants

Study 3 participants consisted of 154 adults recruited from a third year undergraduate psychology subject at an Australian university who consented to allow their data to be used for research purposes. All participants had a typing speed above 15 words per minute and completed the main performance task adequately. Participants included 109 females, 44 males, and 1 unknown. Median age was 21 (mean = 22.5, min = 19, max = 44.5). Participants were randomly allocated to one of three conditions. For the purposes of this chapter, only the No Training Condition is described and analysed. Discussion and analysis of the other conditions is presented in Chapter 4. The No Training Condition analysed in this chapter had 61 participants.

Text Editing Task

The text editing task in Study 3 was similar to Study 2. As with Study 2, the task was programmed in Visual Basic.Net. The program was run on Dell Optiplex GX520 computers with Intel Pentium 4, 3.00 GHz CPUs and 504 MB ram, running Windows XP Pro SP3. The displays were 15 in. LCD displays with 60 hz refresh rates set at a screen resolution of 1200 width by 1024 height. A screenshot of the main display is shown in Figure 2.4

Each trial required the participant to use the keyboard to delete a single continuous set of words marked in red. This required the participant to navigate the cursor to the location of the highlighted words and delete the words.

Each trial involved one randomly selected editing item drawn from a set of 500 items. Each item required the deletion of one to ten consecutive words. The number of words to be deleted on the trial was labelled the *deletion length*. The set of editing items was generated by taking an initial five-paragraph passage of text and adding *deletion text* after an *insertion point*. The deletion text consisted of words drawn from the 500 most common words in the English language (as sourced from, Fry, Kress, & Fountoukidis, 1993). Fifty insertion points were selected so that the deletion text was in different positions on different trials. Thus, the 500 items were generated by crossing the 50 insertion points with the ten deletion lengths. An example of the text to be edited with deletion text included is shown in Table 2.10. The complete list of the editing items is shown in Appendix C.2.4.

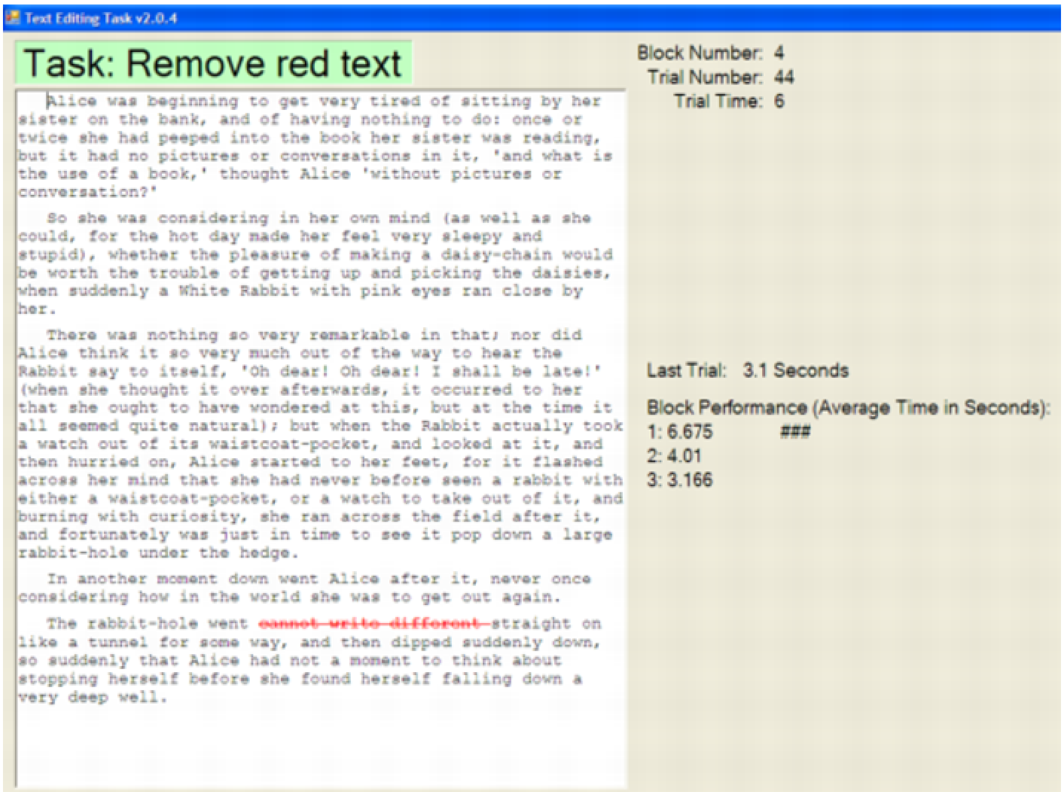

Figure 2.4: Text editing task during a representative trial. The task goal was presented to participants as "Remove red text". Information about the block number, trial number, and seconds past (Trial Time) are displayed in the top right of the display. Performance feedback is displayed on the middle-right of the screen. This feedback includes the number of seconds to complete the previous trial and the average time in seconds to complete all preceding blocks. In the space between top-right and middle-right, messages were displayed when certain events occurs. These included when there was no user activity for longer than five seconds and when the participant made an error related to having too many or too few spaces.

Table 2.10: Study 3 Example of Editing Requirements

Text to be Edited
Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversation?' So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her. There was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually took a watch out of its waistcoat-pocket, and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop down a large rabbit-hole under the hedge. ~~food war that sometimes ten from up sing problem~~ In another moment down went Alice after it, never once considering how in the world she was to get out again. The rabbit-hole went straight on like a tunnel for some way, and then dipped suddenly down, so suddenly that Alice had not a moment to think about stopping herself before she found herself falling down a very deep well.

Before each trial, there was a one second delay after which the editing box became active and the trial commenced. At the start of each trial the cursor was located at the top-left position of the text box. Trials ended either automatically when an edit was successfully completed or after 30 seconds had passed since trial commencement. At the end of each trial, task completion time for the trial that had just ended was displayed. There were also some additional messages displayed when participants engaged in off-task behaviour or made errors (see Appendix C.2.2).

Trials were grouped into blocks. After the passage of 60 seconds on a block, the ending of the active trial triggered the end of the block. Within a block, the end of each trial and the beginning of the next trial were separated by one second where no edits could be made. Block-level performance was the mean trial completion time in the block excluding inaccurate trials. This system was designed to ensure that the total length of the experiment was similar for each participant. This blocking structure was designed to: (a) get reliable block performance measurement, (b) have a sufficient number of blocks to model learning curves, and (c) enable trial-level modelling if desired without introducing block-level artefacts.

At the end of each block, participants were given feedback on the number of blocks out of 30 that they had completed and the average task completion time for accurate trials for the block that had just ended. Participants could then choose to press `ENTER` to continue to the next block or press `F1` to show a list of text editing keys. Figure 2.5 shows an example screenshot of the end-of-block feedback screen after pressing `F1`.

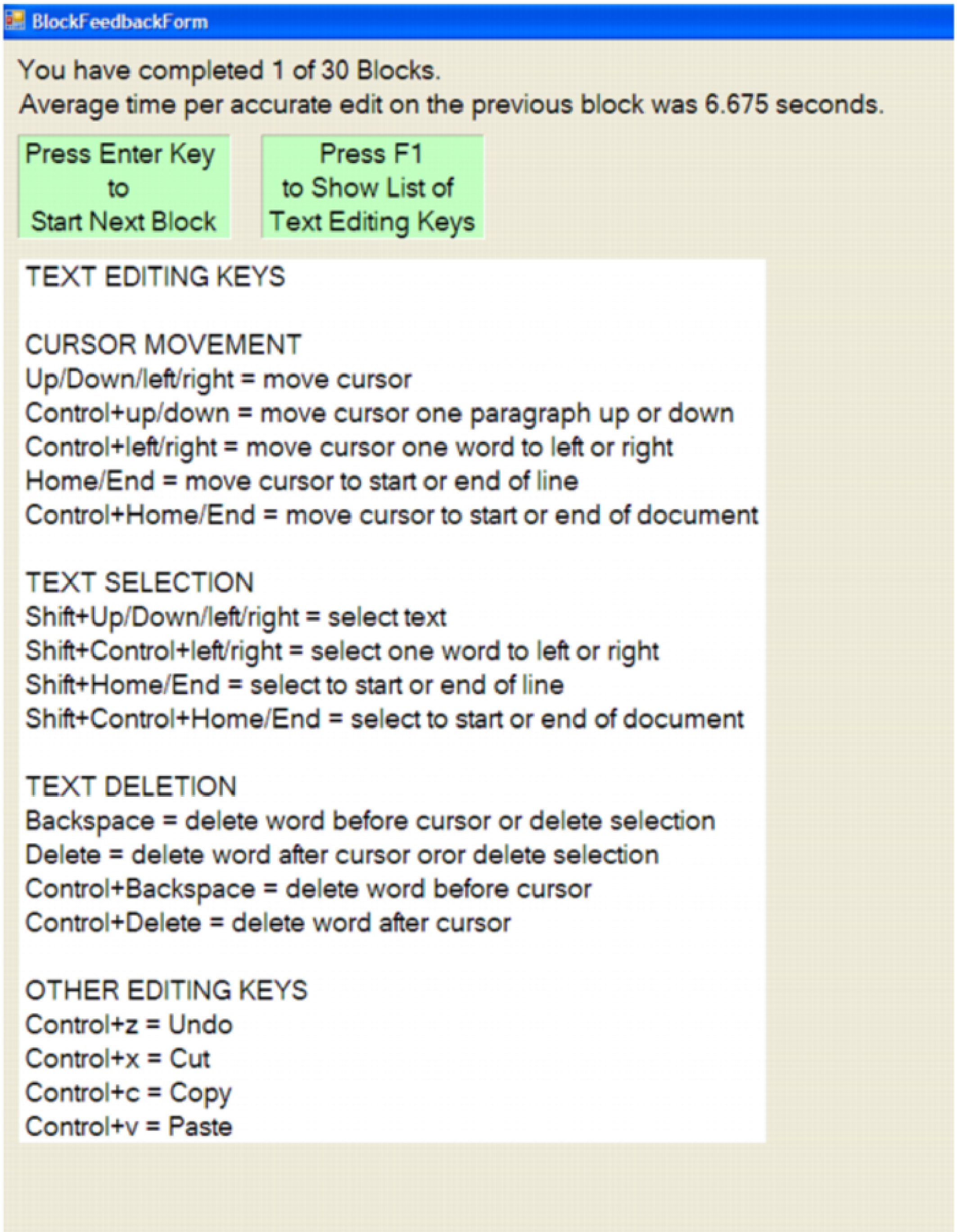

Figure 2.5: Study 3 text editing task screenshot at the end of a block. The display also shows the list of text editing keys displayed after a participant presses F1 at the end of a block.

The main measure of task performance was time to complete a trial. Inaccurate trials (i.e., those that timed out after 30 seconds) were excluded. In order to control for item effects, raw trial completion times were converted to z-scores within each item. Z-score-scaled trial completion times were then rescaled to the typical time for a trial throughout the experiment ($M = 6.2$ sec and $SD = 3.2$ sec). Block completion time for each individual was calculated as the mean of rescaled trial completion times for trials in the given block.

Procedure

Table 2.11 sets out the procedure for Study 3. The personality test was completed in a supervised setting one week prior to the main experimental session and administered on computer using `Inquisit v.3.0`.

Introductory Instructions Initial instructions were read out to participants. The verbatim text of the instructions is presented in Appendix C.2.1. The following summarises these instructions. First, participants were instructed to move to seats so that there was a gap between each participant. Participants were then informed of the nature of the task (keyboard text editing) and were given information about the value of the skill in educational and work domains in order to motivate them to learn. Participants were also told that the quality of the experimental results depended on them trying hard. Participants were then given an overview of the experiment and the time involved in each component. Participants were given a series of rules that they were requested to follow to ensure experimental conditions were maintained. Participants then completed the typing test, demographics, and prior experience measures before commencing the text editing task.

Table 2.11: Study 3 Experimental Protocol

Task	Approximate Duration
Two Weeks Prior to Main Experimental Session	
PLS and Consent	5 min
One Week Prior to Main Experimental Session	
50 Item IPIP Personality Test	10 min
Initial Section	
Introductory Instructions	10 min
Anglim Typing Test	3 min
Demographics Questionnaire	1 min
Prior Experience Questionnaire	1 min
Text Editing Task	
Instructions	4 min
Initial Training	4 min
Practice Trials Blocks 1 to 15	25 min
Mid-Practice Condition	
If Training: Mid-Practice Training	5 min
If Control: 4 Choice RT Task	5 min
If No Training: Continue to practice	0 min
Practice Trials Blocks 16 to 30	25 min
Participant Debriefing	

Note. Exact timing of most sections varied based on participant behaviour.

Initial Text Editing Instructions Instructions for the text editing task were administered on computer. The exact wording of these instructions is presented in Appendix C.2.1. In summary, participants were told: (a) of the importance of reading the instructions carefully; (b) the nature of the task and the real-world value of text editing; (c) to not use the mouse and place it behind the computer; (d) about the instructions and structure of the task including number of blocks, relationship between trials and blocks, nature of trial pausing, and what to do if an error is made; (e) about the rules to follow in order to ensure experimental rigour; (f) about proper finger placement for effective text editing; and

(g) about what to do on the initial training trials.

Initial Training Participants then completed initial training trials on keyboard text editing. Each practice trial introduced one or more text editing key combinations as listed in Table 2.12. Each trial displayed a practice paragraph where participants were required to apply the introduced text editing keys. Participants could only proceed to the next trial once they had pressed each of the keys a specified number of times. This ensured that participants did not skip the practice period.

Table 2.12: Study 3 Keys Practiced in Initial Training Trials

Trial	Keys Practiced
1.	`Up`, `Down`, `Left`, `Right`
2.	`Ctrl+Left`, `Ctrl+Right`
3.	`Ctrl+Up`, `Ctrl+Down`
4.	`Home`, `End`
5.	`Ctrl+Home`, `Ctrl+End`
6.	`Shift+Up`, `Shift+Down`, `Shift+Left`, `Shift+Right`
7.	`Ctrl+Shift+Left`, `Ctrl+Shift+Right`
8.	`Shift+Home`, `Shift+End`
9.	`Backspace`, `Del`
10.	`Ctrl+Backspace`, `Ctrl+Del`
11.	`Ctrl+X`, `Ctrl+C`, `Ctrl+V`, `Ctrl+Z`

Once participants completed the training trials, the following message was displayed:

> Key instructions have ended. You can bring up the list of text editing keys at the end of each block. On the next screen the main task begins. Your task is to use text editing keys to remove the text marked in red as quickly as possible. Press Enter To Continue.

Main Task Participants in the condition presented in this chapter (i.e., the No Training condition) then completed 30 blocks of trials of the text editing task without interruption. Details about the other conditions are described in the method section of Chapter 4.

2.2.6 Analysis Plan

The following section sets out the quantitative methods used to test the proposed hypotheses. The general orientation combined model fitting with extensive use of graphics. Models were broadly evaluated in terms: (a) fit to the data, (b) parsimony as indicated by fewer free parameters, and (c) theoretically meaningfulness as indicated by plausible predictions outside the range of the data. All analyses were performed using `R 2.11.1` (R Development Core Team, 2010). In addition to `base R` the following `R` packages were used: (a) `psych` (Revelle, 2010) for some descriptive statistics and test scoring, (b) `lattice` (Sarkar, 2010) for generating trellis plots of individual-level learning curves. (c) `Sweave` (Leisch, 2002) for integrating tables, figures, and results text into the thesis in a reproducible manner.

All functions proposed in the introduction for modelling practice and strategy use and practice and task completion time were modelled using nonlinear least squares regression (for theoretical treatment, see Bates & Watts, 1988; Seber & Wild, 2003) using the `nls` function (Bates & Watts, 1988) in `R` (for discussion of use of nls, see Huet et al., 2004; Ritz & Streibig, 2008).

Estimation was performed using the `nls` function in `R` (Bates & Watts, 1988) using the Gauss-Newton algorithm. For each model, several different starting strategies were adopted, which involved (a) estimating the full model, (b) estimating a set of models with one parameter fixed for each model but varying over the set of models, (c) fixing one or more parameters that would otherwise be free to vary. The resulting model with the smallest residual sums of squares was retained as the actual model for subsequent model summary purposes. These alternative strategies were particularly important because (a) individual-level data was often noisy, (b) model estimation was desired even for models that were clearly not representative of the data generating process, and (c) some of the models had a relatively large number of parameters relative to the number of data points.

The models estimated and reported in this thesis are all designed to incorporate parameters that lead to plausible predictions outside the range of the data and reflect reasonably plausible psychological processes. This raises a range of issues regarding how possible values of parameters should be constrained to enforce the above properties. While other parameter constraints might be considered, in this thesis the only constraint applied to parameter estimates was that parameters representing breakpoints were forced to occur after the fifth block/trial (note Study 1 uses trials, and Study 2 and 3 uses blocks of trials), and before the fifth last block/trial. Such breakpoint parameters occur in CnSE, CnCn, P2BP2, and E3B models. This meant for example that the CnSE model was required to predict constant strategy sophistication at least in the first five blocks, whereas without this constraint, it could discard the constant component and perfectly replicate the SE model. This constraint was employed to increase the chance that the breakpoint would have a psychologically meaningful interpretation, and was not just used to give additional freedom to model fitting.

When comparing models, smaller AIC and root mean squared error (RMSE or just se) values were used to indicate model superiority. Where two main models had the same number of parameters, R^2 was also used as a basis for comparison. At the individual-level, paired-samples t-tests were used to compare the RMSEs for selected model comparisons: (a) three parameter power with three parameter exponential; (b) four parameter power with APEX; and (c) APEX with three parameter exponential with breakpoint. Also, a chi-square test was used to assess whether the number of fits at the individual-level supporting one model compared to another was greater than chance.

The subsequent results section is at first instance organised around studies. Within each study results for practice and performance are presented followed by results for practice and strategy sophistication. Within each of these sections group-level, and then individual-level results are presented. Within each of these subsections, graphical and model fitting approaches are presented.

2.3 Study 1 Results

2.3.1 Practice and Performance

Group-level

Figure 2.6 shows the group-level relationship between practice and task completion time. Despite a few bumps the plot shows a monotonically decreasing (consistent with Hypothesis 2.3) and decelerating function that approaches an asymptote towards the end of practice.

Table 2.13 shows the model fits statistics for group-level data. Consistent with Hypothesis 2.1, the three parameter power function provided better fit than the three parameter exponential function as evident in the larger R^2 and smaller AIC.

To determine whether results were broadly robust to the particular accuracy adjustment performed, several additional analyses were performed. Models were fit to raw reaction time, and the same general pattern persisted when comparing three parameter power with three parameter exponential models with power being superior at the group-level and exponential being superior at the individual-level. While there are reasons to believe accuracy adjusted reaction times are a more valid measure of performance on the task, core results were robust whether this adjustment was applied or not.

Table 2.13: Study 1 Fit Statistics for Models of the Effect of Trial on Task Completion Time at the Group-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	6.039	.000	350.44
2.	P3	3	0.730	.986	124.24
3.	E3	3	1.384	.949	193.27
4.	APEX	4	0.738	.986	126.22
5.	P4	4	0.737	.986	126.07
6.	CnCn	3	3.379	.699	289.65
7.	P2BP2	5	0.693	.988	120.44
8.	E3B	5	1.114	.969	171.66

Note. Model name abbreviations were previously defined in Section 2.1.2
[a] k = Number of parameters.

Individual-level

Description and Plots Figure 2.7 shows the relationship between practice and task completion time for each participant. As expected, trial-to-trial variation in performance is relatively small but still greater than the group-level. All participants improved with practice. For most participants the rate of improvement appears to be monotonically decelerating. However, some participants seem to approach a personal asymptote earlier than others, and there are a few cases where improvement is almost linear (e.g., cases 14, 33).

Figure 2.8 shows examples of cases for each of the modelled functions. Cases were chosen that generally fit well for the function relative to other individuals.

Power versus Exponential Function Table 2.14 summarises individual-level model fit information for the relationship between practice and task completion time. Supporting Hypothesis 2.2, the exponential function provided substantially superior fit in comparison to the power function ($rmse_{P3} = 5.24, rmse_{E3} = 5.10$). Using a paired samples t-test, this was a statistically significant difference, $t(62) = 2.46, p = .02$.

In terms of individual cases, the exponential function beat the power function in 39 cases, compared to only 24 wins for the power function. This was almost statistically significant when compared to a chi-square goodness of fit test assuming 50% wins for both functions, $\chi^2(df = 1) = 3.57, p = .06$.

APEX versus P4 Function Table 2.14 summarises individual model fit information for three parameter power and exponential functions. There was little difference in the fits of the four parameter power and the APEX function, $rmse_{APEX} = 5.19, rmse_{P4} = 5.19$. Using a paired samples t-test, this was not a statistically significant difference, $t(62) = 0.144, p = .89$. In terms of individual cases, the APEX function beat the power function in 33 cases, compared to 30 wins for the power function. This was not a statistically significant difference when compared to a chi-square goodness of fit test assuming 50% wins for both functions, $\chi^2(df = 1) = 0.143, p = .71$.

E3B versus APEX Function The three parameter exponential break function provided superior fit in comparison to the APEX function ($rmse_{E3B} = 4.73, rmse_{APEX} = 5.19$). Using a paired samples t-test, this was a statistically significant difference, $t(62) = 6.12, p < .001$. In terms of individual cases, the E3B function beat the APEX function in 58 cases, compared to 5 wins for the APEX function. This was statistically significant difference when compared to a chi-square goodness of fit test assuming 50% wins for both functions, $\chi^2(df = 1) = 44.59, p < .001$.

Table 2.14: Study 1 Mean Model Fit Statistics for the Effect of Trial on Task Completion Time at the Individual-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	8.207	.000	370.30
2.	P3	3	5.239	.562	323.53
3.	E3	3	5.100	.579	321.17
4.	APEX	4	5.194	.595	323.10
5.	P4	4	5.186	.581	323.14
6.	CnCn	3	5.754	.484	335.63
7.	P2BP2	5	4.754	.646	315.74
8.	E3B	5	4.730	.647	315.16

Note. Number of individual learning curves = 63.
[a] k = Number of parameters.

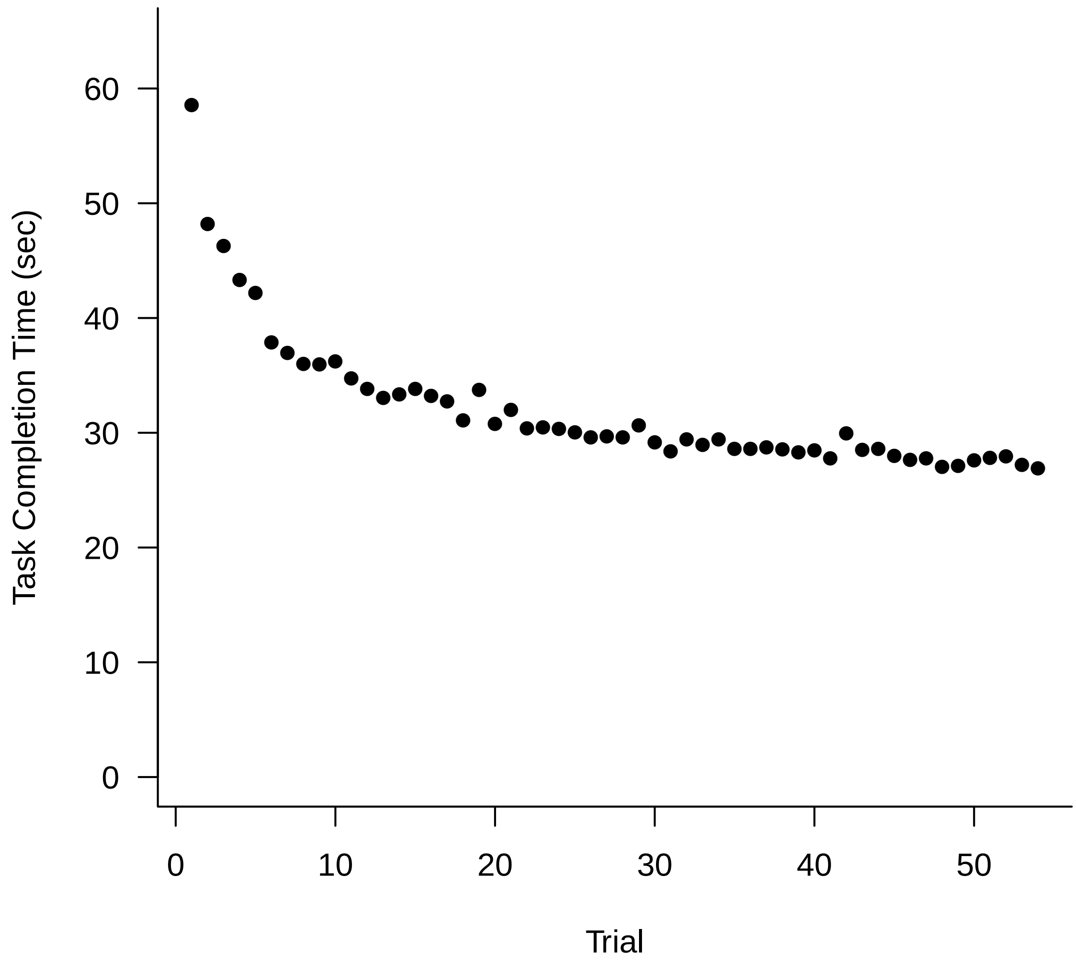

Figure 2.6: Study 1 task completion time by trial at the group-level.

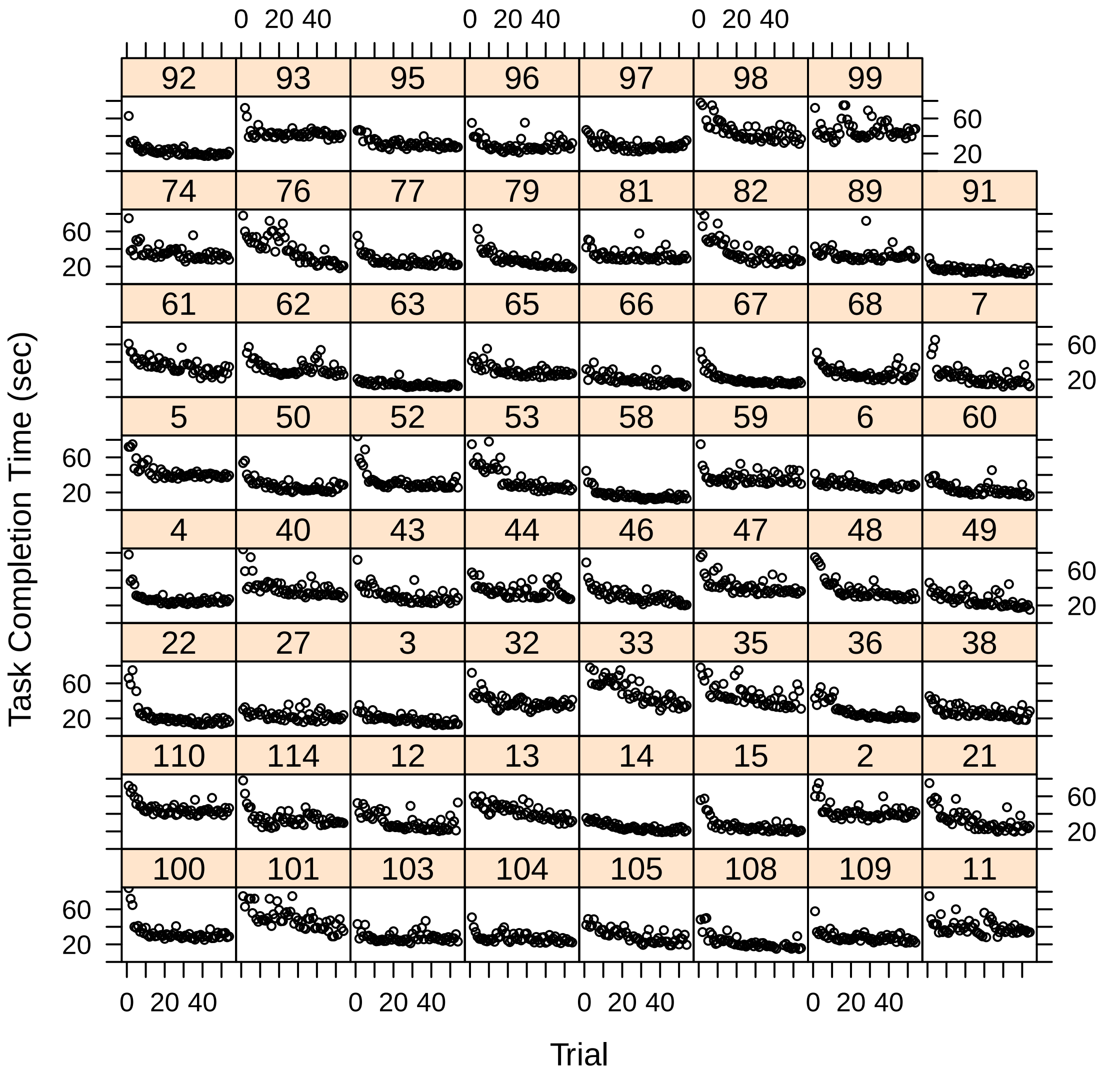

Figure 2.7: Study 1 task completion time by trial at the individual-level.

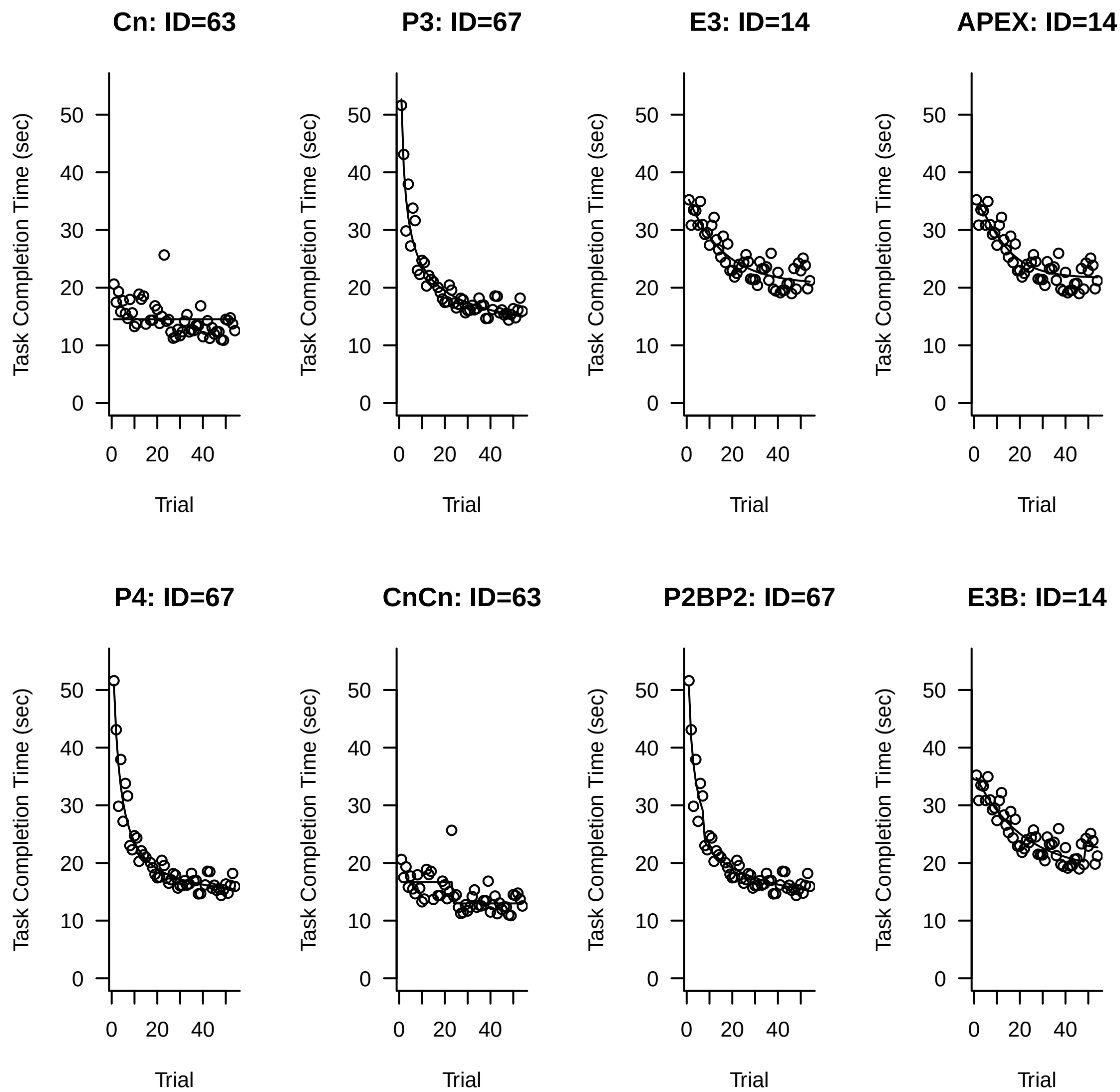

Figure 2.8: Study 1 examples of model fits of the relationship between practice and task completion time at the individual-level.

2.3.2 Practice and Strategy

Group-level

Figure 2.9 shows the relationship between practice and strategy sophistication at the group-level. The relationship is approximately linearly increasing with a slight deceleration of the rate of change over the course of practice. Model fits are shown in Table 2.15. Consistent with Hypothesis 2.10 the Michaelis–Menten function provides a good fit. The almost linear increase in strategy sophistication is somewhat surprising given that task completion time rapidly approached an asymptote. In contrast, the strategy sophistication curve suggests that substantially more improvement would have occurred had the period of practice been longer. While a linear model would provide reasonable fit, it is not reported because of the previously discussed problems associated with polynomial models, such as their failure to predict effectively outside the range of the data. It is also worth remembering that the lower r-squared observed for the CnSE model relative to the SE model is due to the parameter constraint on the CnSE model that required the breakpoint to occur at least five trials into practice.

At the group-level many of the models provided similar levels of fit. However, the Michaelis–Menten and Saturated Exponential function seemed the most reasonable. There was no obvious constant at the start ruling out the Constant–Saturated Exponential model, and there was no obvious S-shape, ruling out the Logistic model.

Individual-level

Figure 2.10 shows the relationship between practice and strategy sophistication at the individual-level. Table 2.16 summarises the support for the proposed models at the individual-level.

Consistent with Hypothesis 2.9, several participants never used so-

Table 2.15: Study 1 Fit Statistics for Models of the Effect of Trial on Strategy Sophistication at the Group-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	0.060	.000	-147.58
2.	MM	3	0.009	.980	-354.14
3.	SE	3	0.009	.980	-354.66
4.	Lg	3	0.009	.980	-355.50
5.	CnCn	3	0.029	.771	-223.24
6.	CnSE	5	0.009	.980	-352.07

Note. Model name abbreviations were previously defined in Section 2.1.3
[a] k = Number of parameters.

Table 2.16: Study 1 Mean Model Fit Statistics for the Effect of Trial on Strategy Sophistication at the Individual-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	0.088	.000	-225.87
2.	MM	3	0.060	.299	-249.46
3.	SE	3	0.056	.364	-253.04
4.	Lg	3	0.052	.308	-254.70
5.	CnCn	3	0.048	.435	-264.03
6.	CnSE	5	0.052	.358	-207.07

[a] k = Number of parameters.

phisticated strategies (e.g., cases 93, 96, 97). Consistent with Hypothesis 2.8 some participants almost always used sophisticated strategies (e.g., cases 11, 110). Some participants shifted to sophisticated strategies fairly quickly in the early trials (e.g., cases 3, 58, 63). Consistent with Hypothesis 2.5 the speed of transition varied between participants. Constant–Constant suggests an abrupt onset and shift. Constant–Saturating Exponential suggests an abrupt onset but a more gradual shift. Consistent with Hypothesis 2.6 several participants only partially adopted sophisticated strategies (e.g., cases 35, 46, 68). Figure 2.11 illustrates for each function one participant with the model fit overlayed.

Broadly consistent with Hypothesis 2.7, many of the participants who increased in their strategy sophistication throughout practice started their increase in the first trial (e.g., cases 7, 22, 79, 105). However, there was also a large number of participants where the onset of a strategy shift occurred halfway through practice. A small number of participants appear to have more than one strategy onset point (e.g., cases 43, 49). Also a couple of participants briefly used more sophisticated strategies before reverting back to simpler strategies. Participant 74 provides a particularly clear example of this.

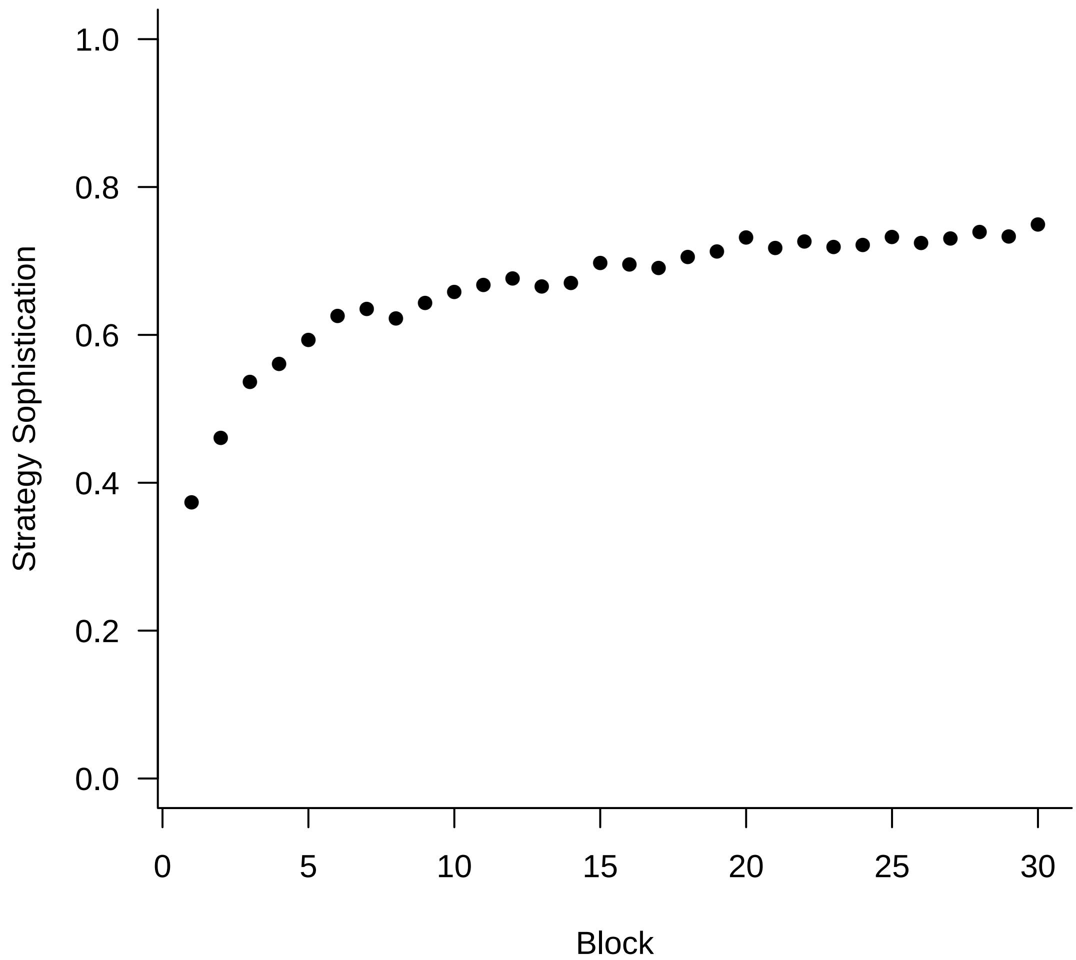

Figure 2.9: Study 1 strategy sophistication by trial at the group-level.

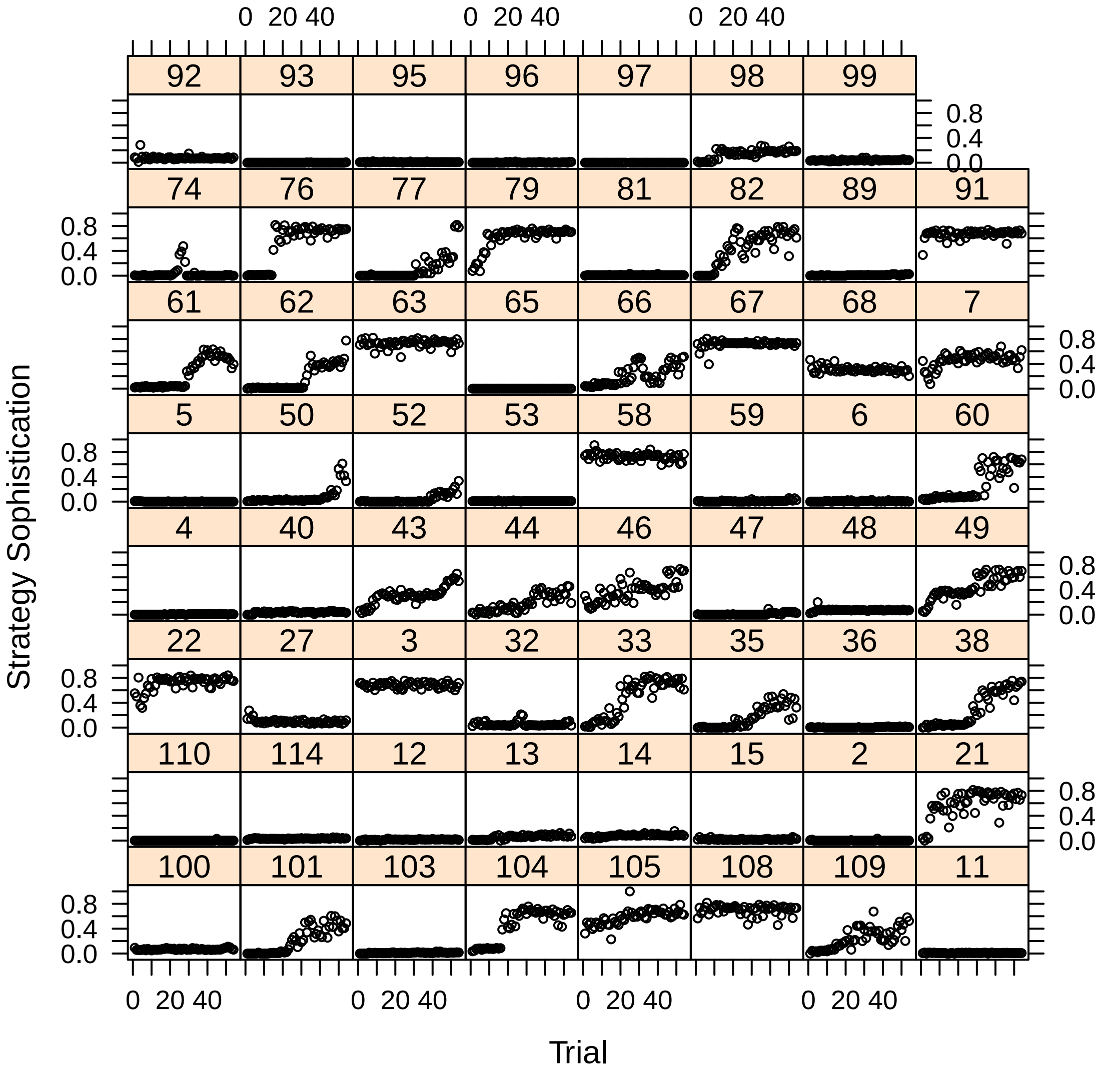

Figure 2.10: Study 1 strategy sophistication by trial at the individual-level.

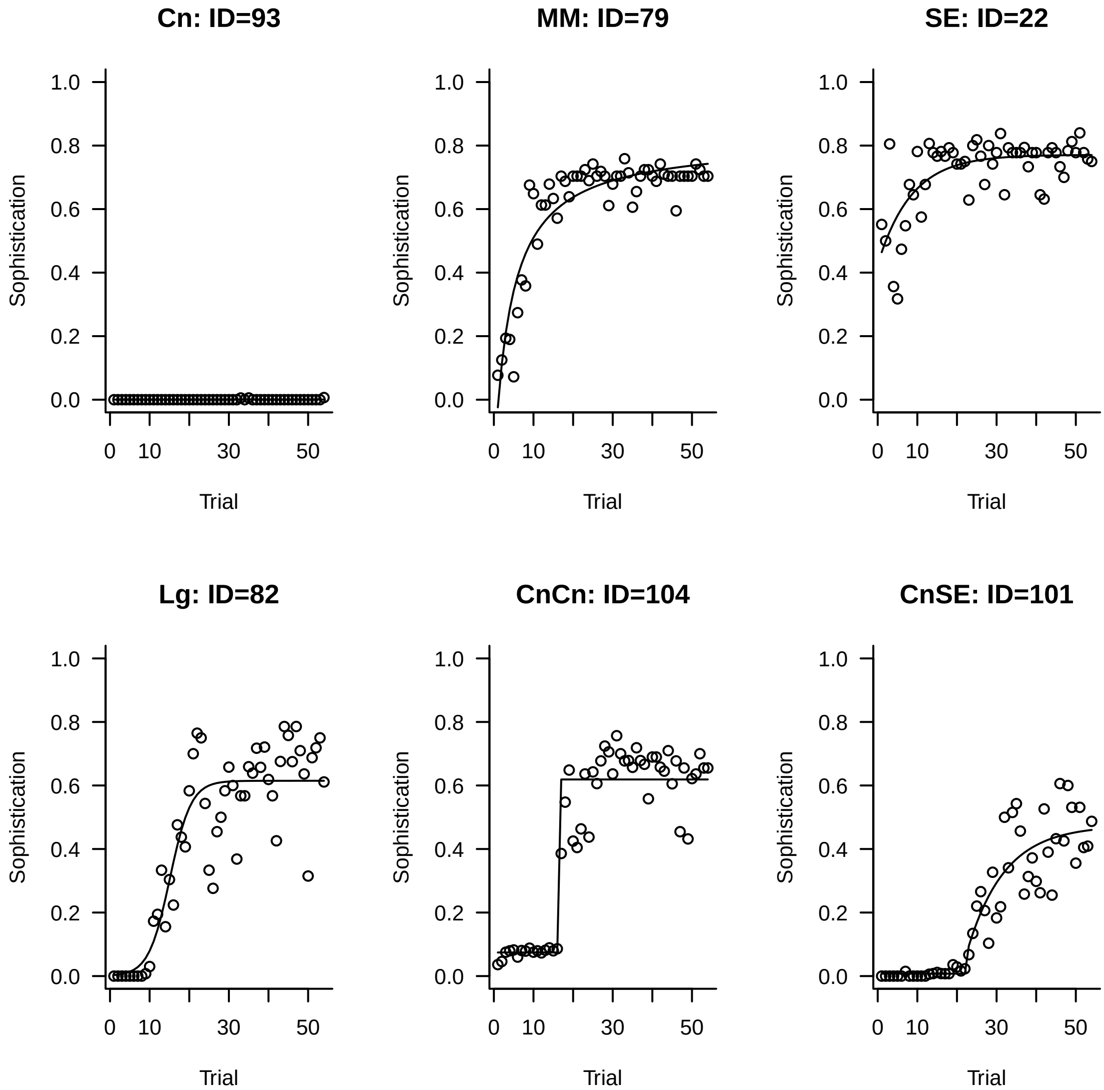

Figure 2.11: Study 1 examples of model fits of the effect of trial on strategy sophistication at the individual-level.

2.3.3 Practice, Strategy, and Performance

Cases with standard deviations above 0.10 were selected for correlation analysis. Of the 63 cases, 23 satisfied this criteria. Within individual correlations between log task completion time and strategy sophistication were calculated for each of these participants separately. Correlations tended to be large and negative (M = -0.55, SD = 0.22, min = -0.86, max = -0.07).

2.4 Study 2 Results

2.4.1 Practice and Performance

Group-level

Figure 2.12 shows the group-level relationship between practice and task completion time. Besides a few minor deviations, the figure shows data representative of a monotonically decreasing and decelerating function typical of group-level learning curves.

Table 2.17 shows the model fits statistics for group-level data. Consistent with Hypothesis 2.1 the three parameter power function provided superior fit relative to the three power exponential function.

Table 2.17: Study 2 Fit Statistics for Models of the Effect of Block on Task Completion Time at the Group-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	1.281	.000	102.98
2.	P3	3	0.159	.986	-20.52
3.	E3	3	0.281	.955	13.72
4.	APEX	4	0.162	.986	-18.25
5.	P4	4	0.161	.986	-18.85
6.	CnCn	3	0.777	.657	74.87
7.	P2BP2	5	0.171	.985	-14.19
8.	E3B	5	0.236	.971	4.91

[a] k = Number of parameters.

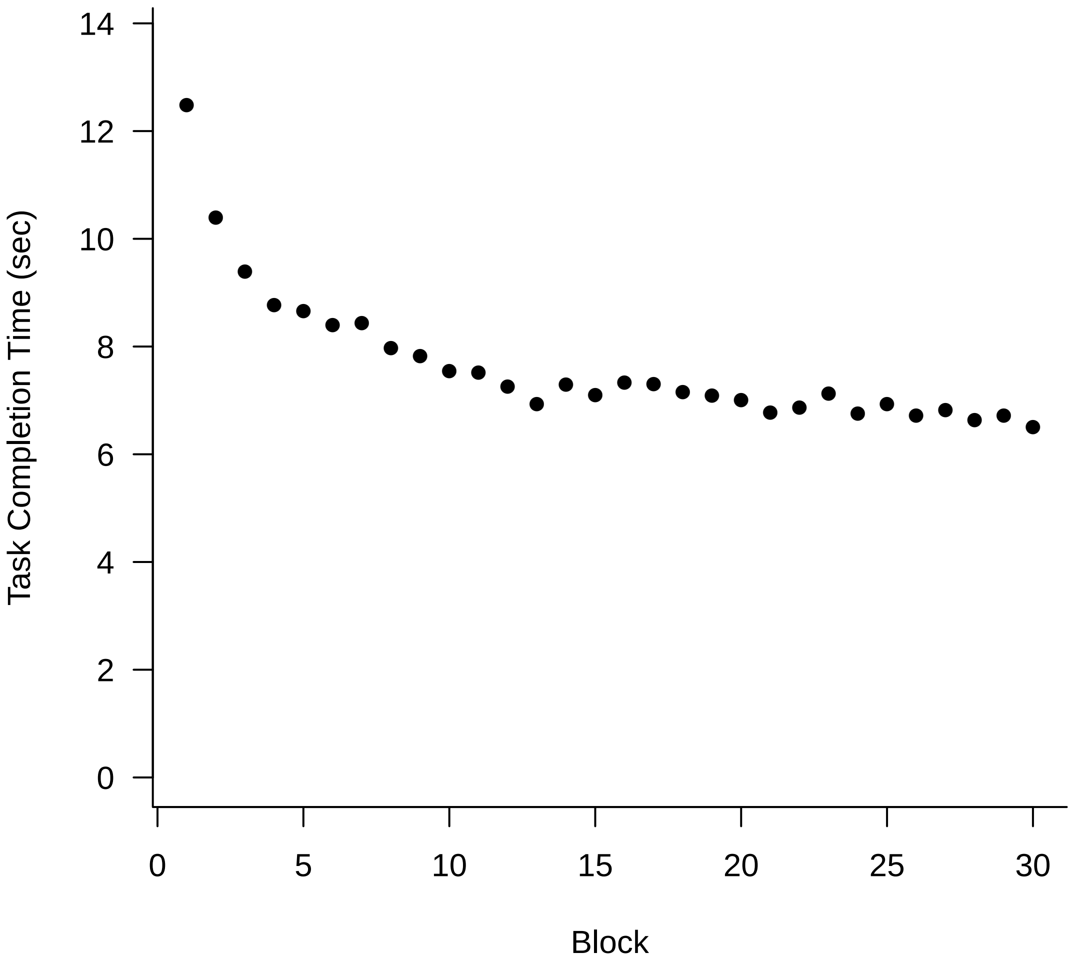

Figure 2.12: Study 2 task completion time by block at the group-level.

Individual-level

Description and plots Figures 2.13 and 2.14 show the relationship between practice and task completion time at the individual-level. Trial-to-trial variation was substantially larger than Study 1. This can be attributed to trials with only a single edit and the variable editing content. Learning curves varied between individuals in several respects including (a) initial performance, (b) final performance, (c) shape of the learning curve, (d) the frequency and timing of outliers, and (e) the consistency of performance from trial to trial. Figure 2.15 shows example function fits for individual participants.

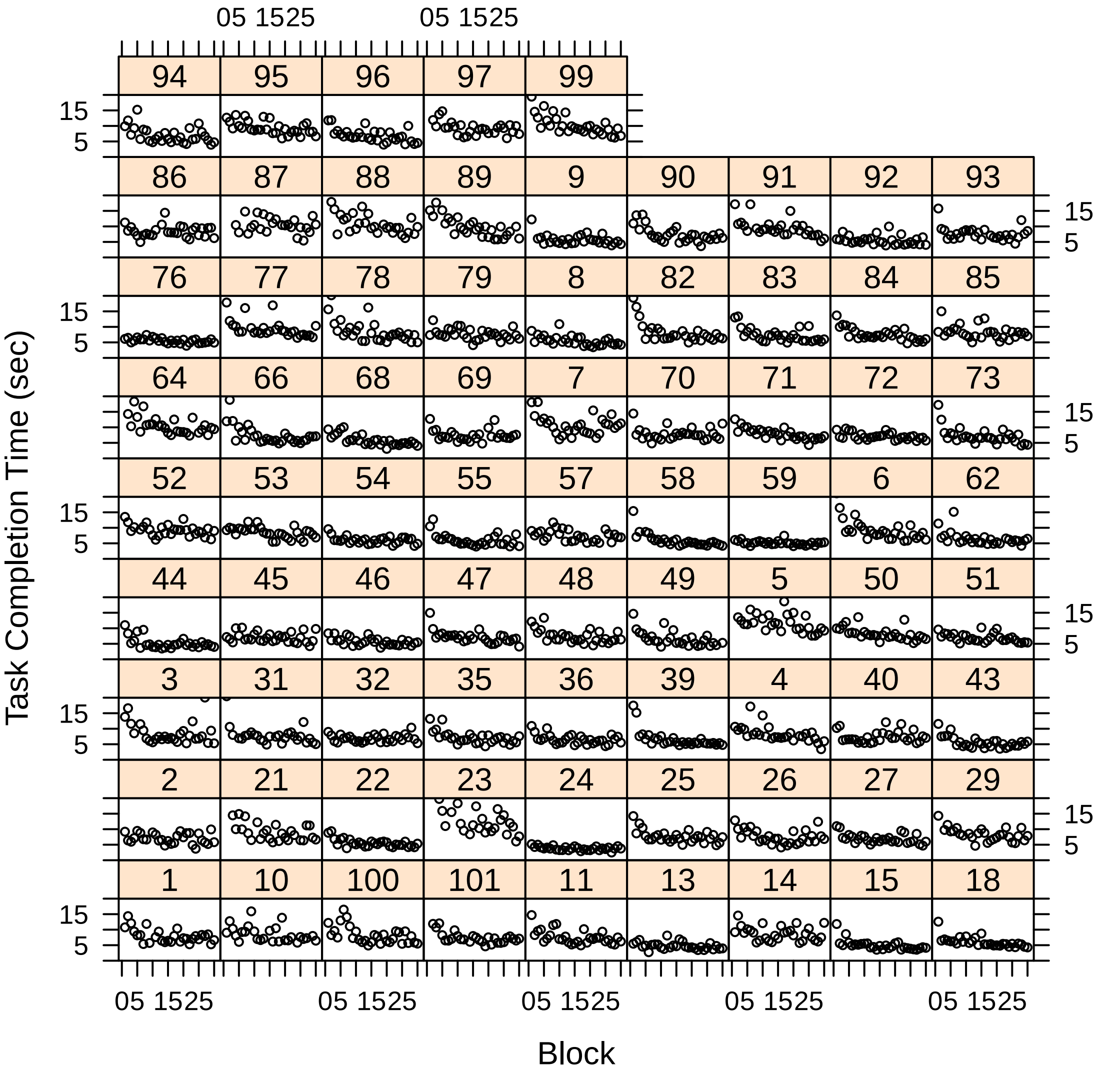

Figure 2.13: Study 2 task completion time by block at the individual-level — Part 1 of 2.

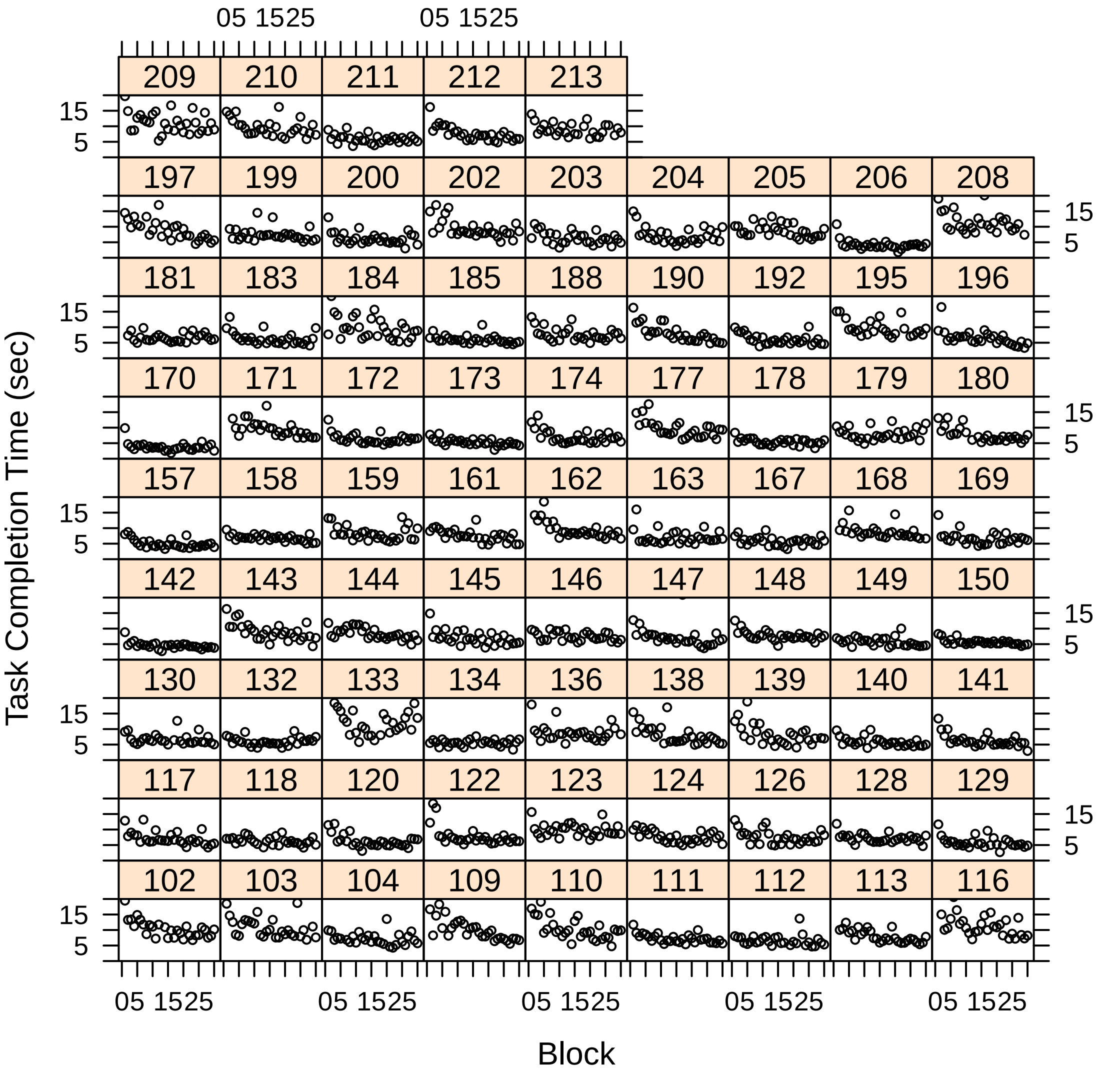

Figure 2.14: Study 2 task completion time by block at the individual-level — Part 2 of 2.

Power versus Exponential Function Table 2.18 summarises individual-level model for the relationship between practice and task completion time. Supporting Hypothesis 2.2, the exponential function provides superior fit in comparison to the power function ($rmse_{P3} = 1.75, rmse_{E3} = 1.74$). Using a paired samples t-test, this was a statistically significant difference, $t(153) = 2.93, p = .004$. In terms of individual cases, the exponential function beat the power function in 96 cases, compared to only 58 wins for the power function. This was significantly different using a chi-square goodness of fit test assuming 50% wins for both functions, $\chi^2(df = 1) = 9.38, p = .002$.

APEX versus P4 Function The APEX function had superior fit to the four parameter power function ($rmse_{APEX} = 1.73, rmse_{P4} = 1.78$). Using a paired samples t-test, this was a statistically significant difference, $t(153) = 4.73, p < .001$. In terms of individual cases, the APEX function beat the power function in 110 cases, compared to 44 wins for the power function. This was a statistically significant difference when compared to a chi-square goodness of fit test assuming 50% wins for both functions, $\chi^2(df = 1) = 28.29, p < .001$.

E3B versus APEX Function The three parameter exponential break function provided superior fit in comparison to the APEX function, ($rmse_{E3B} = 1.66, rmse_{APEX} = 1.73$). Using a paired samples t-test, this was a statistically significant difference, $t(153) = 6.89, p < .001$. In terms of individual cases, the E3B function beat the APEX function in 133 cases, compared to 21 wins for the APEX function. This was a statistically significant difference when compared to a chi-square goodness of fit test assuming 50% wins for both functions, $\chi^2(df = 1) = 81.45, p < .001$.

Table 2.18: Study 2 Mean Model Fit Statistics for the Effect of Block on Task Completion Time at the Individual-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	2.224	.000	129.55
2.	P3	3	1.755	.385	117.06
3.	E3	3	1.738	.397	116.68
4.	APEX	4	1.734	.424	117.32
5.	P4	4	1.776	.397	118.68
6.	CnCn	3	1.856	.331	120.84
7.	P2BP2	5	1.652	.501	115.36
8.	E3B	5	1.659	.489	115.53

Note. Number of individual learning curves = 154.
[a] k = Number of parameters.

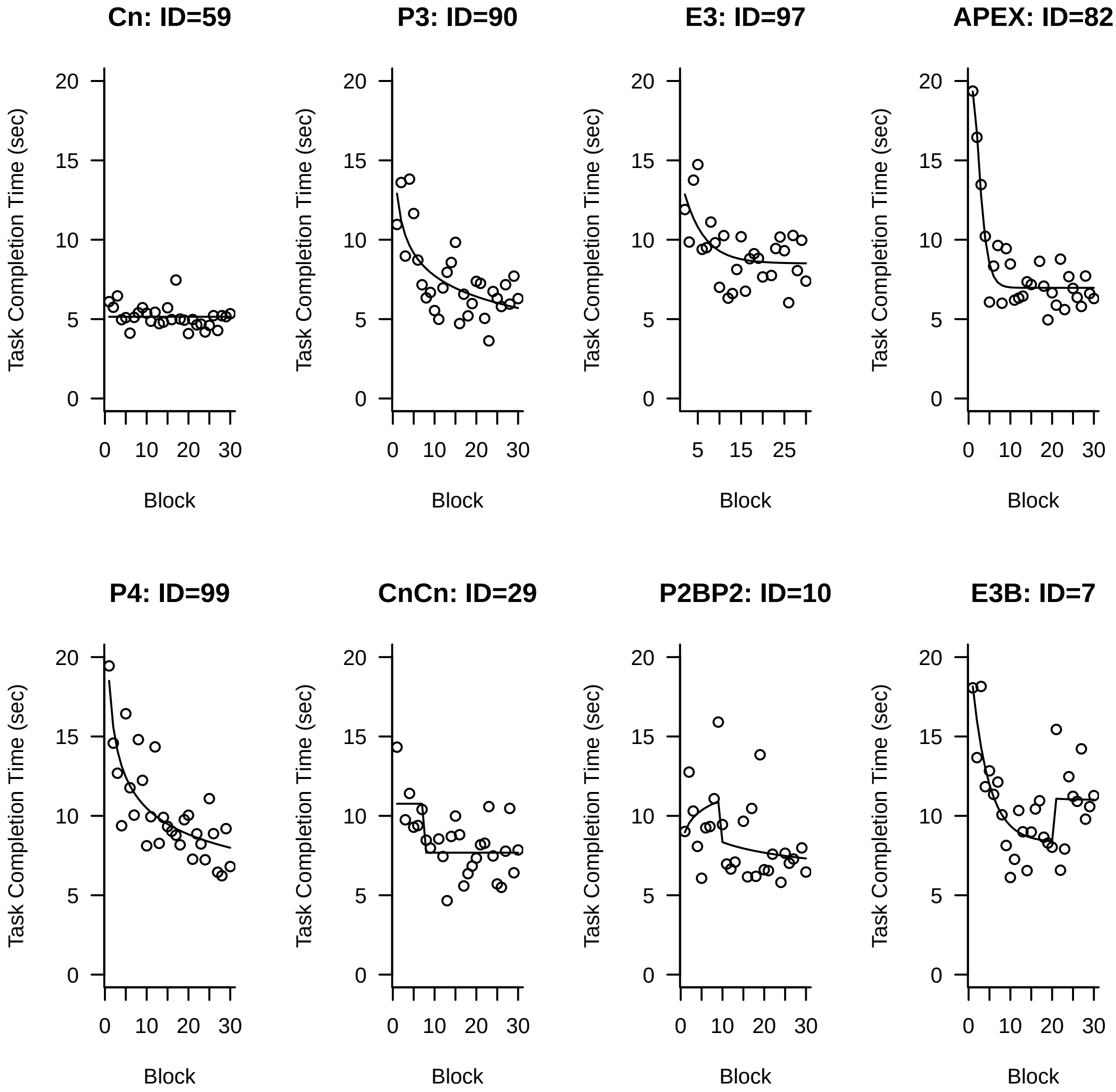

Figure 2.15: Study 2 examples of model fits of the relationship between block and task completion time.

2.4.2 Practice and Strategy

Group-Level

Figure 2.16 shows the relationship between practice and strategy sophistication at the group-level. Both the initial and final levels of strategy sophistication was much greater than in Study 1. Table 2.19 provides a table of model fit statistics. Consistent with Hypothesis 2.10 a Michaelis-Menten function provided a good fit to the data at the group-level. The data did not appear to exhibit an S–shape.

Table 2.19: Study 2 Fit Statistics for Models of the Effect of Block on Strategy Sophistication at the Group-Level.

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	0.077	.000	-65.95
2.	MM	3	0.014	.971	-168.00
3.	SE	3	0.014	.969	-166.64
4.	Lg	3	0.015	.965	-163.12
5.	CnCn	3	0.038	.773	-106.41
6.	CnSE	5	0.025	.910	-130.24

[a] k = Number of parameters.

Individual-Level

Figure 2.17 and 2.18 shows the relationship between practice and strategy sophistication at the individual-level. Table 2.20 summarises the fit at the indiviudal-level. The individual-level strategy sophistication is both greater and more noisy than Study 1.

Consistent with Hypothesis 2.9 several participants consistently had strategy sophistication around zero (e.g., cases 54, 85). Consistent with Hypothesis 2.8 some participants had consistently high levels of strategy sophistication (e.g., cases 132, 149, 206). Some shifted to sophisticated strategies fairly quickly in the early trials (e.g., cases 112, 120). Consistent with Hypothesis 2.5 the speed of transition varied. Constant–

Table 2.20: Study 2 Mean Model Fit Statistics for the Effect of Block on Strategy Sophistication at the Individual-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	0.196	.000	-14.88
2.	MM	3	0.171	.235	-20.86
3.	SE	3	0.169	.248	-21.08
4.	Lg	3	0.170	.234	-20.66
5.	CnCn	3	0.163	.287	-22.77
6.	CnSE	5	0.177	.227	-8.52

[a] k = Number of parameters.

Constant suggests an abrupt onset and shift. Constant–Saturating Exponential suggests and abrupt onset but a more gradual shift. Consistent with Hypothesis 2.6 several individuals only partially adopted sophisticated strategies (e.g., cases 122, 123, 167). Figure 2.19 illustrates for each function one participant where that function fit relatively well. Consistent with Hypothesis 2.7, the onset of the strategy shifts was greatest in the first block. In broad terms, the frequency of shifts decreased after this.

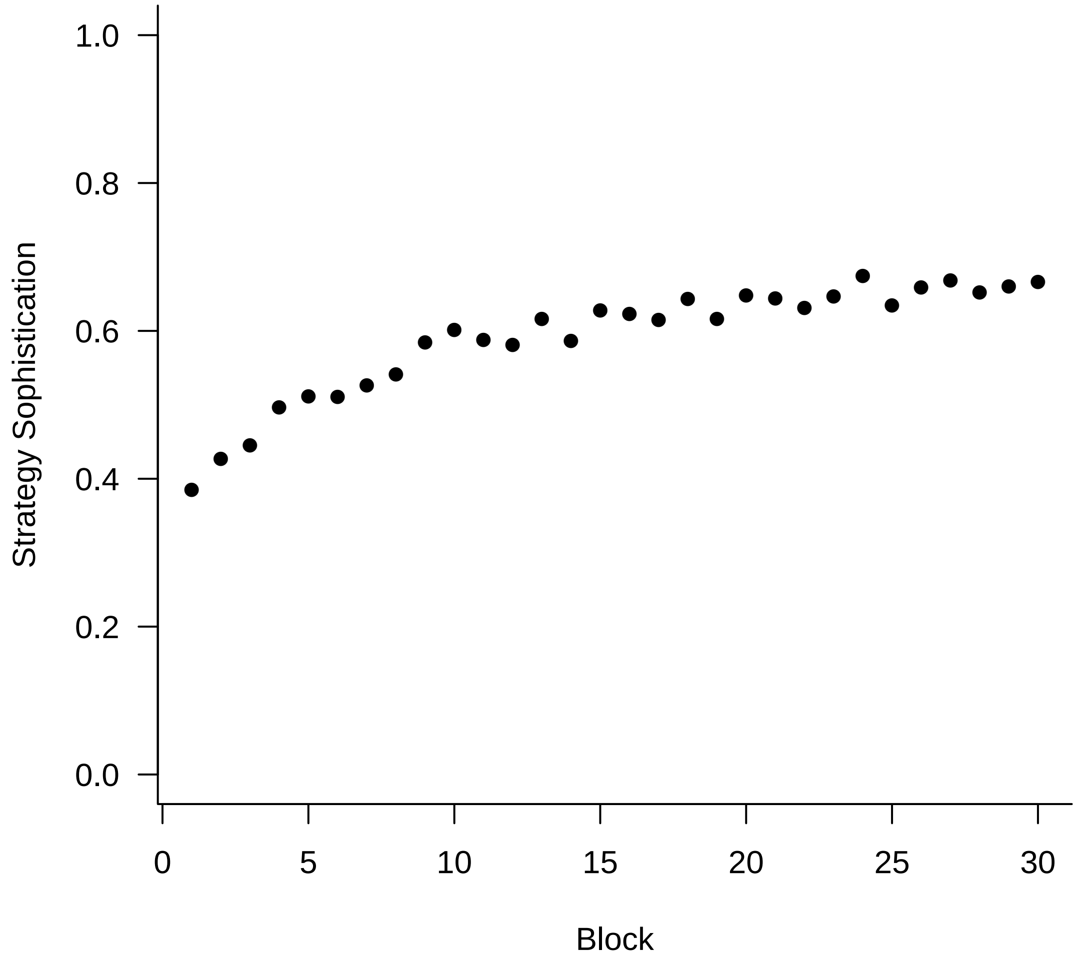

Figure 2.16: Study 2 strategy sophistication by trial at the group-level.

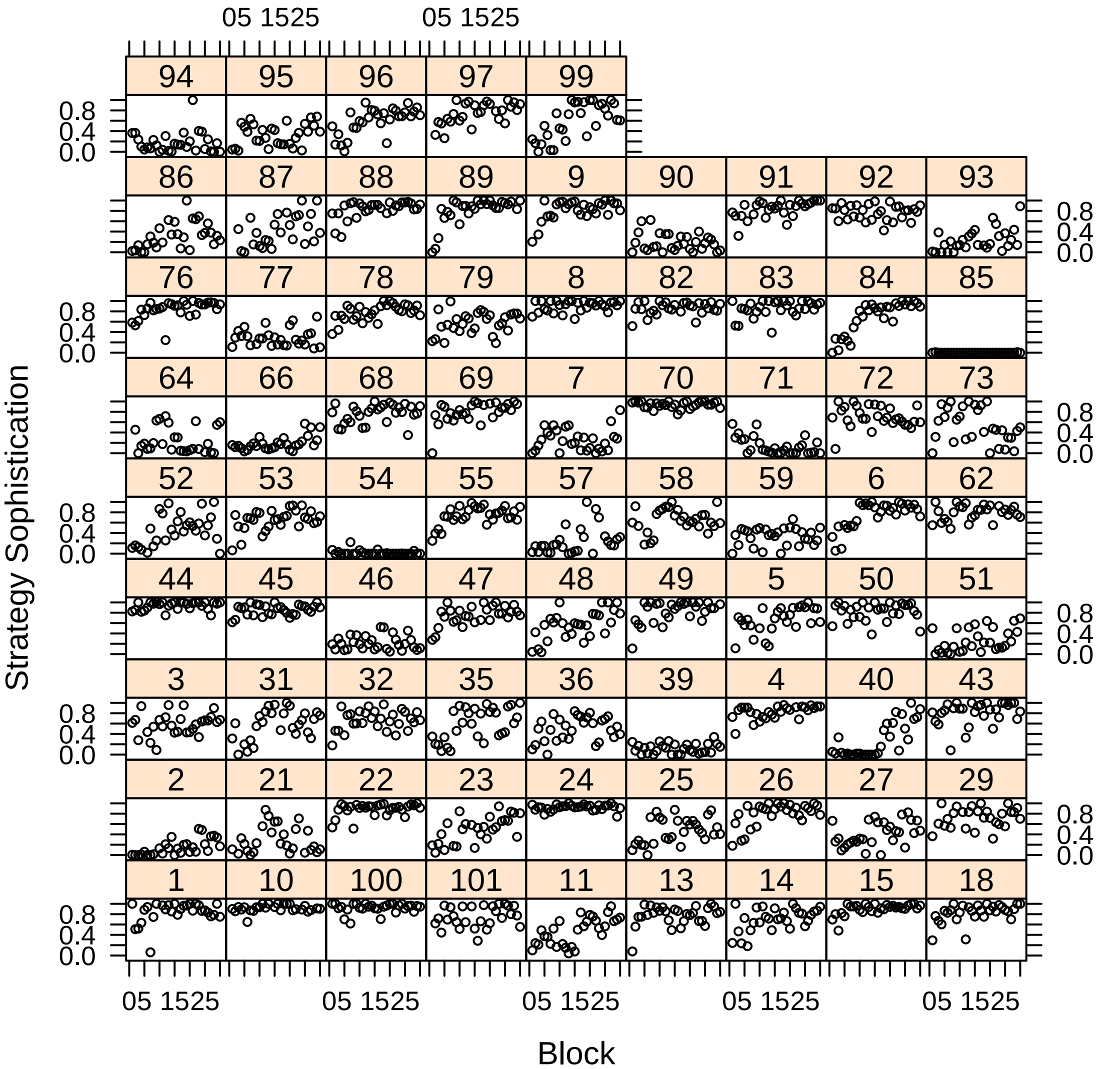

Figure 2.17: Study 2 strategy sophistication by block at the individual-level — Part 1 of 2.

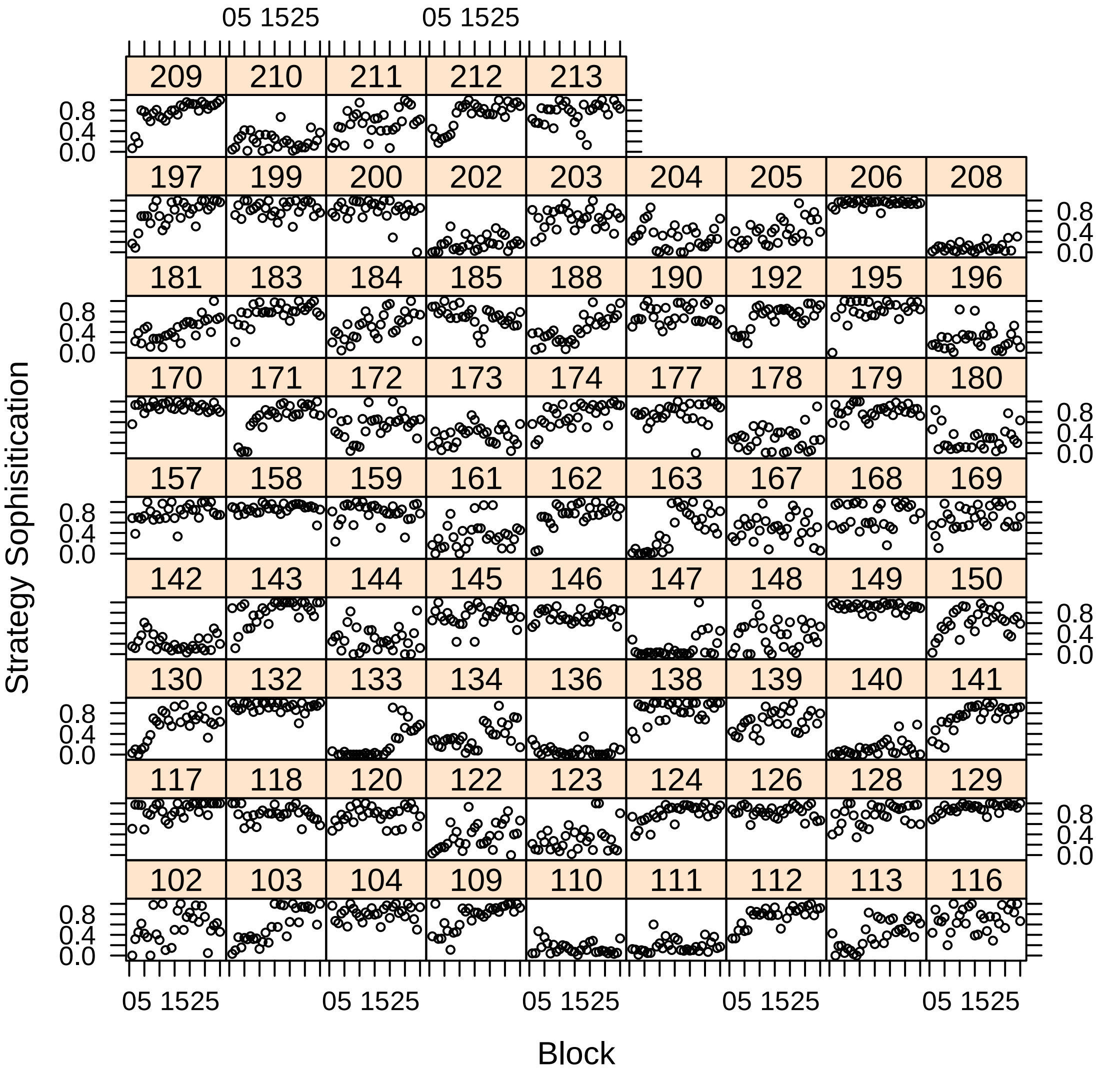

Figure 2.18: Study 2 strategy sophistication by block at the individual-level — Part 2 of 2.

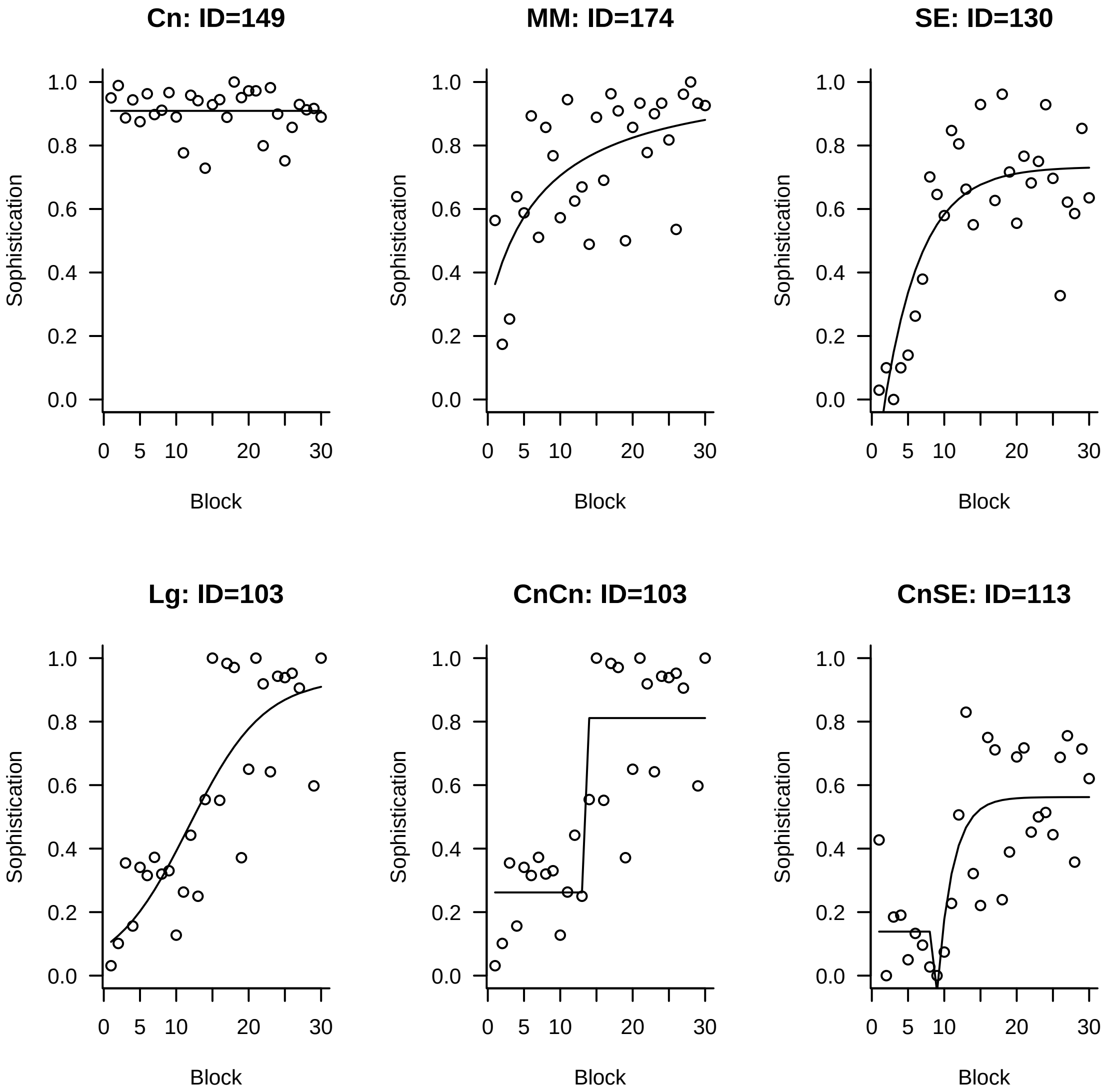

Figure 2.19: Study 2 examples of model fits of the effect of block on strategy sophistication at the individual-level.

2.4.3 Practice, Strategy, and Performance

Within-individual correlations between strategy sophistication and log trial completion times were calculated for participants with a standard deviation in strategy sophistication greater than 0.20 (129 out 154 participants were above the cut-off). The typical correlation was negative and small to medium (M= −0.23, SD = 0.18, range: −0.74 to 0.25).

2.5 Study 3 Results

2.5.1 Practice and Performance

Group-Level

Figure 2.20 shows the relationship between practice and performance at the group-level. Table 2.21 shows the fit information for the group-level learning curve. Consistent with Hypothesis 2.1 the three parameter power function had superior fit in comparison to the three parameter exponential function.

Table 2.21: Study 3 Fit Statistics for Models of the Effect of Block on Task Completion Time at the Group-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	1.250	.000	101.50
2.	P3	3	0.141	.988	-27.41
3.	E3	3	0.219	.971	-1.12
4.	APEX	4	0.144	.988	-25.43
5.	P4	4	0.144	.988	-25.53
6.	CnCn	3	0.694	.713	68.10
7.	P2BP2	5	0.140	.989	-26.12
8.	E3B	5	0.191	.980	-7.68

[a] k = Number of parameters.

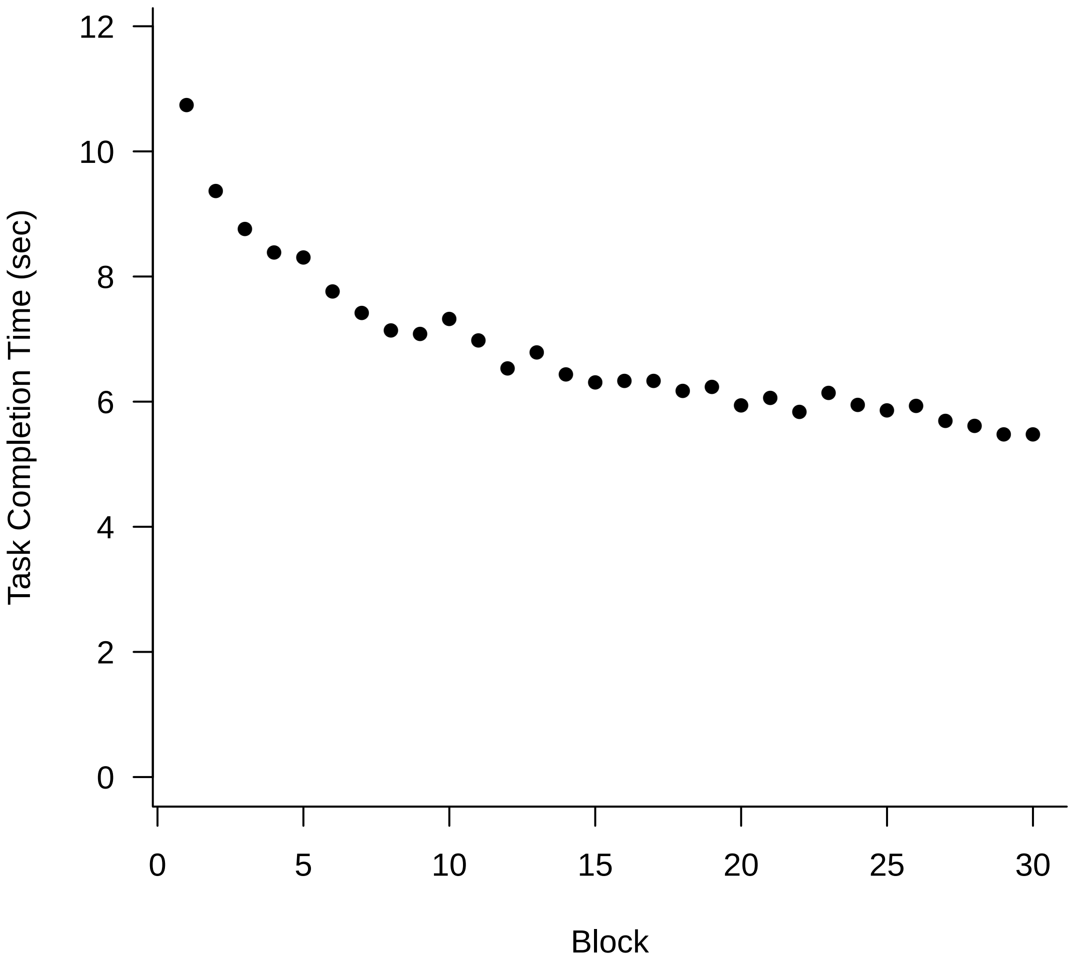

Figure 2.20: Study 3 task completion time by block at the group-level.

Individual-level

Overview Figure 2.21 shows the relationship between practice and task completion time at the individual-level. While participants tended to get quicker over time, large individual differences existed in average performance, amount of improvement, shape of improvement, and trial-to-trial variability.

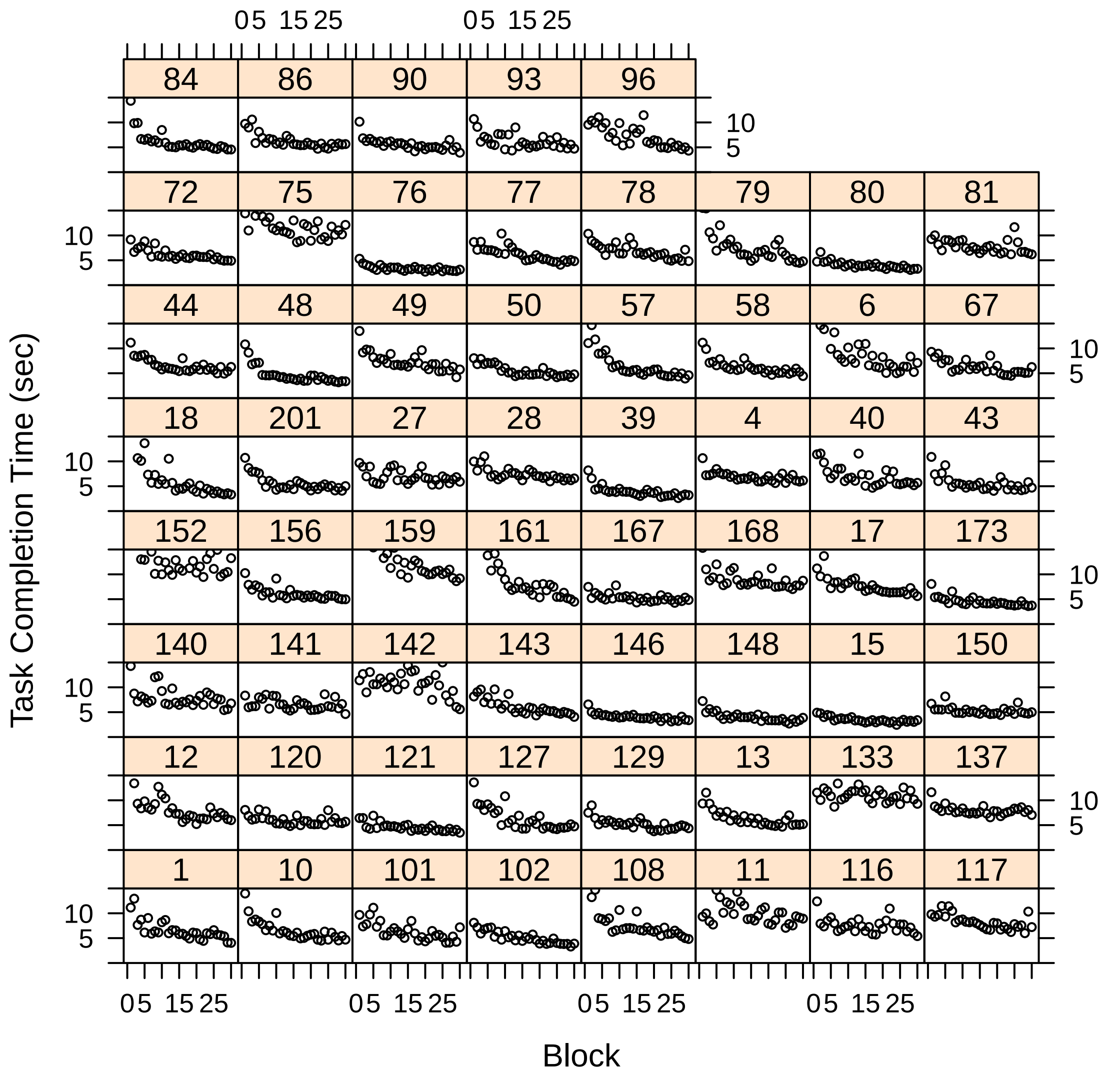

Figure 2.21: Study 3 task completion time by block at the individual-level.

Power versus Exponential Function Table 2.22 summarises individual-level model fit information for three parameter power and exponential functions. While the sample results were in a direction supporting Hypothesis 2.2 the exponential function fit was not significantly better than the fit for the power function ($rmse_{P3} = 1.02, rmse_{E3} = 1.01$). Using a paired samples t-test, this was not statistically significant different, $t(60) = 1.14, p = .26$. In terms of individual cases, the exponential function beat the power function in 29 cases, compared to 32 wins for the power function. This was not significantly different using a chi-square goodness of fit test assuming 50% wins for both functions, $\chi^2(df = 1) = 0.148, p = .70$.

APEX versus P4 Function The APEX function provided somewhat better fit than the four parameter power function in terms of mean error averaged across participants, ($rmse_{APEX} = 1.00, rmse_{P4} = 1.02$. Using a paired samples t-test, this was not quite a statistically significant difference, $t(60) = 1.81, p = .08$. In contrast, the APEX function beat the power function in only 24 cases, compared to 37 wins for the power function. This also was not a statistically significant difference when compared to a chi-square goodness of fit test assuming 50% wins for both functions, $\chi^2(df = 1) = 2.77, p = .10$. The reason for this seemingly contradictory result is that for the majority of participants, the fit for the P4 and APEX models was almost identical, but with a slight tendency for better fit for P4, yet for six cases the APEX function was far superior.

E3B versus APEX Function The three parameter exponential break function provided superior fit in comparison to the APEX function, ($rmse_{E3B} = 0.932, rmse_{APEX} = 0.999$). Using a paired samples t-test, this was a statistically significant difference, $t(60) = 6.66, p < .001$. In terms of individual cases, the E3B function beat the APEX function in 58 cases, compared to 3 wins for the APEX function. This was a statistically significant difference when compared to a chi-square goodness of fit test assuming 50% wins for both functions, $\chi^2(df = 1) = 49.59, p < .001$.

Table 2.22: Study 3 Mean Model Fit Statistics for the Effect of Block on Trial Completion Time at the Individual-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	1.671	.000	112.47
2.	P3	3	1.021	.599	84.16
3.	E3	3	1.007	.606	83.97
4.	APEX	4	0.999	.628	84.12
5.	P4	4	1.015	.614	84.68
6.	CnCn	3	1.152	.516	92.91
7.	P2BP2	5	0.931	.689	80.83
8.	E3B	5	0.932	.684	80.86

Note. Number of individual learning curves = 61.
[a] k = Number of parameters.

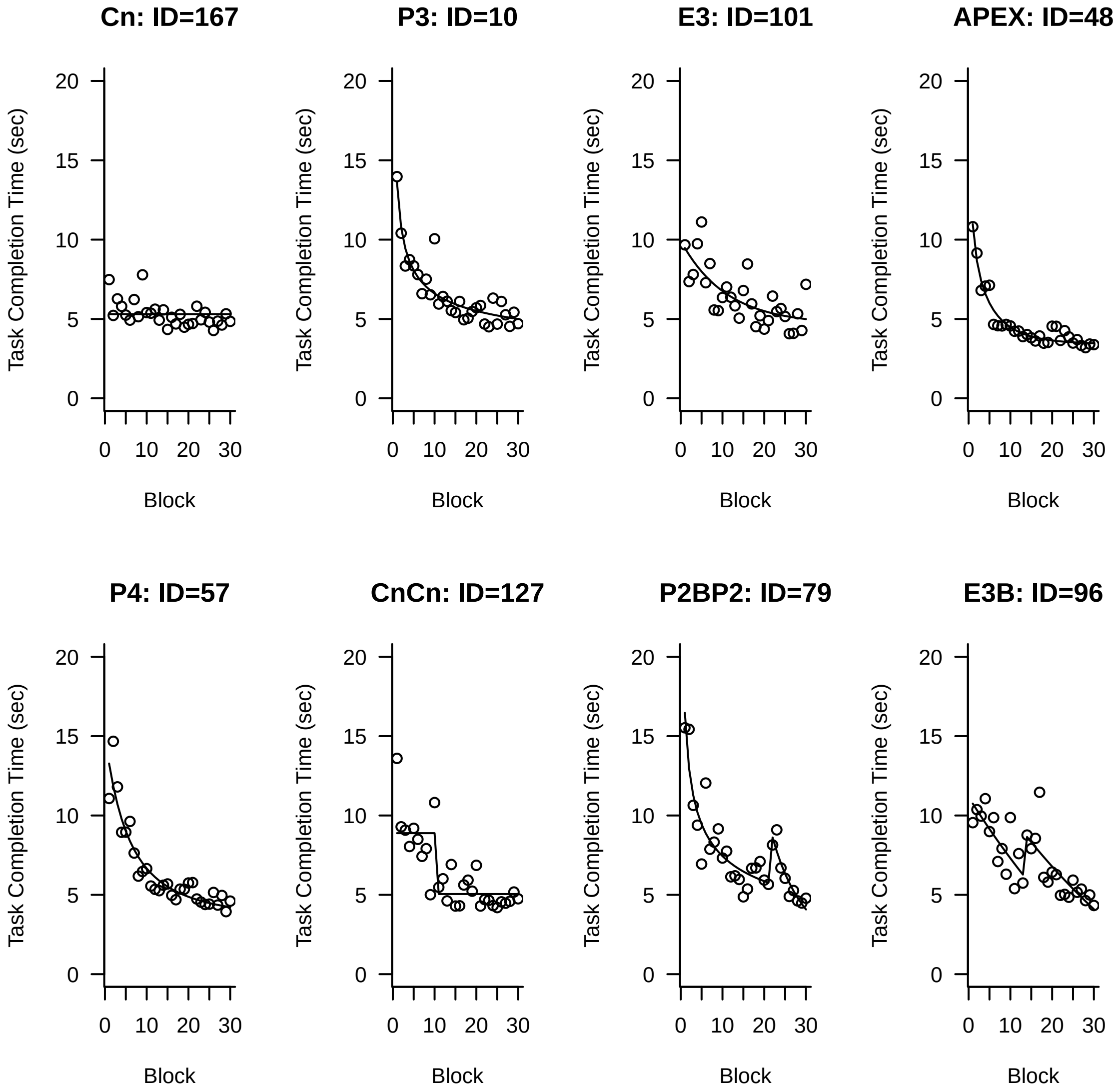

Figure 2.22: Study 3 examples of model fits of the effect of block on task completion time at the individual-level.

2.5.2 Practice and Strategy

Group-level

Figure 2.23 shows the relationship between practice and strategy sophistication at the group-level. Table 2.23 shows the model fit information. Supporting Hypothesis 2.10, a Michaelis-Menten function provided good fit to the group-level data. Interestingly the Michaelis–Menten function was clearly better than both the Saturating Exponential and the Logistic functions. The level of strategy sophistication was greater than both the previous two studies.

Table 2.23: Study 3 Fit Statistics for Models of the Effect of Block on Strategy Sophistication at the Group-Level

	Model	k[a]	se	R^2	*AIC*
1.	Cn	1	0.087	.000	-58.68
2.	MM	3	0.010	.987	-183.92
3.	SE	3	0.017	.966	-155.73
4.	Lg	3	0.020	.949	-144.55
5.	CnCn	3	0.049	.696	-90.45
6.	CnSE	5	0.036	.850	-107.63

[a] k = Number of parameters.

Individual-level

Figure 2.24 shows the relationship between practice and strategy sophistication at the individual-level. The figure reveals several distinct patterns of strategy use and strategy change at the individual-level. Table 2.24 shows a summary of model fits for each of the modelled functions.

Consistent with Hypothesis 2.9 several participants consistently had strategy sophistication around zero (e.g., cases 156, 168). Consistent with Hypothesis 2.8 some participants had consistently high levels of strategy sophistication (e.g., cases 80, 167). Some participants shifted to sophisticated strategies fairly quickly in the early trials (e.g., cases 12,

Table 2.24: Study 3 Mean Model Fit Statistics of the Effect of Block on Strategy Sophistication at the Individual-Level

	Model	k[a]	se	R^2	AIC
1.	Cn	1	0.125	.000	-55.66
2.	MM	3	0.071	.553	-85.89
3.	SE	3	0.068	.569	-86.67
4.	Lg	3	0.072	.512	-83.58
5.	CnCn	3	0.078	.484	-77.13
6.	CnSE	5	0.085	.432	-64.27

[a] k = Number of parameters.

44, 48). Consistent with Hypothesis 2.5 the speed of transition varied. Consistent with Hypothesis 2.6 several individuals only partially adopted sophisticated strategies (e.g., case 27). Figure 2.25 illustrates for each function one participant where that function fit relatively well.

Consistent with Hypothesis 2.7, the onset of the strategy shifts was greatest in early blocks. In broad terms, the frequency of shifts decreased with practice.

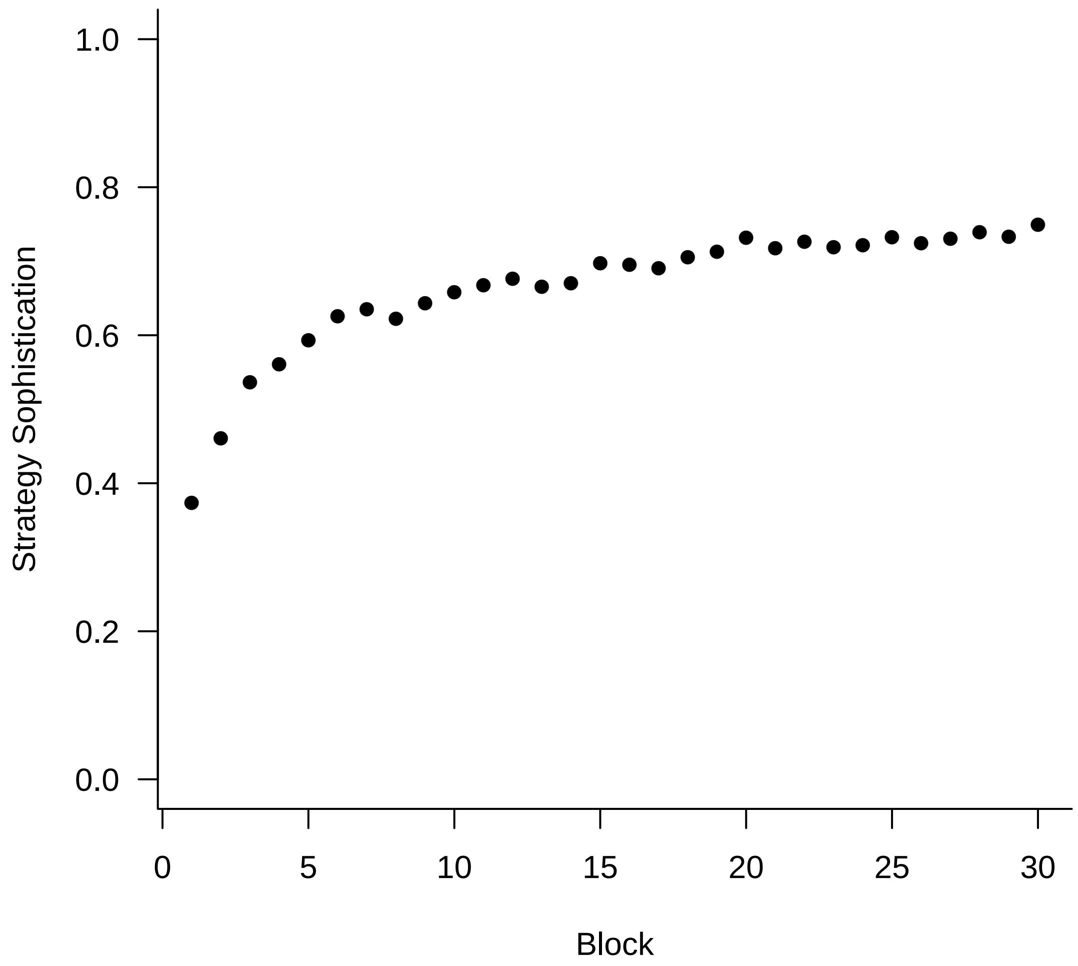

Figure 2.23: Study 3 strategy sophistication by block at the group-level.

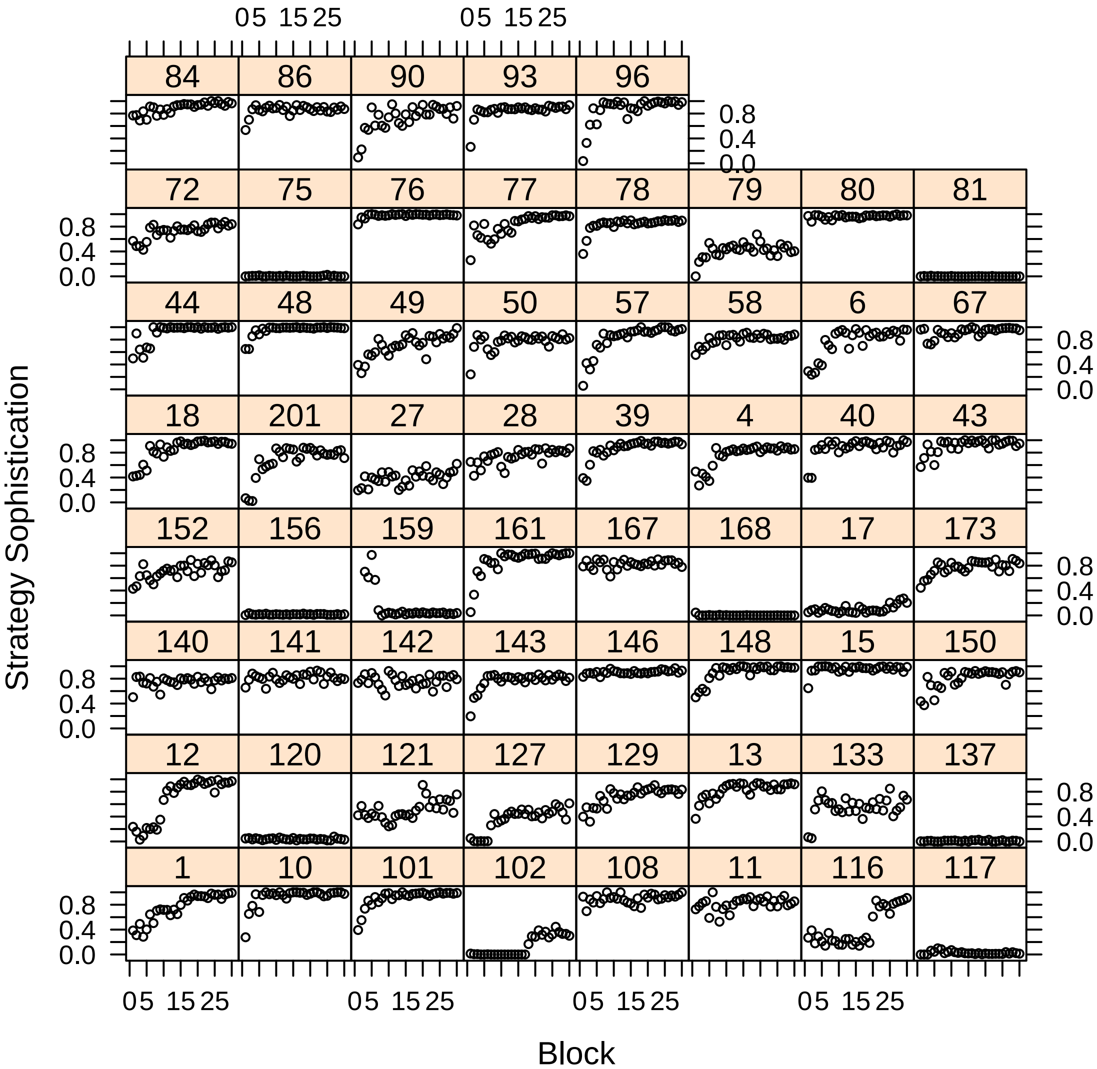

Figure 2.24: Study 3 strategy sophistication by block at the individual-level.

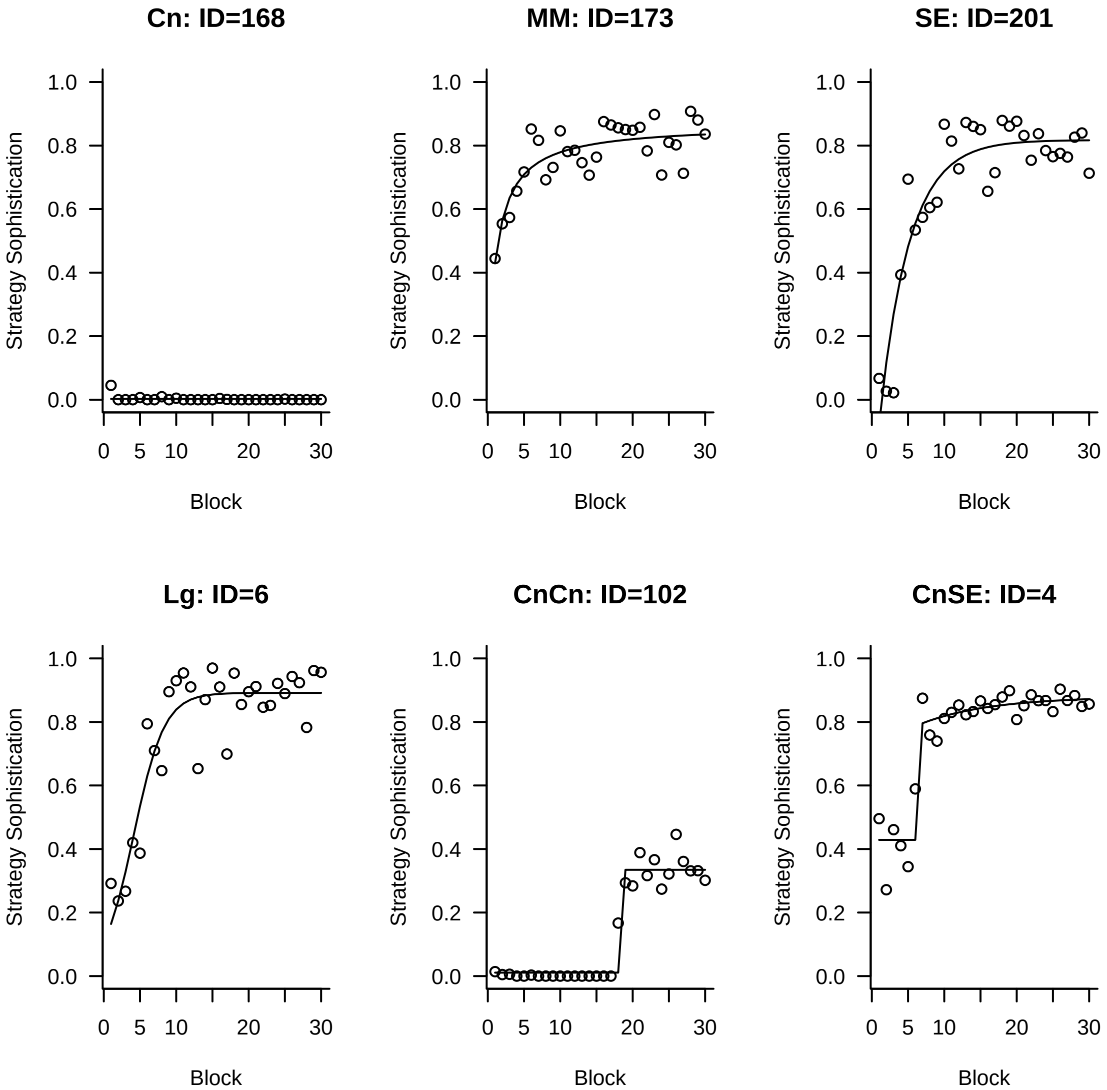

Figure 2.25: Study 3 examples of model fits of the effect of trial on strategy sophistication at the individual-level.

2.5.3 Practice, Strategy, and Performance

Within individual correlations between strategy sophistication and log trial completion time over trials were obtained for participants in the No Training Condition who had a standard deviation in strategy sophistication above 0.15 (45 out of 61 participants). The correlation tended to be moderate and negative (M = -0.34, SD = 0.17, range: -0.58 to 0.44).

2.6 Discussion

2.6.1 Overview

This chapter aimed to model the relationship between practice, strategy, and performance. To achieve this aim, hypotheses were tested and models were examined across three studies.

2.6.2 Practice and Performance

Group-Level

In all three studies, the relationship between practice and task completion time at the group-level was better explained by a three parameter power function than by a three parameter exponential function. The three parameter power function also provided excellent fit at the group-level in absolute terms explaining between 98% and 99% of variance. The degree of superiority was greatest in Study 1. There was almost no difference between four parameter power functions and APEX functions at the group-level for all three studies. Functions with discontinuities generally had worse fit than the continuous three parameter power function.

These results are consistent with the group-level findings of Newell and Rosenbloom (1981) and Heathcote et al. (2000). Thus, they support a limited version of the Power Law of Practice that is confined to group-level analysis. It is likely that the pooling of many noisy individual learning curves with their varying functional forms, learning slopes, timing of outliers, and occasional variably timed mild discontinuities lead to an averaged curve well explained by a three parameter power function. While small deviations in group-level curves were obtained, these are likely to be further smoothed over as sample sizes increase. Nonetheless, this says little about the individual-level, which should be the level of interest when developing a theory of psychological learning processes.

Individual-Level: Power versus Exponential

In studies one and two the relationship between practice and task completion time was better explained by a three parameter exponential function than a three parameter power function at the individual-level. Thus, the superior model at the individual-level differed from what applied at the group-level. The size of the difference in average fit was relatively small. However, given the limit on the amount of systematic variance to be explained the observed differences are still noteworthy.

These results are consistent with those of Heathcote et al. (2000). They support the repealing of the Power Law of Practice. Given that it is the individual-level which is of fundamental importance for understanding psychological change and learning, the results further highlight the problems associated with only examining the group-level relationship.

The general diversity in shapes and patterns in learning curves reiterated the importance of individual-level analyses. Graphical plots of individual curves were also a powerful tool for understanding this variation. While this thesis focuses on modelling expected task completion times, the plots showed clearly that other factors varied between individuals including frequency of outliers, trial-to-trial variability, and the degree to which one model provided superior fit.

Nonetheless, power and exponential functions are both monotonically decreasing and declerating in their functional form, and both approach an asymptote. Both are consistent with a model of skill acquisition as the progressive optimisation of an individual to a task and an environment. Large improvements are attained initially before additional refinements are made to smaller and smaller degrees and with less and less frequency with practice. Thus, the differences between power and exponential models are not as profound as the differences that both power and exponential have with discontinuous models.

Individual-Level: Discontinuities

The analysis of discontinuities yielded mixed results. Graphical inspection of learning curves suggested that meaningful discontinuities were rare. There were only a couple of clear jump discontinuities where a participant abruptly improved performance. However, there were other cases where the learning curve looked to be made up of segments of varying learning rates, arguably separated by a point where the learning rate slowed fairly abruptly.

The graphical observation that discontinuities were rare contrasted with the model fits for tested discontinuity models. While the Constant-Constant model provided poor fit, the P2BP2 and E3B models both provided far superior fit than the monotonically decreasing and decelerating three and four parameter functions. The superiority of E3B and P2B2B functions applied both to biased measures of fit such as R-squared and to parsimony adjusted measures such as AIC and RMSE.

There are several plausible reasons for the superior model fits of the discontinuous functions. First, the distribution of residuals was both skewed and often included outliers. Such outliers were often slow trials, although once discontinuous models are under consideration, discriminating outliers from discontinuities can be challenging. Discontinuous models were seemingly able to capture this noise more effectively than the continuous monotonic functions. To reduce the effect of outliers on model comparisons, future research could consider modelling outliers using some form of mixture model. Task completion time could be drawn from a main model with some large probability as well as some outlier generating process with a much smaller probability.

Second, the learning curves of many individuals had segment-like qualities. Thus, even if the learning curves lacked jump discontinuities, many learning curves seemed to contain discontinuities in the rate of learning. This suggests that learning is a continuous process of accumulating skill that occurs in a stochastic manner. Thus, even if learning is typically more rapid early in practice, learning may accelerate and decelerate over time. Discontinuous models are better able to capture such shifts in the rate of learning, which differ from the smooth monotonically decelerating functions.

Thus, it appears that discontinuous models have some merit in modelling individual-level learning curves. However, in terms of the text editing task studied in this thesis, the observed discontinuities were rarely of the form discussed by Haider and Frensch (2002), i.e., abrupt improvement in performance. Despite the text editing task permitting strategic insights, and despite these insights occurring, abrupt shifts in performance were almost never observed. Thus, while group-level learning curves provide a biased representation of the learning curve at the individual-level, the current study supports the idea that even on tasks permitting discontinuous strategies, performance tends to remain cumulative.

Differences between the studies appear to be related to the structure of the task. Study 1 involved a consistent set of editing requirements and multiple edits per trial. Both these factors reduced the trial-to-trial variation in task completion. The clearest case of an observed jump discontinuity in performance occurred in this study. This is consistent with the idea that such discontinuities depend on a meaningful simple and single discontinuity allowing a jump to occur. Study 2 and Study 3 had varying editing requirements across trials which increased the trial-to-trial variation in task completion time. Study 2 in particular had different types of edits which led to the greatest trial to trial variation. Statistically, such variation makes it harder to detect discontinuities. Theoretically, such variation in task features may make it less likely that a single insight can lead to an abrupt improvement in performance. Any insight that does occur may need to be adapted to the dynamic task environment. Furthermore, dynamic real-world environments may be more likely to require a degree of generalisation.

Overall, the findings are consistent with the idea that discontinuous improvements are a rare occurrence. Individual-level performance over short periods of temporal aggregation and on a variable task is fairly variable. The distribution of task completion time is often skewed and involves outliers which can lead to the impression of discontinuities. However, there are instances, although rare where meaningful discontinuous improvements do occur.

Implications

The above findings support the large scale analyses of learning curves by Heathcote et al. (2000). They consolidate the critique of the proposed Power Law of Practice (Newell & Rosenbloom, 1981). Given that the individual-level is the psychologically meaningful level of analysis, if anything, an exponential law seems more appropriate. This also has implications for researchers who have developed models which assume that the learning curve is a power function such as Logan's Instance

Model (Logan, 1988, 1992) and various cognitive architectures such as Act-R (Anderson et al., 2004). However, given that both power and exponential functions are continuous and monotonically decelerating, the theoretical adjustments may be relatively minor. The greater challenge may be to capture the variability in learning curves and to model the trial-to-trial variability.

The above findings also relate to researchers who have proposed that learning curves at the individual-level contain discontinuities. Haider and Frensch (2002) presented data on an alphabet string verification task and found discontinuities in the learning curves of many of their participants. They then showed analytically and by simulation that group-level curves can smooth out discontinuities at the individual-level. They implied that given the prevalence of strategy shifts, discontinuities in individual learning curves are likely to be prevalent.

However, the results of the three studies suggested discontinuities were rare. Group-level analyses did produce a biased representation of the relationship and trial-to-trial variability, but discontinuities at the individual-level were quite rare. As will be discussed shortly, abrupt strategy shifts do not equate to performance discontinuities. Also, discontinuities that are observed are just as likely, if not more likely, to be the result of some form of temporary drop in performance as opposed to a performance improvement. Given that few researchers in the literature have reported discontinuities, it may be that the dynamics in the study by Haider and Frensch (2002) which led to the observed discontinuities are unusual.

A major contribution of this thesis was the use of formal models of discontinuities. Neither Haider and Frensch (2002) nor Heathcote and colleagues (2000) tested such models. It would be interesting if future researchers applied such models to a greater number of learning curves at the individual-level.

Future research could further explore the role of parameter constraints in modelling learning curves. The present thesis took a minimalist approach to parameter constraints, constraining only breakpoint parameters to ensure that breakpoints were meaningful. Further constratints could be extended to ensure that predicted task completion time and strategy sophistication are within plausible values beyond observed practice levels and toward asymptotic levels. Such constraints include requiring that asymptotic performance be above zero, and perhaps even above some plausible minimum task completion time. Likewise, strategy sophistication models could have asymptotes constrained to be less than or equal to one. Decisions about choosing constraints also raise challenges about the purpose of modelling learning curves and the implications of good fit achieved by psychologically implausible parameters.

Related to parameter constraints, future research could also explore the issues related to model fit statistics. This thesis largely relied on mean square error and AIC to compare models. However, once model constraints are incorporated, or where the number of parameters differ, then substantial debate exists about what fit statistic is most useful. In particular, future research could explore BIC and cross validation approaches. Models that capture outliers and non-normal residuals might also provide a more accurate representation of the data and might assist in evaluating the relative merits of competing models. The challenge is for the model fit statistics to reward not just statistical fit, but also a psychologically meaningful model.

2.6.3 Practice and Strategy

Group-Level

In all three studies, at the group-level the three parameter Michaelis–Menten function provided excellent fit to the relationship between practice and strategy sophistication explaining over 97 percent of variance. Four key aspects of the observed relationship captured by the function were that: (a) strategy use showed some level of sophistication at the start of practice, (b) strategy sophistication increased monotonically with practice, (c) the rate of increase in strategy sophistication decreased monotonically with practice, and (d) strategy sophistication appeared to be approaching an upper level asymptote toward the end of practice that was less than one.

The superiority of Michaelis–Menten varied across the three studies. In Study 1, several functions including the Logistic and the Saturated Exponential provided almost as good fit as the Michaelis–Menten. In Study 2 the superiority of Michaelis–Menten was larger but still small, and in Study 3 the superiority was relatively large. This superiority also mirrored the relative ordering of key models, which is to say that Study 2 and Study 3 emphasised more clearly which models were second, third, and fourth best. A clear ordering of models emerged in Study 3 with Michaelis–Menten fitting best followed by Saturated Exponential, Logistic, and then Constant–Saturated Exponential.

The increase in model differentiation across the studies mirrored the size of strategy change in the three studies. Study 3 had the greatest increase in strategy sophistication with practice, whereas Study 1 had the smallest increase. The strategy sophistication curve in Study 3 was prominently decelerating with rapid strategy change occurring in early blocks. In contrast, Study 1 had lower levels of strategy sophistication and lower levels of strategy change. The curve departed minimally from linearity.

There are several explanations for these differences. First, Study 3 had the most comprehensive initial strategy training of the three studies. Second, the repetitive nature of the deletion task in Study 3 relative to Study 2 may also have facilitated use of sophisticated strategies. Simplicity should reduce attentional demands, reduce the complexity of generalisation, and make strategies specific to the particular problem (e.g., deletion) more effective. Third, the sample of third year university students used in Studies 2 and 3 appeared to have more prior experience than the mixed university and general population sample used in Study 1.

The above provides a model of how strategy change occurs at the group-level. However, as suggested by theory and supported by results, the group-level provides a misleading representation of the relationship between practice and strategy sophistication at the individual-level.

Individual-Level

Functional form The main hypotheses regarding the relationship between practice and strategy sophistication at the individual-level were supported in all three studies: (a) Some individuals never shifted either because they started with low levels or high levels of strategy sophistication and persisted at that level throughout practice; (b) shifts varied in whether they were abrupt or gradual; (c) probability of a strategy shift decreased with practice; and (d) many participants only partially shifted to the sophisticated strategy.

Results supported the argument put forward by Siegler and others (e.g., Lemaire & Siegler, 1995; Siegler, 1987, 1988a) that the relationship between practice and strategy use needs to be analysed at the individual-level. Trajectories varied substantially between individuals. Also, the individual trajectory tended not to resemble the group-level relationship. For those individuals that changed, strategy use a somewhat abrupt onset and a relatively rapid transition was a common pattern. This contrasts with the gradual increase starting from the first trial as suggested by the group-level model. While such a model did occur at the individual-level, it was only suitable for a subset of participants.

Of those individuals that did shift, the degree to which strategy shift was gradual or abrupt also varied between individuals. While an arbitrary cut-off could be given to classify a shift as either gradual or abrupt, a more meaningful representation would describe the shift in terms of the degree to which it was gradual. Some shifts took only two or three trials to complete whereas others took ten or twenty to complete. Thus, assessment of the degree to which shifts were abrupt is contingent on the number of trials allowed in the transition period.

In addition, the concept of 'shift' varied in its degree of applicability. First, even once a shift had occurred, individual trials with lower levels of sophistication still occurred. In many cases the proportion of these low-sophistication trials tended to decrease with practice. Thus, even when the shift was largely abrupt, it still often took time for it to be complete.

It is also inappropriate to describe the relationship between practice and strategy sophistication in terms of a single functional form. The variation in functional form was much greater than for the relationship between practice and task completion. Each of the six models proposed were effective in modelling the relationship between practice and strategy sophistication for some participants. The models captured various combinations of gradual and abrupt change in strategy sophistication, gradual and abrupt onset of strategy shift, immediate and delayed onset of strategy shift, and whether a shift occurred or not.

Implications for Learning Curves In addition to understanding the relationship between practice and strategy use for its own sake, this thesis was motivated by a desire to resolve the tension between a discontinuous relationship between practice and strategy use, and a continuous relationship between practice and performance. The results suggest both that this tension is real and also that it is not as great as some authors have implied (e.g., Haider & Frensch, 2002). Of the variety of functional forms for the relationship between practice and strategy sophistication, many were more discontinuous than individual-level learning curves. However, the majority of strategy shifts do take a period of time to complete which would tend to reduce the potential for discontinuous improvement in performance.

The tendency for strategy shifts to occur early in practice makes their effect on performance more difficult to discern. In early trials performance is more variable, and a large amount of improvement is occurring due to a range of learning processes. Isolating the unique contribution of strategy shift is often not possible.

Adaptivity Another aspect of the observed relationship between practice and strategy sophistication is that very few individuals moved from a sophisticated strategy to a simpler strategy. In the few cases where this occurred, the change to the simpler strategy seemed to occur after the participant tried the more sophisticated strategy for only a few trials. There are several explanations for this overall pattern of results. First, the strategies were sufficiently superior for most participants to persist with them. Second, participants were adaptive in their strategy choice and sensitive to small improvements in strategy effectiveness. Finally, individuals able to use the strategies well were more likely to consider the strategy and try it.

Comparison with Algorithm-Retrieval Shift The relationship between practice and strategy sophistication was also different to what is typically found in studies looking at algorithm-retrieval shift (e.g., Delaney et al., 1998; Rickard, 1997, 1999). In algorithm-retrieval studies most participants shift to the superior retrieval strategy, and the relationship between practice and probability of using the retrieval strategy tends to be sigmoidal or exponential. The reason for the greater uptake in retrieval studies is probably related to the uptake process of retrieval responses which are facilitated by mere exposure to stimulus and correct response. In contrast, in text editing, the simple strategy does not necessarily suggest the more sophisticated strategy. Greater practice may increase the motivation to find a quicker strategy, but if the strategy is not in the repertoire or its benefits are not known, then it may never be adopted. This was seen in Study 1 where few hints were given regarding sophisticated strategies and there was minimal uptake.

Differences in the timing of algorithm-retrieval and simple-sophisticated strategy shifts may be related to how the strategy is acquired. Sophisticated editing strategies often enter through some form of declarative representation. In addition to being able to execute the strategy, it often takes time to learn to execute the strategy fluently and quickly. In contrast, trying to execute a retrieval strategy is useless without a representation of the correct response in memory. Such a memory representation typically takes many exposures, especially where the size of the set of items is large.

The probability of a shift appears to decrease monotonically with practice. This is consistent with the idea that when engaging in a repetitive activity, participants initially explore the environment and then settle into a satisficing routine (e.g., exploration and exploitation models, Erev et al., 2008). Furthermore, practice using an adopted strategy often leads to greater proficiency in the adopted strategy and, thus, a greater performance and automaticity cost associated with switching strategies.

Text editing strategy shifts may also be similar to other types of strategy shifts that involve an element of discovery (e.g., Luchins, 1942) in that the options are not known by all individuals. The main difference is that on typical insight based tasks, discovery is required to perform a task at all, whereas in the case of text editing strategies, it merely results in suboptimal performance. Furthermore, individuals may not be aware that their performance is suboptimal, because they are unaware of the benefits of the more effective strategy. Further explaining the monotonic reduction in rate of strategy shift is the idea that individuals differ in their readiness to shift strategies. For those ready to shift, the shift occurs early; for those not ready to shift, the shift may not happen at all.

Summary There were also differences at the individual-level between the studies. First, there was the previously mentioned greater uses of sophisticated strategies in Studies 2 and 3 relative to Study 1. Second, the trial-to-trial variation in strategy use was much lower in Study 1 than in Studies 2 and 3. Two factors that are likely to explain most of these differences are first that the text editing requirements were constant in Study 1 and variable in Studies 2 and 3. Constant editing requirements should be more likely to induce a regular pattern of responding. Second, trials in Study 1 involved multiple edits whereas Studies 2 and 3 involved a single edit. Aggregation alone should serve to smooth out variability. This finding is broadly consistent with Siegler's (1987) Overlapping Wave Theory, which suggests that strategies overlap in usage over time with more sophisticated strategies gradually replacing simpler ones.

The overall pattern of results is qualitatively consistent with Lemaire and Siegler's (1995) proposal of four types of strategy change that occur with practice: entry into repertoire, increased use of efficient strategies, improved execution of strategies, and improved selection of strategies. The findings are also broadly consistent with models of how individuals keep track of the success of strategies, as captured in many models of learning such as ACT-R.

Overall, this thesis aimed to improve models of the relationship between practice and strategy use by combining trial-level strategy measurement, graphing, and modelling of strategy sophistication at the individual-level. Previous studies have rarely used such a combination, particularly on tasks characterised by a simple–sophisticated strategy shift. The models, particularly at the individual-level, suggest a more nuanced account of the concept of abrupt versus gradual shift, and provide additional data that cognitive architectures need to consider. In particular, the results provide a strong warning against drawing inferences about individual-level strategy shift processes from group-level data.

2.6.4 Practice, Strategy, and Performance

A number of findings emerged regarding the relationship between strategy use and performance and its potential effect on the learning curve. First, strategy sophistication tended to increase with practice. Second, the period of greatest increase in strategy sophistication was associated with the period with the greatest change in task completion time. Third, while strategy shift was often abrupt, discontinuities in performance were rare. Fourth, at the level of intra-individual variation, strategy sophistication was correlated with task completion time.

Taken together, these findings suggest that increased strategy sophistication contributes to an eventual improvement in performance. Increasing strategy sophistication can be seen as partially mediating the effect of practice on performance.

Given that participants shifted towards more sophisticated strategies and rarely away, such individuals were presumably aware implicitly or explicitly of the benefit in using the more sophisticated strategies. Similarly, it may be that participants who did not adopt more sophisticated strategies were also those less likely to benefit from the sophisticated strategies. However, it is also possible that participants were unaware of the potential benefits of the strategies or had some other reason to persist with simpler strategies.

Several pieces of evidence also suggest that the initiation of a strategy shift is only the beginning of realising the performance benefit. First, while the onset of a strategy shift was often abrupt, it often took several trials for the generalisation of the strategy shift to be complete. Second, the time for strategy sophistication to emerge following initiation suggests that it takes time for participants to learn how to apply the strategy. Third, a theory driven analysis of the benefits offered by some of the more sophisticated strategies suggest that even once expertise is acquired, improvements are modest relative to to other factors, such as learning the task interface, learning the task, learning how to do text editing in general, and sources of other randomness such as variation in the type of trial, and random variation in effectiveness and distraction. In combination, this leads to less abrupt performance benefits from strategy shift.

While not detracting from the actual effect of strategy shift on performance, several factors appear to reduce the ability to detect the effect of strategy shifts on performance. First, because strategy shifts tend to occur earlier in practice, any performance benefit of a shift tend to occur at a time when both other learning processes are occurring and also when error variance is greatest. Second, in general, and particularly in Studies 2 and 3, the error variance tends to be fairly large at the individual-level. These factors add noise to any estimation of the effect of strategy shift, and thus, make it difficult to detect discontinuities in the learning curve.

Several differences between the studies are likely to influence the degree to which the effect of strategy shift results in a discontinuity in the learning curve. First, Study 1 had a consistent set of edits, whereas Studies 2 and 3 varied the editing requirements. Varying the editing requirements appeared to lead to longer periods of strategy generalisation. Second, Study 1 and 2 were more complex than Study 3 in the editing requirements that they entailed. This greater complexity meant that adopting a sophisticated strategy involved making more changes, which were less likely to occur in combination and more likely to occur in parts. The greater complexity also meant that a single strategic change was less likely to result in abrupt improvement. Third, Studies 2 and 3 had greater error variance relative to the overall amount of learning, which would make strategy shifts harder to detect against this backdrop of noise. Fourth, the degree to which the prior knowledge of participants, the task itself, and the instructions makes strategy shift both possible and apparent varied between the studies. Studies 2 and particularly 3 provided greater motivation and instruction on using the shortcut keys. This manifested in a greater number of strategy shifts with the consequent greater potential for a discontinuity to be observed.

These findings suggest modification to the thinking of several researchers on the relationship between practice, strategy use, and performance.

Crossman (1959) made an early attempt to explain the learning curve in terms of changing probabilities of strategy use. Practice was assumed to give rise to feedback of strategy effectiveness which led to an updating process of strategy use probabilities. First, the present results show that strategy use and performance are not synonymous. The findings show that the effectiveness of a strategy varies over time, and between individuals. Second, a strategy may have zero probability of use at a given moment if it is not in the repertoire. Thus, prior knowledge, instruction, motivation, and contextual affordances are likely to be important influences of strategy use.

Similarly, while not treating performance and strategy use as synonymous, Haider and Frensch (2002) seem to argue for a greater correspondence between strategy shift and performance shift than is suggested by the results in this thesis. In this thesis, relatively abrupt strategy shifts sometimes occurred, yet abrupt performance shifts almost never occurred. The theory proposed in this thesis provides a reconciliation of this.

2.6.5 Conclusion

The results of the three studies refine models of the relationship between practice, strategy use, and performance. They reinforce the importance of the individual-level measurement and provide a means of reconciling discontinuous effects of practice on strategy sophistication with continuous effects of practice on performance.

Chapter 3

Individual Differences

Contents

3.1 Introduction

3.1.1 Overview

The previous chapter showed how individuals differ substantially in strategy sophistication and task performance. It showed how group-level analyses lead to a biased representation of the individual-level. It showed how no one functional form can describe the relationships between practice, strategy, and performance.

If individuals differ in their learning processes, a natural question is what predicts this variability. Such an approach fits into the broader research enterprise of building links between psychometrics, individual differences, and cognitive psychology. Such an enterprise is captured in the work of researchers such as Ackerman (e.g., Ackerman, 1992, 1990, 1989, 1988, 1987; Ackerman, Beier, & Boyle, 2002; Ackerman, Kanfer, & Goff, 1995) and many others (e.g., Ecker, Lewandowsky, Oberauer, Chee, & Ecker, 2010; M. D. Lee & Webb, 2005; Lovett, Reder, & Lebiere, 1997).

This chapter builds on this research tradition by examining relationships between predictor and criterion measures on the text editing task, and thereby addressing Aim 2 of this thesis. Specifically, connections between four types of predictors and two criterion measures are examined. The four types of predictors are (a) ability, (b) prior experience, (c) personality, and (d) demographics. The two types of criterion measures are (a) task performance, and (b) strategy sophistication.

Many studies already exist that look at such predictor–criterion relationships. Whilst more modest than the previous chapter, the present chapter still aims to make at least two unique empirical contributions. First, it examines predictors of strategy sophistication, a variable rarely studied by previous researchers. Second, a model of strategy mediating the effect of predictors on task completion time is tested. This chapter also complements the previous chapter by focusing on individual differences at the trait-level.

To achieve these aims, the subsequent literature review is organised as follows. First, general theories and frameworks for predicting task performance are presented. Second, empirical evidence is discussed regarding the prediction of performance from demographics, prior experience, ability, and personality. Third, prediction of strategy sophistication is discussed.

3.1.2 Theories and Frameworks

Predicting performance has a long and contentious history in psychology (for a discussion, see Neisser et al., 1996). In addition to improving decision making in contexts such as employee selection (for a review, see P. Sackett & Lievens, 2008) and allocation of educational opportunities (for a review, see P. Sackett, Schmitt, Ellingson, & Kabin, 2001), researchers have also been interested in using individual differences to understand information processing models of cognition (e.g., Ackerman, 1987, 1988; Adams, 1987; Fleishman, 1972; Keil & Cortina, 2001). The challenge for research on individual differences is to explain why observed relationships vary as a function of factors such as type of predictor, type of criterion, context, and sample characteristics.

Various general frameworks have been proposed. First, meta-analyses and other validity generalisation studies (e.g., Burke, 1984; Schmidt & Hunter, 1977) have revealed the degree of robustness of factors such as intelligence in predicting factors like job performance. However, theory and empirical findings in meta-analyses typically indicate that the true effect size varies between studies. This leads to the challenge of identifying key moderators of this effect.

Second, many but not all (see Patrick, 1992) individual difference researchers endorse a version of the *cognitive correlates* perspective whereby the size of correlation between predictor and performance is assumed to be related to the importance of that predictor for task performance (for a discussion, see Proctor & Dutta, 1995). The cognitive correlates approach applies both to general enabling abilities and models of specific transfer, such as Identical Elements Theory (Thorndike & Woodworth, 1901) and modern day variants (e.g., A. Kramer, Strayer, & Buckely, 1990; MacKay, 1982; M. Singley & Anderson, 1985; M. K. Singley & Anderson, 1989). These ideas regarding ability–performance correlations have been extensively explored in studies by Fleishman (Fleishman & Fruchter, 1960; Fleishman, 1972) and Ackerman (Ackerman, 1987, 1988, 1989, 1990, 1992; Ackerman & Woltz, 1994; Beier & Ackerman, 2005; Kanfer & Ackerman, 1989).

Third, there is a general principle that correlations between predictors and outcomes are stronger if they correspond in terms specificity. Thus, stable traits such as intelligence and personality are assumed to be better predictors of long term measures of performance such as job performance.

In contrast, performance on a specific occasion is assumed to be better predicted by state variables close in time to the performance occasion. As with temporal congruence, there is a similar notion with regards to the generality of the domain, such that domain specific predictors should predict domain specific outcomes, and domain general predictors should predict domain general outcomes. The implication of this for experimental tasks is that relevant prior experience and related skills should be better predictors of task performance than general abilities.

While these general frameworks have value, it is important to be critical of causal interpretations of predictor–criterion correlations. A predictor may be an enabling capacity or it might just be a sign. And if it is a sign, it may just be epiphenomenal or it may have led to the third variable. It is important to consider alternative causal interpretations particularly regarding differences in prior experience of participants and how this might be related to other predictors.

A second issue that arises in relation to testing theories of predictor–criterion relationships is that the hypothesis testing orientation is not always suited. The issue is often not whether there is a relationship, but rather the relative size of the relationship. To deal with this issue, proposed hypotheses are qualified with the terms, small, medium, and large corresponding to approximate expectations of correlations around 0.10, 0.30, and 0.50 respectively. These adjectives corresponds to Jacob Cohen's (1992) influential rules of thumb for correlations in social science research.

3.1.3 Predictors of Performance

Demographics

Age and gender have all been extensively studied as predictors of performance. While researchers have often showed differences between younger and older adults (e.g., Hertzog, Cooper, & Fisk, 1996; Rogers, Hertzog, & Fisk, 2000), differences within younger and middle-aged adults are often expected to be minimal. When differences do arise, it is interesting to consider the role of declining abilities with age as opposed to differences in prior experience often related to cohort effects. Likewise, some studies find gender differences (for a discussion, see Ackerman et al., 1995), and similar questions arise over differences in prior experience related to the kinds of tasks used in skill acquisition experiments. In particular, males often perform better on computer based tasks common in skill acquisition experiments (e.g., Ackerman et al., 1995). Thus, it was expected that on the text editing task, males would be slightly better than females and younger adults would be slightly better than older adults.

Prior Experience

Prior experience is generally a good predictor of task performance. McDaniel, Schmidt, and Hunter (1988) provides a discussion of experience predicting job performance. In the context of experimental designs, despite tasks often being selected for their novelty, prior experience is still relevant. For example, when the task involves using computers, experience with video games is often a good predictor. Rabbitt, Banerji, and Szymanski (1989) obtained a correlation with task performance of .33 for self-rated and .40 for experimenter rated video game experience. This links with theories of transfer, which often show somewhat weak correlations with other specific skills. However, task analytic approaches to selecting predictors of job performance are often used. In particular, where the period of practice is short relative to a life of potentially relevant prior experience, this prediction may be greater.

The present studies used a keyboard based text editing. Based on task analysis, many prior experience measures should be relevant such as amount of experience with keyboards, computers, word processors, and other text manipulation tools. Experience with keyboard-based video games may also be relevant.

Thus, it was hypothesised that

Hypothesis 3.1 *Typing speed is strongly positively correlated with text editing performance.*

Hypothesis 3.2 *Prior experience is strongly positively correlated with text editing performance.*

Ability

Before discussing the prediction of ability, it is worth reviewing what is meant by the term 'ability'. In contrast to skill, and following on from Fleishman (1972), ability is treated as a (a) general, (b) relatively resistant to change, (c) enabling capacity. Intelligence as g is the best known and most studied ability, yet many other representations have emerged. Several multi-faceted and often hierarchical representations of cognitive ability have been proposed (e.g., J. B. Carroll, 1993). Other researchers have focused on non-cognitive abilities (e.g., Fleishman, 1972). Ackerman (1988) adapted the multidimensional scaling analysis of Marshalek, Lohman, and Snow (1983) and divided ability into general cognitive, perceptual speed, and psychomotor ability.

Empirical evidence generally shows strong relationships between ability measures and task performance. Meta-analyses typically show strong correlations between intelligence and job performance. For example, Hunter and Schmidt (1998) report a mean meta-analytic correlation between general mental ability and job performance of $r = .63$. Other researchers such as Ackerman (1988) and Fleishman (1972) have focused on changes in ability–performance correlations over time (for a meta analysis, see Keil & Cortina, 2001). In general the literature has shown logical connections between particular abilities and the apparent demands of tasks (see Farina & Wheaton, 1971).

Thus, it was hypothesised that:

Hypothesis 3.3 *Ability is moderately positively correlated with text editing task performance.*

In this thesis, ability is operationalised in ways consistent with Ackerman (1988) in terms of general cognitive, perceptual speed, and psychomotor abilities. Differential prediction of the three classes of abilities was of interest, but no specific hypotheses were proposed. The meta-analysis by Keil and Cortina (2001) broadly suggested stronger correlations with task performance for cognitive and perceptual speed than for psychomotor ability. However, Ackerman (1990) has shown evidence of prediction for all three types of ability on an air-traffic control simulation, which mirrors some general features of text editing.

Personality

Personality traits have received increased attention over the last 20 years as potential predictors of performance. This can be attributed to two major factors. First, the Big 5 model of extraversion, neuroticism, conscientiousness, agreeableness, and openness (or intellectance) has received increased acceptance and has emerged as a useful means of synthesising the diverse set of non-ability based individual difference constructs (see Digman, 1990; L. Goldberg, 1993). Second, building on the Big 5, influential meta-analyses have tended to show weak but positive (negative for neuroticism) correlations between the Big 5 scales and job performance with conscientiousness in particular emerging as a consistent predictor (see Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991).

Some attempts have also been made to integrate ability and personality perspectives (e.g., Ackerman et al., 1995) within the training context.

In general, theories often suggest that personality is associated with what an individual 'will do', whereas ability is concerned more with what they 'can do'. If performance requires effort and ability, personality is assumed to be a positive predictor of discretionary allocation of effort.

However, there are several reasons to expect self-report measures of personality to be unrelated to task performance in laboratory settings. First, even in applied settings, correlations tend to be small (see Barrick & Mount, 1991; Tett et al., 1991). Second, laboratory experiments minimise the role of discretionary effort. The nature of the experimental environment is typically short enough and sufficiently monitored to align more with what P. R. Sackett, Zedeck, and Fogli (1988) call a *maximum performance* situation, as opposed to a *typical performance* situation. Third, self-reported measures of style and approach are often relatively poor indicators of competencies. Competencies rather than motivational tendencies are likely to be better predictors of performance, particularly in settings where discretionary effort is largely controlled. Thus, it was hypothesised that:

Hypothesis 3.4 *Big 5 personality factors are unrelated to text editing task performance.*

3.1.4 Predictors of Strategy Sophistication

While there is a lot of evidence of what predicts performance, there is much less evidence of what predicts strategy sophistication. A starting position is that the same things that predict task performance also predict strategy sophistication.

Prior Experience

Prior experience represents a more domain specific pathway of acquiring sophisticated strategies. With practice, people often adopt more sophisticated strategies, although it is not uncommon for this to asymptote at suboptimal levels. This would suggest that variables like typing speed and prior experience with text editing would be associated with greater text editing strategy sophistication.

Ability

Kyllonen et al. (1984) provides an early example of studying the relationship between ability and strategy use. Kyllonen et al. (1984) had 30 participants complete various ability tests followed by a task that required the construction of designated shapes from assorted pieces. Participants were classified in terms of the shape construction strategies that they adopted. While results suggested that ability was related to strategy use, the sample size was too small to draw clear conclusions.

In another article, Schunn and Reder (2001) report three studies that looked at correlations between a battery of ability tests and adaptive strategic use of runway landing strategies on a simplified air-traffic control task. Results generally showed that approximately ten percent of measured ability tests were significantly correlated with operationalised strategy adaptivity. However, the choice of analyses was problematic. Running significance tests for each test in a large battery, particularly with small sample sizes, leads to problematic inferences. For example, if the true population correlation with task performance of all tests was .20, sample correlations would vary from test to test. By chance alone, some tests would be statistically significant and others would not. Furthermore, given the large correlations between individual tests, it is unlikely that any one test is a pure measure of an individual cognitive component. Factor analytic approaches that reduce individual tests into general measures are likely to yield more parsimonious and plausible mappings between predictors and strategy sophistication. They also reduce the problems associated with multiple significance tests because fewer tests need to be run. In summary, Schunn and Reder (2001) do suggest correlations exist between practice and strategy adaptivity, but the nature of differential predictions based on ability classes remains unclear.

Ackerman (1988, p. 295) theorised that the "efficacy of the initial productions formulated in Phase 1 of skill acquisition [are] a function of general-broad content abilities". This builds on ideas of Anderson and colleagues (e.g., Anderson, 2000, 1993, 1990, 1987) who propose that an important component of skill acquisition is the process of knowledge entering declaratively before it becomes proceduralised. This would suggest a correlation between general ability and strategy sophistication.

It also seems likely that instruction could moderate the relationship between predictors and strategy sophistication. Considering that strategy sophistication can be manipulated with instruction or practice, the relationship between ability and strategy sophistication may vary. Effective top-down reasoning is a defining feature of cognitive ability. However, instruction or prior knowledge can presumably facilitate the discovery of such insights. Thus, it may be that across the three studies reported in this thesis, as the degree of instruction increases, the correlation between ability and strategy sophistication may decline.

3.1.5 Strategy Sophistication and Performance

In considering models of the relationship between practice, strategy, and performance, it is necessary to consider the correlation between strategy sophistication and performance. As a matter of interest, sophisticated strategies are generally chosen for study because they are superior in some sense. Thus, it is expected that strategy sophistication will correlate positively with performance. However, the size of this correlation should depend on task and strategy characteristics.

In general, studies with larger correlations between strategy sophistication and performance are likely to have a more similar pattern of correlations with other predictors. Researchers looking at age differences on algorithm-retrieval tasks have found that older adults are less likely to use a retrieval strategy and are slower on both retrieval and algorithm strategies (e.g., Touron et al., 2004). Given this, it seems plausible that strategy sophistication partially mediates the effect of predictors on task performance. Ability may lead to greater insight into strategy superiority, and this represents one mechanism by which ability measures produce correlations with performance.

In summary it was hypothesised that:

Hypothesis 3.5 *Strategy use partially mediates the relationship between ability and prior experience on task performance.*

3.1.6 The Current Studies

The results presented in this chapter are based on the same three studies presented in the previous chapter. Each of these studies measured demographics and prior experience. Studies 1 and 2 measured ability, and Studies 2 and 3 measured Big 5 personality. All studies measured strategy sophistication and task completion time on a text editing task. Performance was aggregated over practice to form an overall measure.

3.2 Method

3.2.1 Participants

Participants for Study 1 were the same as reported in Chapter 2. For Study 2 measures were collected at different times points and resulted in some missing data among the retained sample: 131 (85.1%) completed the personality test, 143 (92.9%) completed the paper ability tests, 149

(96.8%) completed the computerised ability, typing, and prior experience measures. By analysing performance only on the performance blocks prior to the mid-practice training, Study 3 analyses were based on the whole sample ($n = 154$) across the three conditions, but only a subset of participants ($n = 104$) completed the personality test.

3.2.2 Predictor Measures

Overview

The following sections describe the individual difference measures in the three studies organised under the categories of demographics, prior experience, typing speed, ability, and personality. Each study used a different but overlapping set of measures as outlined in Table 2.5 on page 58, Table 2.9 on page 65, and Table 2.11 on page 74. In brief, Study 1 did not have personality measures, Study 2 measured both personality and ability, and Study 3 did not measure ability.

Demographics

All studies measured age and gender.

Prior Experience

Study 1 Prior Experience Scale The measure contained 17 items. Each item was one of three types. Five items assessed self-reported experience and physiological readiness, five items measured self-reported competence in word processing tasks, and seven items tested participant knowledge of text editing shortcut keys. Each response option was assigned a weight related to the degree to which it was theorised to reflect prior text editing experience. Items and weights are shown in appendix A.2.1. To calculate a measure of prior experience, a raw score was first calculated as the sum of the weights for the selected response options.

The raw score was converted to a scale score such that the minimum possible score was zero and the maximum possible score was 100. This can be expressed in mathematical notation as

$$PE_{\mathsf{raw}} = \sum_{i=1}^{k} PE_i,$$

and

$$PE_{\mathsf{scaled}} = 100 \times \frac{PE_{\mathsf{raw}} - \min(PE_{\mathsf{raw}})}{\max(PE_{\mathsf{raw}}) - \min(PE_{\mathsf{raw}})},$$

where PE_i is the weight for the $i = 1, ..., k^{th}$ prior experience question, and $\max(PE_{\mathsf{raw}})$ and $\min(PE_{\mathsf{raw}})$ refer to the maximum and minimum possible scores respectively.

Study 2 Prior Experience Scale A quick two-item measure of prior experience was used that showed good predictive validity (see correlations in Table 3.4). Question 1 (Q_1) was "when typing do you look at the keys?" and Question 2 (Q_2) was "prior to the training in the previous class, how often did you use the SHIFT key to select text?" Both questions had response options: 1= "Never", 2 = "Almost Never", 3 = "Sometimes", 4 = "Usually", 5 = "Always". Prior experience was calculated to be $\frac{Q_2+6-Q_1}{2}$. A third question, "On average, how many hours per week do you spend using a computer?", was considered for inclusion but showed no predictive validity and hence was excluded.

Study 3 Prior Experience Scale The scale included 14 questions. Each question was closed-ended and each response option was assigned a non-negative integer weight. Larger weights were assigned to response options that were theorised to be associated with better performance on the text editing task. Question wording, response option wording, weights, and the rationale for the weights are shown in Appendix C.1.1. Items were generally good predictors of task performance. The weights were developed based on knowledge of the factors that influence individual differences in text editing. This knowledge in turn was based on (a) observations of participants in previous text editing experiments, (b) task analysis, and (c) theories of transfer. The method of calculating prior experience was the same as that used for the *Study 1: Prior Experience Measure*.

Typing Speed

Ten Thumbs Typing Test The typing test was locally developed (see Armstrong, 2000). Participants were required to type random grammatically correct sentences for two minutes. If participant typed an incorrect letter the program would not allow the cursor to proceed until the correct letter had been entered. The measure of typing speed was average words typed per minute.

Anglim Typing Test Studies 2 and 3 used an improved measure of typing speed. The typing test was developed and reported in Anglim and Waters (2007) where it obtained a high internal consistency reliability with Cronbach's alpha equal to 0.98. The typing test involved three trials each of one minute duration. Each trial involved typing a specified passage of text. The passages of text were taken from Wikipedia and were selected because they used common English words and standard punctuation. Participants were instructed that their score would be the number of words that they typed in one minute, minus the number of words they typed incorrectly. There were no explicit constraints on what the participants typed and no feedback was provided as to whether they had typed the text correctly or incorrectly. After the passage of one minute the trial automatically ended. Participants pressed `Enter` to start the next trial. The program was implemented in Inquisit version 3.2. See Appendix B.1.2 for further details.

An algorithm was devised to check the accuracy of the text entered by participants. The final score was the number of characters typed multiplied by the accuracy percentage. The metric was converted from characters per minutes to words per minute using the standard conversion of five characters, including spaces, representing one word. A participant's final score was the mean of the three trials.

The task has been used in several additional studies unrelated to this thesis and has shown to have high internal consistency reliability in university samples (i.e., Cronbach's alpha above 0.90, Anglim & Waters, 2007). The improvements over the test used in Study 1 were: (a) the ability to estimate reliability, (b) a longer period of measurement, (c) increased logging of participant responses, (d) more precise control over trial duration, (e) removal of the confusion that occasionally resulted when participants typed the wrong letter of a word.

Ability Tests

Study 1 included nine different ability tests, two psychomotor tests were updated in Study 2. Table 3.1 summarises these tests. Selected tests corresponded to appropriate points on Ackerman's (1988) multidimensional scaling of 31 reference tests based on data from Allison (1960). Consistent with Ackerman's (1988) terminology, the nine tests were classified into one of three ability classes: (a) general cognitive ability, (b) perceptual speed ability, or (c) psychomotor ability. Factor analysis in Armstrong and Langan-Fox (2000) supported the three factor structure. Each test was scored as set out in Table 3.1. A measure for each of the three ability classes was obtained by converting constituent tests into z-scores and summing the z-scores. This sum was then converted to a z-score. For purposes of mediational analyses, an overall ability measure was formed by summing the z-scores of the three ability classes.

Table 3.1: Summary of Ability Tests Used

Abbreviation	Name	Reference	Ability Measured	Description	Scoring	S1[a]	S2[a]
ERV	Extended Range Vocabulary Test	Ekstrom et al. (1976)	General	Two parts; 24 items per part; 6 minutes per part; choose which of five words is most similar in meaning to a target word	Percentage correct after subtracting a quarter mark for each item incorrect	X	X
IT	Inference Test	Ekstrom et al. (1976)	General	Two parts; 10 items per part; 6 minutes per part; choose which of five statement repersenta a logical inference	Percentage correct after subtracting a quarter mark for each item incorrect	X	X
CC	Cube Comparison Test	Ekstrom et al. (1976)	General	Two parts; 21 items per part; timed test with 3 minutes per part; choose whether two cubes are the same or different	Percentage of total possible correct after subtracting one mark for each item incorrect	X	X

Table continues on next page

Abbreviation	Name	Reference	Ability Measured	Description	Scoring	S1[a]	S2[a]
NC	Number Comparison Test	Ekstrom et al. (1976)	Perceptual Speed	Two parts; 90 seconds per part; indicate pairs of numbers that are different	Total count of items correctly marked minus incorrectly marked	X	X
NS	Number Sort	Ekstrom et al. (1976)	Perceptual Speed	Two parts; 90 seconds per part; indicate which of five numbers is largest (part 1) or smallest (part 2)	Total count of items correctly marked minus incorrectly marked	X	X
CS	Clerical Speed and Accuracy	Ekstrom et al. (1976)	Perceptual Speed	Two parts 3 minutes per part; read a manual to find a marked item and mark the marked item in the answer sheet	Total count of items correctly marked minus incorrectly marked	X	
RT1	Simple Reaction Time	Anglim (2000a)	Psychomotor	40 trials; respond by pressing F5 to target as quickly as possible	Mean reaction time for correct trials excluding outlier trials	X	
RT2	2-Choice Reaction Time	Anglim (2000a)	Psychomotor	40 trials; respond by pressing F5 of F6 corresponding to two response options	Mean reaction time for correct trials excluding outlier trials	X	

Table continues on next page

Abbreviation	Name	Reference	Ability Measured	Description	Scoring	S1[a]	S2[a]
RT4	4-Choice Reaction Time	Anglim (2000a)	Psychomotor	40 trials; respond by pressing F5, F6, F7, or F8 corresponding to four response options	Mean reaction time for correct trials excluding outlier trials	X	
RT1V2	Simple Reaction Time Version 2		Psychomotor	40 trials; respond by pressing F5 to target as quickly as possible	Mean reaction time for correct trials excluding outlier trials		X
RT4V2	4-Choice Reaction Time Version 2		Psychomotor	40 trials; respond by pressing F5, F6, F7, or F8 corresponding to four response options	Mean reaction time for correct trials excluding outlier trials		X

[a] X indicates that the test was included in the study.

Personality

IPIP Personality Test The International Personality Item Pool (*IPIP*) is a self-report measure that includes a measure of the Big 5 factors of extraversion, conscientiousness, neuroticism, agreeableness, and openness/intellectance (L. R. Goldberg et al., 2006). Study 2 used a 100 item version, and Study 3 used a 50 item version. Each factor had an equal number of items, and each scale had a similar numbers of positively and negatively worded items. Both versions have shown good internal consistency reliability (L. R. Goldberg, 1999). Each item asked the participant to indicate the degree to which the given statement described himself or herself on a 5-point scale from 1 = very inaccurate, to 5 = very accurate. Scale scores were computed as the mean score after reversing negatively worded items, where a reversed item was $6-x$, and x is the item score. For details of the instructions given see Appendix B.1.1. For a complete list of items and assigned factors, see Appendix B.1.1 for the 100 item version, and Appendix C.1.2 for the 50 item version.

3.2.3 Criterion Measures

Task completion time and strategy sophistication were all calculated as the mean level over the period of practice in the study. In Study 3, the mean was based only on the first 15 blocks in order to use a consistent and complete sample.

3.2.4 Analysis Plan

Correlations were used to examine the strength of relationship between predictor and criterion measures. Several separate mediational models were conducted to test whether the effect of several predictors on task performance was mediated by strategy sophistication. Separate models were conducted for typing speed, prior experience, and a composite ability measure. The Sobel Test (Sobel, 1982) was used to evaluate the significance of the indirect effect.

3.3 Study 1 Results

3.3.1 Preliminary Analyses

Split-half reliability estimates of general ability tests (Inference = 0.78; Vocabulary = 0.62; Cube Comparison = 0.65) and perceptual speed tests (Number Comparison = 0.62; Number Sort = 0.63; Clerical Speed = 0.89) varied between reasonable and very good. An initial factor analysis was performed on the nine ability tests to assess whether it was appropriate to combine the tests into three ability classes. Psychomotor tests were reversed such that higher scores meant more of the ability. A maximum likelihood factor analysis with Promax rotation was performed extracting three factors. Factor loadings and correlations between factors are shown in Table 3.2. The first three factors accounted for 38.1, 15.7, and 9.0 percent of variance respectively. Given the reasonable factor analytic support and the theoretical rationale, the three ability composites were created as set out in the Method.

The ability tests showed reasonable consistency with the proposed factor structure with three main deviations. First, the cube comparison test did not correlate highly with the other cognitive measures, and showed some cross-loadings with perceptual speed tests. Possibly, the requirements for spatial rotation and speeded completion were distinct from the Inference and Vocabulary tests with their associated greater requirement for verbal knowledge and reasoning ability. Second, 2–Choice and 4–Choice RT were more closely related to each other than they were with Simple RT. Third, Number Sort and Number Comparison were more correlated with each other than they were with Clerical Speed and Accuracy. These latter two points are consistent with similarities in information processing requirements of the tests. Despite these deviations from a clean factor structure, the three ability measures were still constructed as the z-score of the sum of z-scores of constituent tests. Although not reported, multidimensional scaling and hierarchical cluster analyses were also performed on the reversed correlation matrices between ability tests and told a similar story as the factor analysis.

Table 3.2: Study 1 Factor Loadings and Intercorrelations among Ability Tests

	Ability test	I[a]	II[b]	III[c]
1.	Inference	.08	−.03	**.99**
2.	Vocabulary	−.16	.03	**.61**
3.	Cube Comparison	.09	.18	.29
4.	Number Comparison	−.07	**.87**	−.01
5.	Number Sort	−.07	**.75**	.09
6.	Clerical Speed and Accuracy	.20	**.65**	−.18
7.	Simple RT	**.77**	−.05	.15
8.	2-Choice RT	**.95**	−.02	−.02
9.	4-choice RT	**.88**	.01	.02
	Factor intercorrelations			
I.	Psychomotor			
II.	Perceptual Speed	.45		
III.	General	.22	.44	

Note. Loadings > |.30| are shown in bold.
[a] Psychomotor Ability; [b] Perceptual Speed Ability; [c] General Ability.

3.3.2 Hypothesis Testing

Table 3.3 shows descriptive statistics and intercorrelations for predictor and criterion variables. Typing speed (supporting Hypothesis 3.1) and prior experience (supporting Hypothesis 3.2) were both strong predictors of task completion time. Moderate to strong prediction was obtained between the three ability tests and task performance, broadly supporting Hypothesis 3.3. The trend in the data was consistent with expectations that younger adults and males would have faster task completion times, although the correlations were not statistically significant.

Table 3.3: Study 1 Descriptive Statistics and Intercorrelations for Task Performance, Demographics, Ability, and Prior Experience Variables

	Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9
1.	Task RT (s)	32.81	8.71	—								
2.	Strategy Sophistication (0 - 1)	0.22	0.25	−.47	—							
3.	Male[a]	0.30	0.46	−.18	.04	—						
4.	Age (years)	23.05	4.43	.16	−.16	.30	—					
5.	Typing Speed (wpm)	33.30	10.65	−.53	.14	−.19	−.07	—				
6.	Prior Experience (0 - 100)	66.08	12.78	−.53	.21	.08	−.01	.49	—			
7.	General Ability (z-score)	0.28	0.98	−.29	.11	−.06	−.01	.23	.06	—		
8.	Perceptual Speed Ability (z-score)	0.24	0.89	−.41	−.04	.10	−.10	.22	.02	.26	—	
9.	Pscyhomotor Ability (z-score)	0.30	0.73	−.50	.36	.28	−.31	.13	.11	.04	.31	—

Note. $N = 63$; $r > |0.25|$ is significant at $\alpha = .05$.
[a] Female = 0 and Male = 1.

Three mediation analyses were performed assessing whether the effects of ability, typing speed, and prior experience, respectively, on task performance were mediated by strategy sophistication. The standardised indirect effects were all non-significant: ability (indirect effect $= -0.905, z = -1.50, ns$), typing speed (indirect effect $= -0.0458, z = -1.05, ns$), prior experience (indirect effect $= -0.054, z = -1.53, ns$). Strategy sophistication was strongly related to task performance. Psychomotor ability was the only statistically significant predictor of strategy sophistication, although prior experience was approaching significance. Failing to support Hypothesis 3.5, results were broadly consistent with the idea that strategy sophistication and various forms of aptitude provide independent contributions to performance. Results also indicate that ability and prior experience predict task performance better than strategy sophistication.

3.4 Study 2 Results

3.4.1 Preliminary Analyses

Split-half reliabilities for the five ability tests were as follows: Inference = 0.69; Vocabulary = 0.49; Cube Comparison = 0.61; Number Comparison = 0.77; Number Sort = 0.36. Cronbach's alpha reliability for the five personality factors were all reasonably good: extraversion = 0.93; agreeableness = 0.84; conscientiousness = 0.88; emotional stability = 0.93; openness = 0.91. Typing speed had a Cronbach's alpha of 0.96.

3.4.2 Hypothesis Testing

Table 3.4 shows descriptive statistics and correlations for predictors and criterion measures. Typing speed (supporting Hypothesis 3.1) and prior experience (supporting Hypothesis 3.2) were both strong predictors of task performance. Moderate to strong predictions were obtained between the three ability tests and task performance, broadly supporting Hypothesis 3.3. Males performed better on the task, although no relationship was observed with age, perhaps reflecting the minimal variation in age in the sample.

Consistent with Hypothesis 3.4 there was minimal evidence of a correlation between personality and task performance. Openness was the one exception to this. One interpretation of this significant correlation is that it is a Type I error resulting partially from running five significance tests. Another interpretation is that the IPIP measure of openness is related to self-perceived intelligence. This latter interpretation is supported by the relatively strong correlation between openness and general ability.

Table 3.4: Study 2 Descriptive Statistics and Intercorrelations for Task Performance, Demographic, Ability, and Prior Experience Variables

	Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	Task RT (s)	7.63	1.76	—													
2.	Strategy Sophistication (0 - 1)	0.59	0.26	−.18	—												
3.	Male[a]	0.26	0.44	−.32	.05	—											
4.	Age (years)	21.14	1.60	−.02	.09	.17	—										
5.	Typing Speed (wpm)	53.40	15.33	−.46	.25	−.08	−.03	—									
6.	Prior Experience (0 - 100)	3.02	0.96	−.51	.14	.16	.12	.52	—								
7.	General Ability (z-score)	0.11	0.97	−.45	.17	.18	.05	.29	.15	—							
8.	Perceptual Speed Ability (z-score)	0.05	1.02	−.34	.17	.04	−.17	.22	.11	.15	—						
9.	Psychomotor Ability (z-score)	0.04	1.06	−.24	.28	.16	−.10	.12	.09	.28	.11	—					
10.	Extraversion (1 - 5)	3.43	0.64	.08	.07	−.03	−.02	.09	.00	.03	.00	−.18	—				
11.	Agreeableness (1 - 5)	3.96	0.41	−.03	.21	−.09	.09	.07	−.05	.16	.12	−.05	.43	—			
12.	Conscientiousness (1 - 5)	3.42	0.55	−.05	.04	−.16	.05	.13	−.05	−.05	.11	−.16	.24	.20	—		
13.	Emotional Stability (1 - 5)	3.09	0.71	−.13	.14	.20	.07	.03	.14	.02	−.03	.11	.23	.06	.02	—	
14.	Openess (1 - 5)	3.68	0.54	−.20	.06	.19	.12	.15	.19	.49	−.01	−.04	.43	.49	.18	.18	—

Note. Correlations were calculated based on pairwise deletion of missing data. Personality variables are variables 10 to 14. Correlations not involving a personality variable had $N = 154$; if $r > |0.16|$, then $p < .05$. Correlations involving a personality variable had $N = 131$; if $r > |0.18|$, then $p < .05$.

[b]The variable 'Male' is coded such that 1 is Male and 0 is Female.

Predictors of psychomotor ability, typing speed, and surprisingly, agreeableness, correlated with strategy sophistication. Strategy sophistication correlated to a small to moderate extent with task performance. Three mediation analyses were performed assessing whether the effect of ability, typing speed, and prior experience, respectively, on task performance were mediated by strategy sophistication. The standardised indirect effects were all non-significant: ability (indirect effect $= -0.0100, z = -0.220, ns$), typing speed (indirect effect $= -0.002, z = -0.875, ns$), and prior experience (indirect effect $= -0.0288, z = -1.15, ns$). Thus, Hypothesis 3.5 was not supported.

3.5 Study 3 Results

3.5.1 Preliminary Analyses

Cronbach's Alpha reliabilities for the personality scales were good: extraversion = 0.92, agreeableness = 0.78, conscientiousness = 0.85, emotional stability = 0.9, openness = 0.84. Typing speed had a Cronbach's alpha of 0.96.

3.5.2 Hypothesis Testing

Table 3.5 shows descriptive statistics and intercorrelations between predictor and criterion measures. Although correlations were slightly less than expected, typing speed (supporting Hypothesis 3.1) and prior experience (supporting Hypothesis 3.2) were both moderately strong predictors of task performance. Consistent with expectations, males and younger adults performed better on the task. Consistent with Hypothesis 3.4, there was minimal evidence of a correlation between personality and task performance. Surprisingly conscientiousness had a small negative correlation with task performance.

Table 3.5: Study 3 Descriptive Statistics and Intercorrelations for Task Performance, Demographic, Prior Experience, and Personality Variables

	Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1.	Task RT (s)	7.51	1.94	—										
2.	Strategy Sophistication (0 - 1)	0.60	0.30	−.34	—									
3.	Male[a]	0.29	0.45	−.37	.27	—								
4.	Age (years)	22.46	5.05	.31	.06	.05	—							
5.	Typing Speed (wpm)	51.73	15.82	−.41	.29	.06	−.20	—						
6.	Prior Experience (0 - 100)	46.37	12.28	−.46	.42	.44	−.11	.35	—					
7.	Extraversion (1 - 5)	3.32	0.92	.04	−.07	−.04	−.14	−.05	−.08	—				
8.	Agreeableness (1 - 5)	4.19	0.49	.03	−.04	−.17	−.00	−.02	−.04	.17	—			
9.	Conscientiousness (1 - 5)	3.48	0.74	.23	−.14	−.07	.17	−.03	−.23	−.01	.07	—		
10.	Emotional Stability (1 - 5)	2.97	0.88	−.16	.10	.19	.18	.13	.16	.26	.09	.19	—	
11.	Openness (1 - 5)	3.81	0.65	−.09	.23	.16	.04	.04	.16	.02	.32	.11	.07	—

Note. Correlations were calculated based on pairwise deletion of missing data. Personality variables are variables 7 to 11. Correlations not involving a personality variable had $N = 149$; if $r > |0.17|$, then $p < .05$. Correlations involving a personality variable had $N = 104$; if $r > |0.20|$, then $p < .05$.

[a]The variable 'Male' is coded such that 1 is Male and 0 is Female.

Two mediation analyses were performed assessing whether the effects of typing speed and prior experience, respectively, on task performance were mediated by strategy sophistication. Typing speed was a moderate predictor and prior experience was a strong predictor of strategy sophistication. Strategy sophistication correlated to a small to moderate extent with task performance. The standardised indirect effects were all non-significant: typing speed (indirect effect $= -0.00793, z = -2.30, p < .05$), prior experience (indirect effect $= -0.0110, z = -1.93, ns$). Thus, Hypothesis 3.5 was not supported.

3.6 Discussion

3.6.1 Overview

This chapter examined the relationship between predictor and criterion measures on the text editing task. Overall, with the exception of the mediational hypothesis, hypotheses and expectations were supported. The discussion is organised in terms of predictors of task performance, and the mediational model of predictors, strategy sophistication, and performance.

3.6.2 Performance

Demographics

The general pattern of results were consistent with expectations. Males and younger adults generally performed better at the text editing task.

Gender differences were observed in all three studies, albeit only as a trend towards significance in Study 1. In Study 3, males had greater prior experience, a strong predictor of performance, yet in the other two studies, gender showed minimal correlations with other predictors. It may be that males were more competitive, or had greater prior experience with computers, computer games, and text editing. Study 3 provided some support for attributing differences to known relevant sources of prior experience. However, in Studies 1 and 2 other possible predictors such as typing speed did not correlate with gender.

The negative correlation between age and performance was small. It emerged as a trend in Study 1 and a small significant effect in Study 3. Study 2 lacked meaningful variation in age to test the relationship. Equally, the amount of variation in age was also fairly low in Studies 1 an 3. In broad terms the findings are consistent with both cognitive aging and cohort effects related to generational increases in use of computers. However, given the relatively small variance in age across the studies and the convenience sampling, the interpretation of the observed age related differences should be treated with caution.

Ability

Ability was a moderate to strong predictor of task performance. In Study 1 psychomotor ability had the strongest correlation followed by perceptual speed and then general ability. In Study 2 the order of the size of correlations was reversed. Given the large sample size in Study 2 it should be given more weight. Random sampling is one plausible explanation for the differences between the results. It is also possible that abilities and prior experience may have differed between Study 1 with its community sample and Study 2 with its third year university sample.

It is also worth considering the degree to which the ability measures chosen provide a basis to generalise to other potential ability measures. Other researchers might have preferred that other particular ability measures were employed such as an operation-span test to measure multitasking or a working memory measure. The three classes of ability tests were chosen in order to mirror the framework often adopted by Phillip Ackerman. The list of possible ability tests is quite large. Expanding the pool of tests would have been empirically interesting, but the multiplicity of predictors can create challenges to interpretation. It can also be argued that interpretation of the cognitive correlates approach should be treated with caution. Just as correlation is not necessarily causation, the relative size of a correlation does not necessarily correspond to relative size of causation. Correlation with external variables, degrees of test validity and reliability, and the general intercorrelated nature of ability testing should temper causal interpretation of ability-performance correlations.

Prior Experience

Prior experience was a strong predictor of task performance in all three studies both when it was operationalised as typing speed and as self-reported experience. This is consistent with cognitive correlates, transfer, and correlation specificity perspectives.

Personality

In both Studies 2 and 3 where personality was measured, self-reported measures of personality rarely predicted task performance. A small positive correlation with openness in Study 2 and a small negative correlation with conscientiousness in Study 3 with task performance was observed. It seems plausible that when openness measures include items related to self-perceived intelligence, that correlations could emerge. However, given that ten significant tests were performed and the lack of an a priori reason to expect such correlations, a reasonable interpretation is that population correlations are close to zero.

In summary findings are consistent with personality being unrelated to task performance on an experimental text editing task. They are also consistent with the notion that performance on the text editing task is a skill influenced more by 'can do' than by 'will do' factors. In particular, conscientiousness, which has been found by meta-analyses to be related to workplace performance was unrelated to task performance. A related interpretation is that the correlation is particularly small, perhaps less than $r = .10$. Even if this were the case, personality is arguably of minimal importance.

Skill acquisition experiments may better relate to what Paul Sackett (P. Sackett, Zedeck, & Fogli, 1988) calls 'maximal performance' rather than 'typical performance'. Maximal performance is believed to occur over short periods of time where performance is being monitored. Skill acquisition experiments are highly monitored environments where participants are asked, and may often feel obliged, to apply maximal effort.

3.6.3 Mediational Model

It was hypothesised that the effect of ability, prior experience, and typing speed on task performance would be mediated by strategy sophistication. In all three studies, and for all three predictor variables, the mediation model was not supported. Prior experience was a reasonable predictor of strategy sophistication particularly in Study 3. Likewise typing speed showed some mixed evidence of a relationship with strategy sophistication. Strategy sophistication also correlated with task performance, although the correlation was relatively small given the potential efficiency gains that can result from the more sophisticated strategy.

The relationship between individual differences in strategy sophistication and performance in Study 2 was less than in Study 1. While random sampling might explain some of this difference, another intriguing possibility is that increased instruction of sophisticated strategies led to a drop in the observed correlation. In Studies 2 and 3 more participants may have been using the sophisticated strategies for the first time. In contrast, participants in Study 1 received minimal instruction to use the sophisticated keys. As such, use of the keys may have been more of a proxy for prior experience in text editing general.

The correlation between prior experience and strategy sophistication is consistent with the idea that using good strategies is a relatively domain specific adaptation that may often evolve slowly over time in natural environments.

Schunn and Reder (2001) proposed a theory that suggested that ability would correlate with adaptive strategy selection. In some respects, strategy sophistication captures the concept of adaptive strategy selection on the text editing task. If so, present results provide only mixed support for Schunn and Reder's theory. Psychomotor ability was the better predictor of strategy sophistication across the two studies that measured ability, although some small correlations were observed in Study 2 with general cognitive ability. It seems that factors related to prior experience, including prior experience itself, typing speed, and possible proxy measures, such as gender, are better predictors of strategy adaptivity. In this framework, strategy adaptivity is a domain specific adaptation, where in adult samples, prior experience is likely to overwhelm effects of more general adaptive abilities.

3.6.4 Conclusion

In broad terms the findings are consistent with several general principles regarding predictor–criterion relationships. First, the cognitive correlates perspective was supported with overlap between prior experience, typing speed, and text editing reflected in observed correlations. Second, the fact that prior experience and typing speed were generally stronger predictors than ability is consistent with the specificity perspective. Third, the lack of prediction by personality is consistent with the idea that task performance is about skill and not self-reported dispositions.

However, caution should still be applied when applying a causal interpretation to the cognitive correlates perspective. Theory and the observed results suggest several alternative interpretations of observed correlations. In particular, correlations may be inflated due to the nature of the sample or a third variable. In particular more plausible measures of past experience and skill may correlate with ability based measures in surprising ways.

Chapter 4

Instructed Strategy Shift

Contents

4.1 Introduction

4.1.1 Overview

Results presented in Chapter 2, showed that under self-directed conditions, individuals often initiate shifts from simple to sophisticated strategies. Such shifts are (a) often gradual but can be abrupt, (b) more likely to occur early in practice than later, and (c) rarely lead to observable discontinuities in performance. However, research suggests that different dynamics operate when strategies are introduced by external forces, such as instruction (e.g., Delaney et al., 1998; Yechiam et al., 2004). In such situations, following external instruction, strategy shifts may be more abrupt, and disruptions to task performance may occur. Such externally introduced strategies may result in what may appear as a new learning curve (e.g., Delaney et al., 1998).

This chapter addresses Aim 3 of the thesis. It examines the effect of instructed strategy shifts on task completion time and contrasts this with self-initiated strategy shifts. It examines both abruptness of strategy shifts and effects on task completion time following the introduction of a new strategy after a period of practice. It examines both the immediate effect on task completion time and the effect following subsequent practice. The literature review that follows is organised around the above two topics first focusing on strategy use, and then discussing effects on performance of instructed strategy shift.

4.1.2 Strategy

As discussed in Chapter 2 most models of strategy use and performance (e.g., Lovett, 2005; Lovett & Anderson, 1996; J. W. Payne et al., 1988; Schunn et al., 1997; Siegler & Shipley, 1995) share qualitatively similar features. In order for a strategy to be used, an individual must either explicitly or implicitly know both of its existence and how to use it. Individuals are more likely to use a strategy if they implicitly or explicitly believe that the strategy is superior to an alternative. Thus, strategy change is often assumed to be caused by a corresponding change to information about strategy existence, execution, or relative effectiveness. This is typically assumed to be stochastic, operating at the level of probabilities of strategy use. Information regarding strategy use can come from many sources including both internal sources, through interaction with the task environment and through external instruction.

In this chapter a distinction is made between self-initiated and instructed strategy shifts. The literature review focuses on how instructed strategy shifts operate, and how they may have different strategy use and performance dynamics. With regards to instructed strategy shifts, various distinctions can be made. Important types of information are (a) strategy existence, (b) how to execute a strategy, and (c) why to execute a strategy (i.e., the relative effectiveness of the strategy). The instructor can also vary the format of the instructions to be a command, request, suggestion, or provide information only.

An important literature for understanding the process of strategy shift and instruction is the concept of training needs. The first step before implementing a training program often involves implementing a training needs analysis (e.g., Moore & Dutton, 1978; Rossett, 1987). Training needs analysis is grounded in rational principles of utility maximisation. Training needs analysis generally recommends training for an individual where: (a) the individual does not currently have the skills taught, (b) the individual is in a zone of development where they would be able to learn the skills taught, (c) the skills taught are superior to any alternative that the learner currently uses to achieve a relevant goal, (d) the learner will have the opportunity to apply the skills taught, and (e) weighing the costs of training in terms of time spent engaged in off task behaviour, and in reduced effectiveness, and all the benefits of training in terms of improved effectiveness in the future, the training provides a rational net benefit. Such a perspective forms a backdrop for strategy instruction decisions. These principles of training needs analysis have implications for the expected effect of introducing instruction on strategy use and performance.

It seems likely that self-initiated strategy shifts will lead to greater improvement in performance. First, relatively automatic associative processes are likely to pull participants towards strategies that are effective in the short term. Second, understanding that a more sophisticated strategy exists may indicate a readiness to perform the strategy. Third, initiation and persistence with a more sophisticated strategy is more likely when the individual experiences the strategy as adaptive in the short term. In contrast, instructed-strategy shift is likely to lead individuals who may not be ready in the short term to adopt the strategy.

It should be noted that when participants are explicitly told to use a particular strategy, they generally will use it (e.g., last study reported in, Delaney et al., 1998). Delaney et al's (1998) final study involved teaching a single participant how to be an expert in mental calculation. When the participant was first taught and told to use a more sophisticated strategy, they temporarily dropped in performance before eventually improving to levels superior to before the introduction of the more sophisticated strategy.

When developing a theory of the effect of instructed versus self-initiated strategy shift, it is useful to consider the differences between implicit and explicit strategy use and strategy change. Several researchers have suggested that when strategies enter explicitly they are more likely to be abrupt (for a discussion, see Siegler & Stern, 1998). Instructed strategy shift almost by definition has to enter explicitly. Thus, at a basic level it is hypothesised that:

Hypothesis 4.1 *The combination of a request to use a strategy and information on how to use a strategy increases strategy use.*

It is also interesting to consider how features of a strategy influence the effect of instruction on subsequent strategy use. In the case of algorithm–retrieval strategy shifts, a learner can not immediately shift to retrieval by mere instruction. Research has shown that explicit instruction can increase the degree and speed of retrieval (Ackerman & Woltz, 1994), but such an increase is predicated on having an accessible representation in memory. With simple-sophisticated strategies it is generally possible to perform the sophisticated strategy following instruction, but performance may decline while the more sophisticated strategy is being learnt. In other cases, the sophisticated strategy may be too complex for the individual to perform immediately and may require additional off task behaviour to understand. Yechiam et al. (2004) provides and example of where using a more sophisticated scripting strategy required reading a manual and going through a process of trial and error in order to use the scripting strategy effectively.

In summary, research suggests that instruction can change awareness and motivation to implement a new strategy. Such instruction may often lead to strategy selection operating explicitly. It may also override normal strategy selection mechanisms that emphasise effectiveness of a strategy in the short term.

4.1.3 Performance

The previous chapters in this thesis showed that when individuals self-initiate a strategy shift, there is typically minimal change to performance. One explanation for this is that strategy selection is relatively adaptive. Individuals are inclined to shift when the short term rewards suggest that the strategy is effective. Thus, individuals who shift are more likely to shift when they are ready. It may also be that being aware of a strategy is related to a capacity to perform it at an adequate level. It is also possible that individuals already know the strategy, and the observed shift in strategy use is simply a change in perceptions about what is appropriate in the circumstances.

Delaney et al. (1998) proposed the idea that learning curves should be considered as learning curves within strategies. The final study in Delaney et al. (1998) was a case study that examined performance on complex multiplication. Phase one involved one strategy and phase two introduced a new strategy that was designed to be superior to the first. Several important findings emerged. First, performance on the first strategy improved in accordance with something approximating a power law. Second, performance on the second strategy involved an initial decline relative to the first strategy. Third, performance on the second strategy showed a pattern consistent with the commencement of a new learning curve.

Thus, learning a new strategy is like learning a new skill. In the case of strategies introduced by others, it is likely that the learner will be less prepared to use the strategy than those that naturally shift to the strategy. The requirement to use the new strategy can override normal strategy selection rules which limit the selection of slower strategies.

The amount of performance decline and the potential for a strategy to ultimately surpass performance of the initial strategy are likely to be specific to aspects of individual, task, and context. Nonetheless, the following hypotheses were formed as applicable to the transition from simple to sophisticated strategies on the text editing task and also to many other tasks permitting a simple–sophisticated strategy shift.

Hypothesis 4.2 *When sophisticated strategies are externally introduced later in practice, performance declines immediately after introduction.*

Hypothesis 4.3 *When sophisticated strategies are externally introduced later in practice, performance eventually surpasses the level that would have been attained had instruction not been received.*

4.1.4 The Current Study

This chapter reports additional results from Study 3. In addition to the group that practiced text editing without interruption, a second group was given additional instruction and training half way through practice on the text editing task. Chapter 2 only presented results for the No Training group. This chapter examines the differences in strategy use and performance between the Training and No Training groups over practice following the introduction of additional training.

4.2 Method

4.2.1 Design

The study used a 3×30 mixed design. Participants were randomly allocated to one of three conditions (Training, No Training, and Control). All participants completed 30 blocks of practice on the text editing task. After completing 15 blocks the experience of participants differed based on condition. Participants in the Training condition were given additional instruction on sophisticated text editing strategies, as well as explicit instructions to use these strategies on subsequent trials. Participants in the Control condition completed a four-choice reaction time task instead of receiving mid-practice training. Participants in the No Training condition proceeded to block 16 immediately after block 15. More details on the procedure was presented previously in Section 2.2.5.

4.2.2 Participants

Seventy six (49.4%) participants completed the Training condition, 17 (11.0%) completed the Control condition, and 61 (39.6%) completed the No Training condition. The sample size for the Control condition was deliberately smaller than the other two groups. The purpose of the Control condition was to ensure that the break alone was not sufficient to change strategy use or performance. The important research questions related to differences between the Training condition and the No Training condition as these conditions assessed most clearly the differences in the effect of practice and strategy shift on the learning curve under instructed and non-instructed settings.

4.2.3 Mid-Practice Condition

Training Condition

Participants in the Training condition following block 15 were given additional instructions on additional text editing strategies followed by an opportunity to practice these strategies. This condition was included in order to examine strategy use and performance following explicit instructions to use particular strategies in contrast to self-discovered strategy shifts.

Training condition instructions were presented on the computer over five screens. To maximise the chance that participants actually read the instructions a specified time was required to pass before participants were permitted to advance through the instruction screens. The verbatim text of the instructions is presented in Appendix C.2.3. In brief the instructions (a) outlined what the training involved; (b) presented theory on text editing, describing text editing in terms of movement, selection, and manipulation, and encouraging participants to consider long term efficiency that results from reducing the number of key presses required to perform a given editing task; (c) outlined strategies to speed up movement; (d) outlined strategies to speed up deletion; and (e) explained Training condition trials.

Participants in the Training condition completed six training trials. These were further composed into three sub-trials where participants were asked to (a) attempt to minimise key presses, (b) follow a specific strategy displayed on the screen, and (c) repeat the previously displayed sequence without assistance. Each of the six training trials involved different text. At the end of each sub-trial, participants were given feedback on the number of key presses they made. If they completed the task in the minimum number of key presses, they were also given the feedback “Well Done”. Otherwise they were given feedback, “Keep trying to reduce your key presses”. At the end of the six trials, participants were presented with the following message: “The additional training has ended. For remaining trials, try to use the strategies used in the training trials.”

Control Condition

The Control condition consisted of a filler task after block 15 of the text editing task. Participants were given the following message:

> You are now having a break from the main text editing task in order to restore your energy. The following task is a four-

> choice reaction time task. Place your left middle and index fingers on the D and F and your right index and middle fingers on J and K. Press the appropriate key in response to the red light as quickly as possible. Press Enter To Continue.

The reaction time task went for two minutes and was not used for measurement purposes. The condition was included for comparison purposes with the Training condition. It was designed to help attribute changes in strategy use and performance in the Training condition to the instructions as opposed to the break from task performance.

No Training Condition

The No Training condition involved no break from the text editing task. The condition was included in order to explore how strategies and performance change with practice in the absence of explicit instruction. In this sense, it is similar to the design of Studies 1 and 2.

4.3 Results

4.3.1 Practice and Strategy

Group-level

Figure 4.1 shows the relationship between practice and strategy sophistication for each condition at the group-level. The No Training and Control conditions showed decelerating functional relationships seemingly approaching an asymptote. In contrast, the Training condition had an abrupt increase in strategy sophistication following the mid-practice training. Supporting Hypothesis 4.1, between Blocks 15 and 16 the Training condition increased by 11.2 percentage points, whereas the other conditions showed little change (No Training changed by -0.2 percentage points; Control changed by -1.3 percentage points). This abrupt shift in strategy sophistication suggests that the mid-practice training had at least some of the intended effect of increasing strategy sophistication. The data also highlights that many participants were already using sophisticated strategies prior to training and that training did not make all participants sophisticated strategy users.

While the Control condition did not change immediately following the Mid-Practice break, it is interesting to note that in the second block following the break, strategy sophistication was often more similar to the Training condition than to the Control condition. Given the substantially smaller sample size in the Control group, this may be attributed to random sampling. The Control condition also seemed to have slightly higher levels of strategy sophistication prior to the break. When comparing blocks 12 to 15 with blocks 16 to 19, the Training condition increased by 13.5 percentage points, the No Training condition by 2.4 percentage points, and the Control condition by 7.2 percentage points. Pairwise comparisons of these change scores between conditions using independent groups t-tests showed that only Training and No Training conditions were significantly different ($p < .001$). The unadjusted p-value for Training versus Control was 0.07 and No Training versus Control was 0.07.

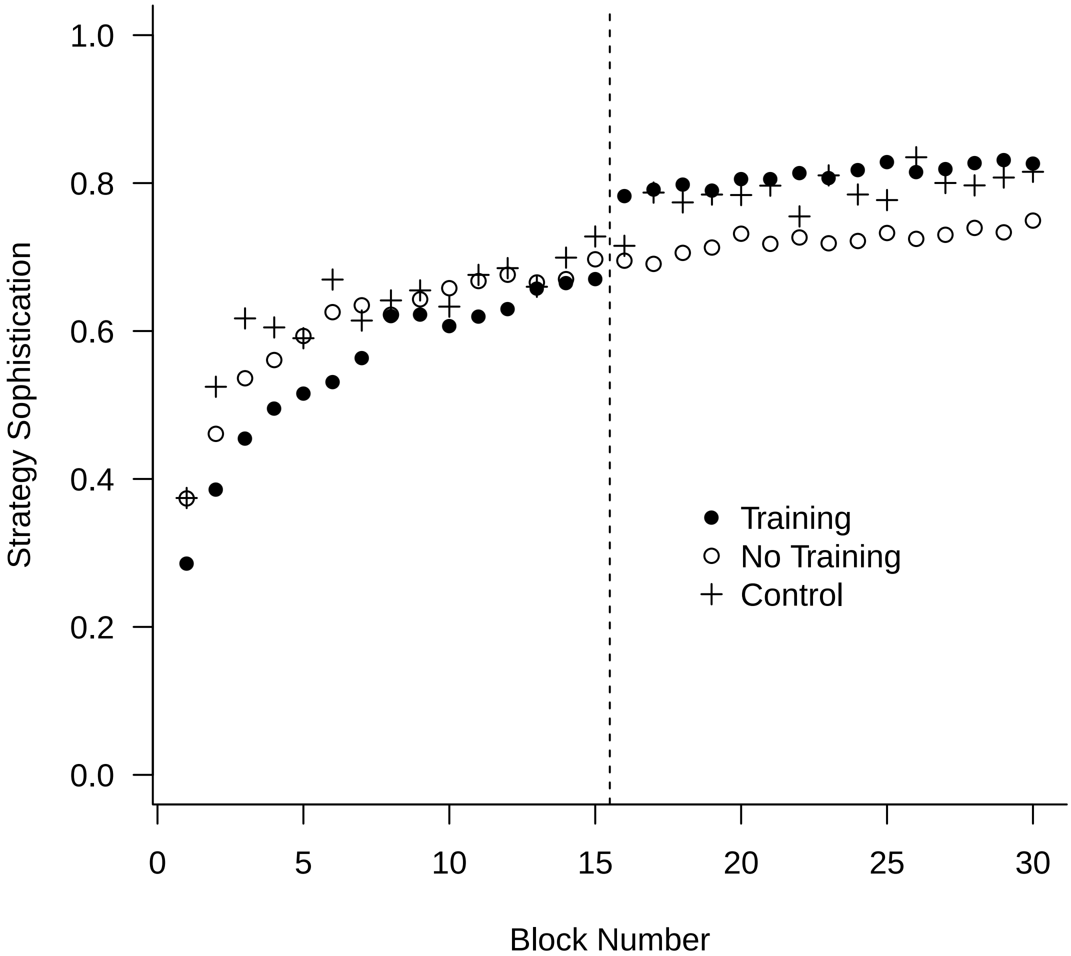

Figure 4.1: Strategy sophistication by block and condition at the group-level. Dotted vertical line indicates timing of mid-practice training in the Training condition.

Individual-Level

Figure 4.2 shows the relationship between practice and strategy sophistication at the individual-level. Many of the same patterns observed in the No Training condition (see Figure 2.24) can also be observed in the Training condition. The main distinction is the change in strategy sophistication observed after the Mid Practice training.

Many participants had already adopted a relatively sophisticated strategy prior to the mid-practice training. In some cases participants used a sophisticated strategy from the first block (e.g., cases 2, 16). Other participants had attained a sophisticated level over the course of the first 15 blocks. In such cases, participants were unable to shift to a more sophisticated strategy because they had no training need. Other participants were at an intermediate level of strategy sophistication at the time of the mid-practice training and did not increase sophistication further (e.g., cases 113, 176). It is unclear whether such participants were unaware of the potential increases in sophistication that could be attained, or whether they chose not to adopt such approaches.

Participant 14 dabbled with more sophisticated strategies after the Mid Practice Training and then reverted to simpler strategies. Inspection of Figure 4.4, shows how the participant experienced a fairly abrupt decline in performance while trying to use the more sophisticated strategy. One interpretation is that external instruction was insufficient to overcome the substantial drop in performance the participant experienced following the adoption of the more sophisticated strategy.

Several participants permanently increased strategy sophistication following the mid-practice training (e.g., cases 23, 24, 34, 35, 45, 46, 53, 71, etc.). In particular, many of these shifts were abrupt both in onset and in the level of strategy sophistication. The overall pattern highlights how external instruction can lead to more abrupt strategy shifts than when strategy shifts are self-initiated.

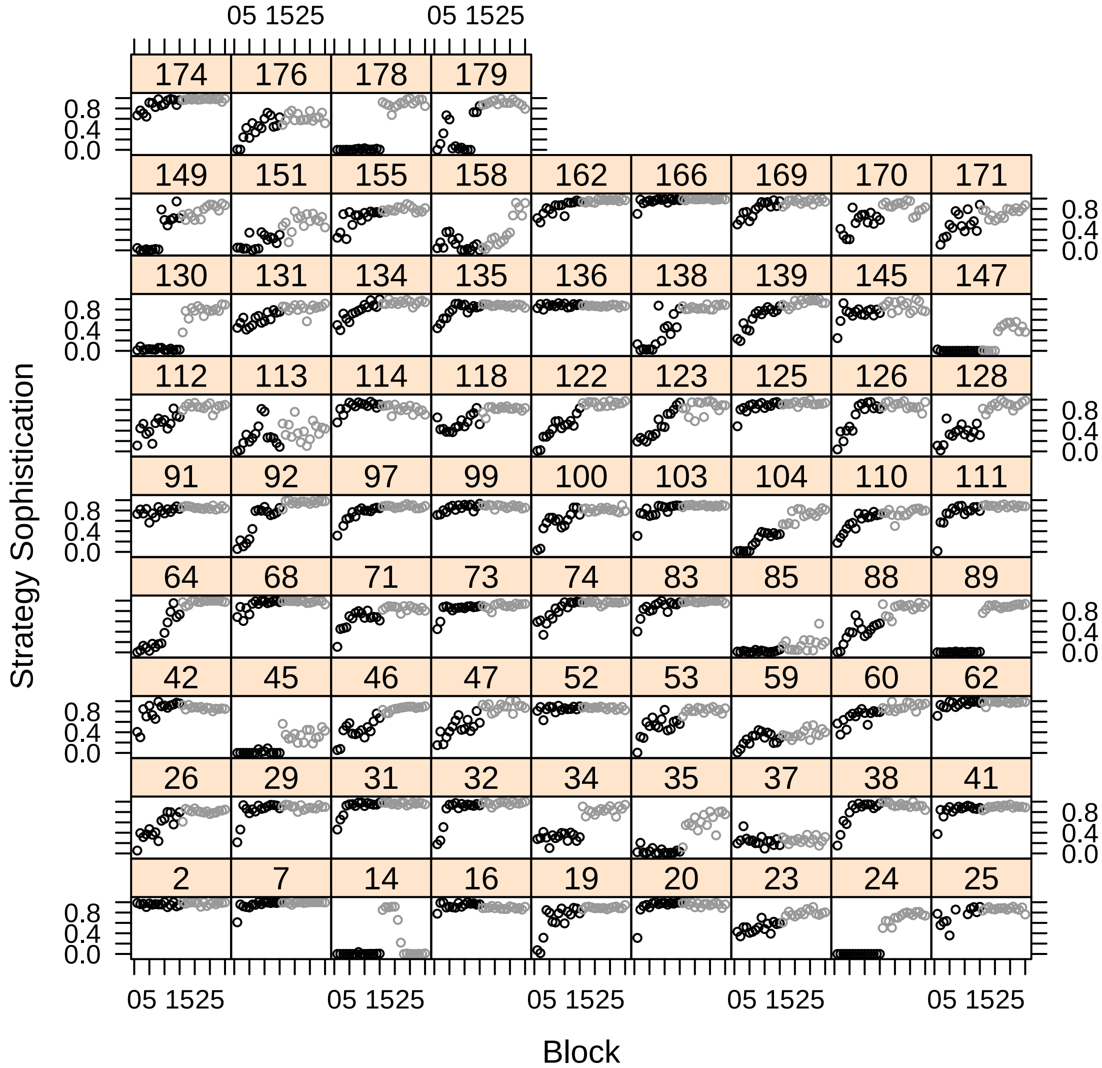

Figure 4.2: Strategy sophistication by block in the Training condition at the individual-level. Black points indicate blocks before the mid-practice training and grey points after the mid-practice training.

4.3.2 Practice and Performance

Group-level

Figure 4.3 shows the relationship between practice and task completion time by condition at the group-level. The functional form of the No Training condition was well fit by a three parameter power function. The Control condition was similar to the No Training condition. The greater volatility in the Control condition can be explained in terms of the smaller sample size. Importantly, the break from task performance in the Control condition did not appear to have an effect on task completion time.

Supporting Hypothesis 4.2, performance dropped substantially in the Training condition following the mid-practice training. A significant effect of condition was obtained at block 16, $F(2, 151) = 12.01, p < .001, \eta^2 = 0.137$. Tukey's Post Hoc Test showed a significant difference between the Training and the other two conditions ($p < .05$). By the end of practice in block 30, performance in the Training condition was similar to block 15 (block 15 = 6.17 block 30 = 6.11, and slightly worse than the other two conditions (Control block 30 = 5.00, No Training block 30 = 5.48), although not significantly so, $F(2, 151) = 2.67, p = .07, \eta^2 = 0.034$; Tukey's HSD was also non-significant for all pairwise comparisons at the .05 level. Thus, Hypothesis 4.3, which proposed that performance would be superior in the Training condition by the end of practice, was not supported.

The drop in performance in the Training condition is consistent with the change in strategy following the mid-practice training, which put individuals on a new learning curve. The absence of a similar change following Block 15 in the Control condition suggests that the drop in performance in the Training condition was not simply due to the break from the task.

Individual level: Training Condition

Figure 4.4 shows the relationship between practice and performance in the Training condition before and after the mid-practice training. The figure highlights how the immediate and final effect of the mid-practice training varied between individuals. Some participants had little initial or final effect (e.g., cases 151, 155). Some participants had an initial drop, which was sustained to the end of the experiment (e.g., cases 89, 158). Some participants dropped initially, but ultimately restored their performance (e.g., case 145) or surpassed their prior performance (e.g., case 88). A small number of participants obtained immediate benefits (e.g., case 45). Many participants were better at the end of practice than they were prior to the mid-practice training, although it is not clear whether this was due to the mid-practice training or to more general learning processes that may have occurred anyway.

Table 4.1 presents some quantitative summaries of these effects. In particular, the table shows that most participants in the Training condition were slower following the extra training. After four blocks, approximately half of the participants were at pre-training levels. By the final trials performance was similar to before the mid-practice training for participants on average. This is consistent with the idea that performance drops when there is a shift from a well-known to a novel strategy. It may also be that the strategies taught in training were only minimally more effective than those adopted naturally by participants.

Table 4.1: Change in Task Completion Time Pre- and Post- Training in the Training Condition

Comparison	M_{Δ}^{a}	SD_{Δ}^{b}	$M_{\%\Delta}^{c}$	$SD_{\%\Delta}^{d}$	% improved[e]
12–15 vs 16–17	1.80	2.16	29.10	35.54	18.42
12–15 vs 18–20	0.90	1.91	14.30	30.82	43.42
12–15 vs 28–30	-0.17	1.26	-2.07	20.37	60.53

[a] M_{Δ} is mean raw increase in RT (secs).
[b] SD_{Δ} is SD of raw increase in RT (secs).
[c] $M_{\%\Delta}$ is mean of percentage increase in RT.
[d] $SD_{\%\Delta}$ is SD of percentage increase in RT.
[e] "% improved" is percentage of the sample that improved after training.

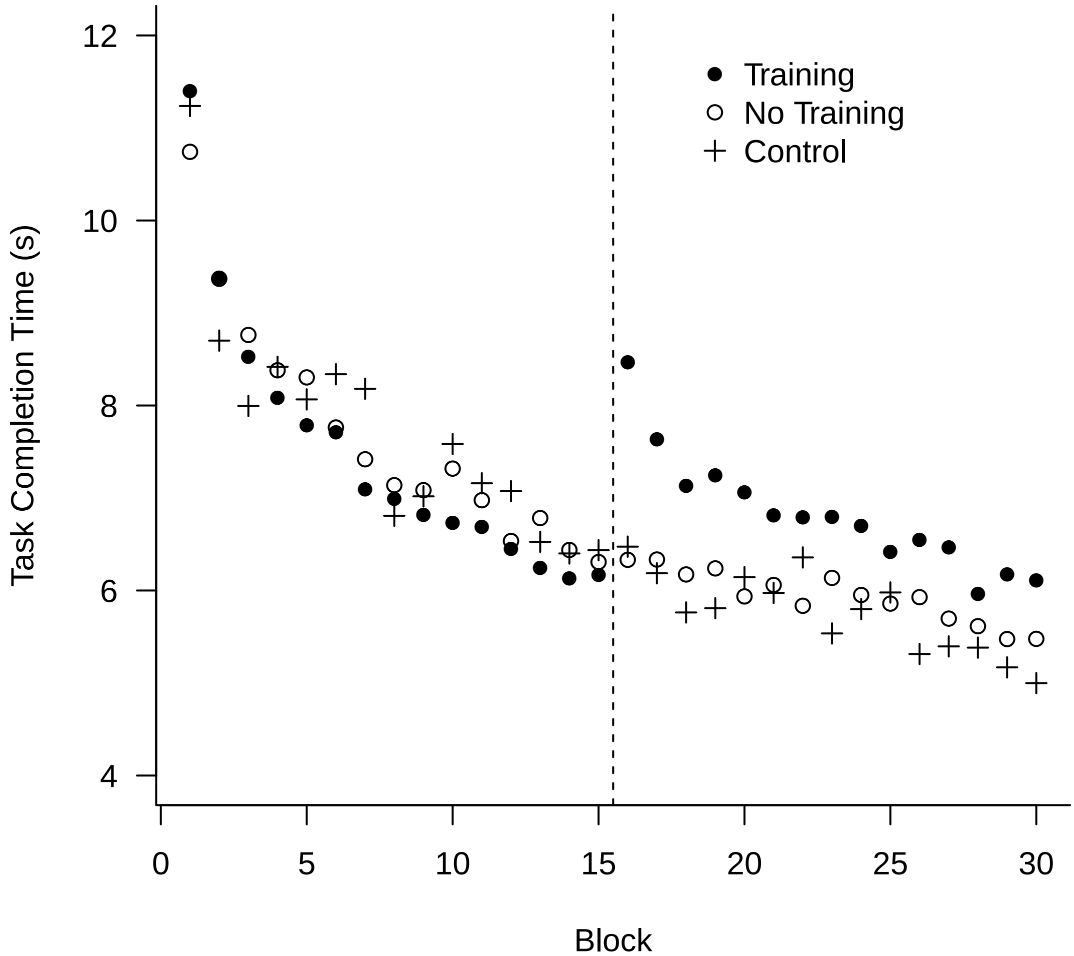

Figure 4.3: Task completion time by block and condition at the group-level. Dotted vertical line indicates the onset of mid-practice training.

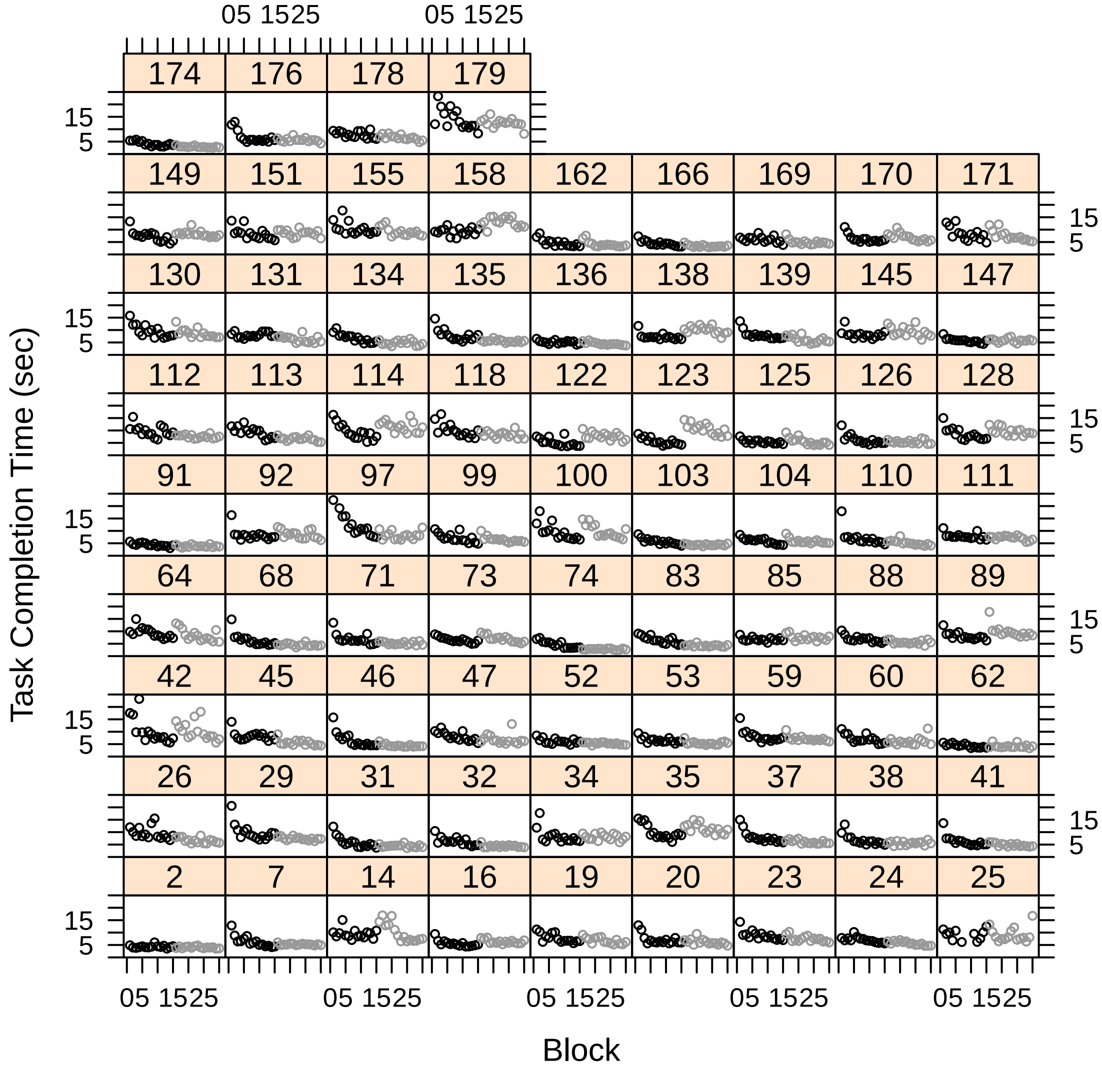

Figure 4.4: Task completion time by block at the individual-level. Blocks before mid-practice training are shown in black, and blocks after are shown in grey.

Figure 4.5 examines whether participants in the Training condition who changed their strategy following training also had the greatest change to their task completion time. This appears to be the case. The figure compares data from the four blocks before training with the four blocks after training. The smoothed line of fit shows that participants with minimal change in strategy use generally had minimal change to their task completion time. In contrast, participants who changed strategies typically were slower following the change. Interestingly, participants who reduced their strategy sophistication also appeared to decline in performance.

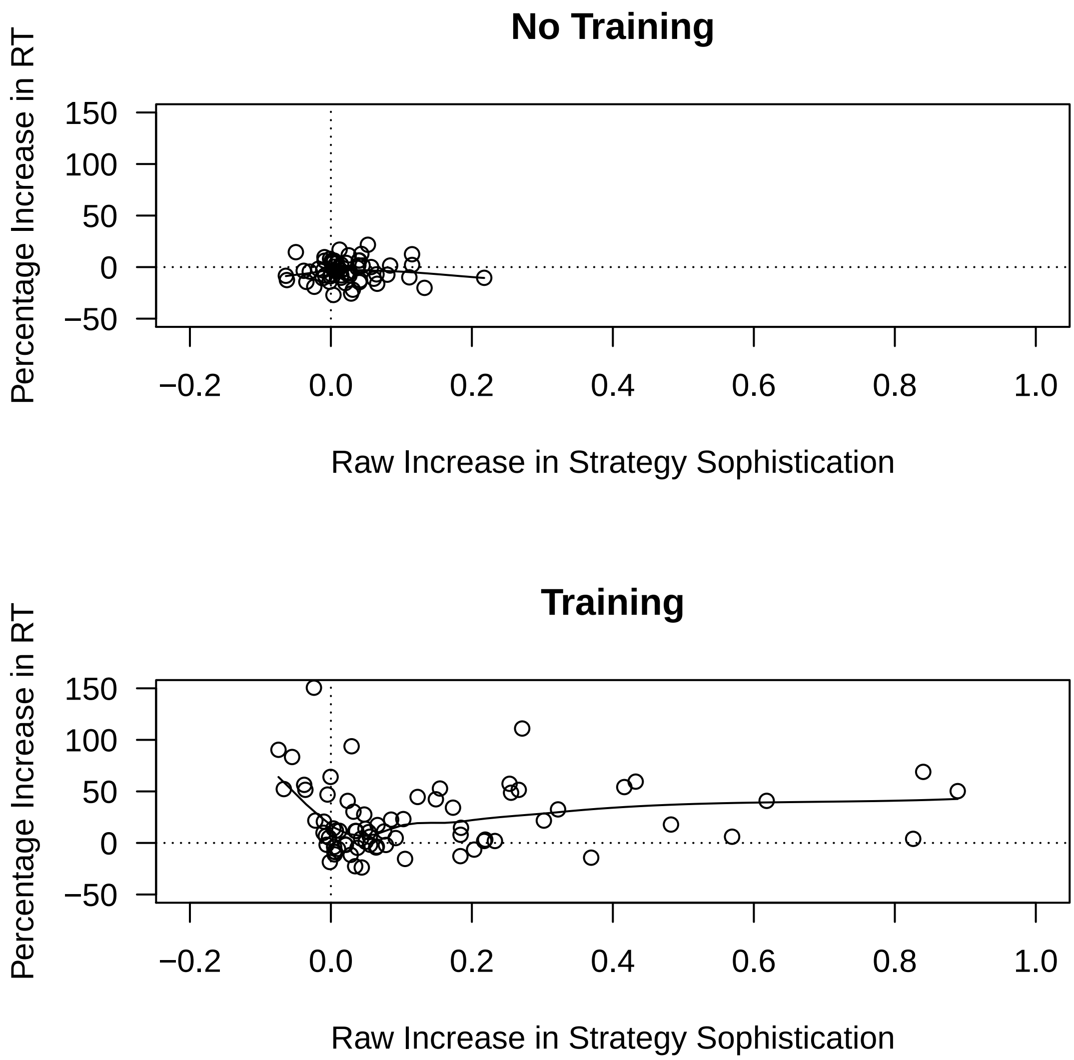

Figure 4.5: Scatter plot showing percentage increase in mean trial reaction time by increase in strategy sophistication from blocks immediately before training (i.e., 12–15) with those immediately after training (i.e., 16–18). Top graph shows No Training condition and bottom graph shows training condition. Loess smoother is overlayed. Dotted line indicates no change.

4.4 Discussion

4.4.1 Overview

This chapter aimed to improve understanding of the dynamics of strategy change and performance following instructions to shift strategy. Previous chapters suggested that self-initiated strategies were often gradual and typically led to no discernible discontinuities in performance. This chapter showed that instructed strategy shifts were more abrupt, led to initial performance decline, and often led to discernible and new learning curves.

4.4.2 Strategy

As expected the mid-practice strategy instruction led to an increase in strategy sophistication. Consistent with expectations, participants were able, at least partially, to learn from the training and follow the instructions.

The actual size of the strategy change at the group-level was modest. An examination of the individual-level suggested several explanations for this. First, a large number of participants were already using a sophisticated strategy at the time of the additional instruction. Initial instructions appear to have been fairly effective in encouraging use of sophisticated strategies. Individuals already using sophisticated strategies were thus largely unable to increase sophistication. Second, for participants using a simple strategy at the time of the mid-practice training the shift was not always complete. Participants may not have been able to absorb all the ideas about the sophisticated strategy at once.

At the individual-level, patterns of shift following mid-practice training were diverse. Of the shifts that did occur, most were abrupt. Several plausible reasons for this abruptness can be proposed. First, the mid-practice training included several opportunities to consolidate the increased strategy sophistication. Thus, strategy change may have happened gradually but over the period of the mid-practice training rather than on the measured performance trials. Second, participants were given explicit instruction to use the more sophisticated strategy. Third, in the Training condition prior to the mid-practice training, and in the other conditions at all times, participants were requested to optimise task completion time. Following the mid-practice training participants were asked to optimise strategy sophistication and then performance. This may have led participants who received the training to persist in using sophisticated strategies in spite of a drop in performance. Whether this persistence was due to perceived long term benefits of text editing strategy sophistication or to general compliance with experimental instructions is difficult to determine. Finally, most participants were relatively stable in their strategy use once mid-practice training occurred, which means that the effect of instruction was more discernible.

In one case, an individual abruptly shifted to the sophisticated strategy and then reverted to the simpler strategy, possibly because of performance difficulties. Examination of the participant's performance data revealed a rather large drop in performance while the participant used the more sophisticated strategy. Arguably, such a situation mirrors a common occurrence across training situations where an individual lacks the prerequisites to implement a strategy effectively, or where an individual perceives the time to attain competence in the new strategy as too long or requiring too much effort. In the language of local minima, the participant is unwilling to traverse the distance to an ultimately superior minimum. Alternatively, the particular individual may not perceive the required investment to improve performance in the longer term to be rational.

Many other individuals shifted incompletely with some individuals settling on a strategy that was only partially sophisticated. This may be due to limits in amount of information individuals could absorb. In a more iterative environment of feedback and training typical of personalised instruction or deep self-reflection, such plateaus of strategy sophistication could be removed.

The timing of the mid-practice training also limited the strategy shifts that could be observed. Many participants were already using relatively sophisticated strategies at the onset of the mid-practice training.

It is interesting to note the changes in the control group before and after the mid-practice break. The change did not occur immediately following the break. The increase that did eventually occur was not quite statistically significantly different to the other conditions. Several explanations are possible. One explanation is that it was just random variation. Another is that the disruption encouraged a new orientation to the task and eventual increase in strategy sophistication. Either way, the absense of an immediate shift after the break and the fact that the sample observed shift was smaller than that for the Training condition suggests that the shift observed in the Training condition was a result of the Training and instruction it contained and not merely the break from task performance. Theory would suggest that the increase in strategy sophistication in the training condition was the result of the explicit entry into the repertoire and explicit request to use sophisticated strategies incorporated into the Training Condition.

Overall the results reiterated the biasing effect of group-level analysis. The group-level results combined participants who had abrupt increases in their strategy sophistication with those that changed very little. At the group-level, it appeared that explicit instruction resulted in only a modest increase in strategy sophistication. However, at the individual-level, many participants increased abruptly and substantially in their strategy sophistication. After these participants were combined with others who were already operating at a high level of sophistication, the discernible effect was greatly reduced.

4.4.3 Performance

As hypothesised, performance declined following strategy instructions. The Control group did not appear to decline in performance following a break from task practice. Thus, it seems likely that the instruction induced a change in strategy which disrupted established strategies, and slowed task completion time.

As expected, performance at the group-level and often at the individual-level exhibited characteristics of a new monotonically decelerating learning curve. The results are also consistent with the idea that new strategies are similar to a new task. They can create a new learning curve. It may be that the processes in naturalistic studies minimise this effect by strategy selection processes operating when the strategy is relatively adaptive in the short term.

In contrast to expectations, performance did not surpass either performance prior to the introduction of the strategy or performance relative to the No Training group. However, by the end of practice, performance was relatively similar, yet strategy sophistication remained substantially higher.

These findings have several implications for applied work. Externally introduced strategy shifts typically require longer to learn. It reinforces general management principles that encourage the avoidance of micro-management. It highlights how the expert's view of the relative effectiveness of a sophisticated strategy differs from that of the novice. Effective trainers need to be wary of imposing their preferred strategy on to novices. The expert is in a position to appreciate the long term benefits of investing in the superior strategy, but may need to apply effort to be mindful of the extended period the novice will experience where performance declines. Such strategy adjustment is also likely to result in a drop in automaticity, which may reduce the potential for the novice to focus on broader task goals.

Another reason for the drop in performance may be related to the individuals that shifted. In particular, many participants had already shifted to the more sophisticated strategies. Participants who had already shifted are likely to have either felt intrinsically ready to make the shift, or may have had prior experience relevant to using the same or similar strategies. Thus, the participants who were continuing to use simple strategies at block 15 were less competent in general at text editing, and particularly less ready for the more sophisticated text editing keys.

Along these lines, it would be interesting for future research to repeat the present study with different levels of initial strategy training. With less initial training, strategy sophistication should be lower at the point of the mid-practice training. As such, it may be that the pool of potential shifters that remain would include more participants ready for strategy shift. By comparing low and high initial training conditions following mid-practice training, the relative effect of readiness for shift and the mere fact of introducing a strategy shift could be compared in terms of adoption of the instructions and disruptions to performance.

The findings support the idea that individuals are adaptive in selecting strategies at least given the short term levels of use operating in the present study. In the case of text editing, more sophisticated strategies reduce key presses but also often require more complex coordination or more complex mental models of text editing. Learning to coordinate new movements and acquire new mental models takes time, arguably longer than the duration of practice in the present study.

4.4.4 Comparison with Previous Research

Yechiam et al

The study by Yechiam et al. (2004) suggested that individuals were reluctant to switch to a more sophisticated strategy if they had already invested time in learning a reasonably effective simpler strategy. In the present study many participants shifted to a more sophisticated strategy following instruction and training to do so. There are several explanations of why the results of Yechiam et al. (2004) differ from the present study. First, the duration of the present study was of a defined duration, whereas Yechiam et al's study was ostensibly defined in terms of completing a defined number of trials, regardless of how long it took to complete. Thus, assuming participants wanted to finish the study quickly, in Yechiam et al's study participants had an incentive to avoid strategies with unknown pay-offs whereas in the present study whether participants were fast or slow, the period of practice was of constant duration.

Second, a similar point regarding incentives could be made about the real world relevance of the strategies used. In Yechiam et al's study the script based strategy was invented for the experimental task. It was not a scripting language that participants might use in the future to improve their productivity. In contrast, the sophisticated strategies in the present study had the potential to improve the productivity of participants outside the experiment. This fact was highlighted to participants in the initial instructions. Real-world relevance should increase the incentive to invest in sophisticated strategies.

Third, the immediate cost of switching to the sophisticated strategy differed between the studies. In Yechiam et al's study, the sophisticated strategy involved writing a computer script. Participants needed to read a manual and experiment, with no certainty that they would ever get the script to work. Thus, the real strategic choice was whether to invest time in trying to use the sophisticated strategy. The actual execution of the script once understood was very simple. This is a common model for script based strategies, which are hard to learn, but easy to implement. This contrasts with the sophisticated text editing strategy. Presumably, most participants were able to press the sophisticated text editing keys. However, even once participants were aware of the sophisticated keys, there was still substantial learning required in order to automate the psychomotor routines associated with the key combinations and refine the production rules associated with using the keys in a given context. Performance losses associated with a lack of skill would typically result in reduced speed, but the task goal could still be achieved. This made it possible to dabble with the sophisticated text editing keys and consider their relevance.

Fourth, the training differed between the two studies. In Yechiam et al's study participants were merely given a manual which they could choose to study. In the present study participants were required to work through a set of interactive instructions and exercises. They were also explicitly told at the completion of training that they should try to continue to use the more sophisticated strategies. Most likely, richer training and stronger instructions can override the tendency of people to persist in using simpler strategies.

Finally, the studies differed in the potential for transfer across simple and sophisticated strategies. In Yechiam et al's study the simple mouse based strategy had no overlapping elements with the script based strategy. In contrast, sophisticated text editing share many elements with simpler text editing. This degree of transfer should ease the transition between the two strategies.

In summary, the above discussion interpreted the findings of the present study in terms of incentives, information, and task characteristics. Incentives can flow from an instructor, from the task itself, or from the potential for future benefits. Incentives reflect the degree to which short term performance declines are required for longer term benefits. Incentives also reflect whether longer term benefits even exist.

The central conclusion of Yechiam et al. (2004) was that instructors need to be careful when introducing simple strategies first in training. The argument was that individuals may be reluctant to learn more complicated strategies once they have invested and grown comfortable with simpler strategies. While this argument seems reasonable, the present study suggests a few qualifications. First, external instruction and training can override the tendency to use simpler strategies. It is the role of training to provide reasons, and a pathway, to more sophisticated strategies. Second, for some individuals, who do not perform the task often, it may be rational for them to maintain the simpler framework of thinking about the task. This removes the need for additional learning, a temporary drop in performance, and the increased complexity and cognitive resource requirements that may be initially required to learn the more sophisticated strategies. Third, task interfaces can often be used to facilitate the conversion from simpler to more sophisticated strategies.

Delaney et al

Results of the present study were consistent with Delaney et al. (1998) who found that external introduction of a new strategy can lead to a temporary reduction in performance. The new strategy introduces new task elements to learn and introduces a new learning curve.

The new learning curves are even more apparent when the strategy is externally introduced. Self-initiated strategy shifts on the text editing task appear to be more adaptive. Participants appear less willing to accept temporary drops in performance for potential future gains. Participants who initiated a strategy shift on their own seem to be more ready to benefit from these more sophisticated strategies.

4.4.5 Conclusion

The results have interesting implications for the decision of whether to instruct people to adopt a more sophisticated strategy when they are already comfortable with an existing simpler strategy. A temporary drop in performance is often to be expected. Whether the acceptance of such a temporary drop in performance is rational would depend on the details of the duration of the drop, the ultimate superiority of the more sophisticated strategy, and how often the strategy would be used in the future. It also helps to explain why it is sometimes better to teach more sophisticated strategies from the start.

Overall, the results show that instructed strategy shifts yield discontinuities in performance. However, such discontinuities, at least in the case of text editing, are typically the opposite to those proposed by Haider and Frensch (2002). Rather than improvement in performance, strategy shift when externally introduced can result in a short term drop in performance. In the literature there have been occasions where ultimately superior strategies are seen as being synonymous with performance improvement (e.g., Crossman, 1959). The present study has shown empirically that process and performance are distinct, and that sometimes the benefits of an ultimately superior strategy take a period of extended practice to be realised.

It may be that casual observations help to explain the ongoing belief in discontinuities following strategy shift. First, if an individual who is skilled in both a simple and a sophisticated strategy chooses to compare his or her own performance with the simple and sophisticated strategy, the individual will notice that performance abruptly improves when he or she uses the sophisticated strategy. Second, if an individual is skilled in the simple strategy but not the sophisticated strategy and is told to adopt the sophisticated strategy when they are not ready, they will notice an abrupt drop in performance. However, the strategy shift that typically occurs as part of the learning curve is different to both of the above scenarios. Typically, when an individual shifts strategy as part of the learning process, shifts are likely to occur when they seem appropriate. While different dynamics are possible, it seems that self-initiated strategies are more likely to occur when there are benefits in the short term. Thus, empirically observed strategy shifts are self-initiated and often lead to minimal observed changes to performance. The learner is both ready to shift, but has not yet realised the full benefits of the shift. In combination, these different conditions may help to explain the persistence of ideas of discontinuous effects on performance following strategy shift.

Chapter 5

General Discussion

Contents

5.1 Overview

This thesis aimed to understand several fundamental aspects of the relationship between practice, strategy use, and performance. The first aim was concerned with determining the functional form of the relationships between practice and strategy use, and practice and performance. The second aim examined the role of individual differences in explaining variation in strategy sophistication and performance. The third aim assessed differences between self-initiated and instructed strategy shifts.

To achieve these aims, three studies were conducted. Each study measured individual differences and strategy use and performance on a text editing task. The methodology and analytic approach combined individual-level analysis, consideration of multiple continuous and discontinuous models, and trial-level measurement of strategy sophistication and task completion time. Results confirmed and extended previous research on the relationship between practice, strategy use, and performance. Specifically, several hypotheses were proposed and tested. Table 5.1 sets out these hypotheses indicating over the three studies whether each one was supported. In summary, most hypotheses were supported.

Table 5.1: Summary of Support for Hypotheses

Number	Hypothesis	Study 1	Study 2	Study 3
2.1	The relationship between practice and task completion time at the group-level is better explained by a three parameter power function than a three parameter exponential function.	Supported	Supported	Supported
2.2	The relationship between practice and task completion time at the individual-level is better explained by a three parameter exponential function than a three parameter power function.	Supported	Supported	Non-significant trend in predicted direction
2.3	Discontinuities do not occur in the relationship between practice and task completion time at the group-level.	Supported	Supported	Supported
2.4	Discontinuities occasionally occur in the relationship between practice and task completion time at the individual-level.	Supported	Supported	Supported
2.5	For individuals that shift strategy, some shifts are gradual and others are abrupt.	Supported	Supported	Supported
2.6	Many individuals will be characterised by only partial strategy sophistication.	Supported	Supported	Supported
2.7	The probability of a strategy shift occurring decreases monotonically with practice.	Supported	Supported	Supported

Table continues on next page

Number	Hypothesis	Study 1	Study 2	Study 3
2.8	For some individuals, the relationship between practice and strategy sophistication involves near continuous use of a simple strategy.	Supported	Supported	Supported
2.9	For some individuals, the relationship between practice and strategy sophistication involves near continuous use of a sophisticated strategy.	Supported	Supported	Supported
2.10	The relationship between practice and strategy sophistication at the group-level is effectively modelled by a three parameter Michaelis-Menten function.	Supported	Supported	Supported
3.1	Typing speed is strongly positively correlated with text editing performance.	Supported	Supported	Supported
3.2	Prior experience is strongly positively correlated with text editing performance.	Supported	Supported	Supported
3.3	Ability is moderately positively correlated with text editing task performance.	Supported	Supported	Supported
3.4	Big 5 personality factors are unrelated to text editing task performance.	Supported	Supported	Supported
3.5	Strategy use partially mediates the relationship between ability and prior experience on task performance.	Rejected	Rejected	Rejected
4.1	The combination of a request to use a strategy and information on how to use a strategy increases strategy use.	Not Tested	Not Tested	Partially Supported

Table continues on next page

Number	Hypothesis	Study 1	Study 2	Study 3
4.2	When sophisticated strategies are externally introduced later in practice, performance declines immediately after introduction.	Not Tested	Not Tested	Supported
4.3	When sophisticated strategies are externally introduced later in practice, performance eventually surpasses the level that would have been attained had instruction not been received.	Not Tested	Not Tested	Not Supported

The three previous chapters have discussed specific findings. It is not the purpose of this chapter to repeat these. The remainder of this general discussion chapter focuses on general themes, limitations, and concluding comments. Suggestions for future research are also interspersed.

5.2 General Themes

5.2.1 Biasing Effect of Group-Level Analyses

Throughout this thesis it has been argued that it is the individual-level that is relevant to understanding psychological change. While group-level data are easier to model due to both the need to only model one dataset and the reduced error variance, group-level models are inappropriate for drawing inferences regarding the individual-level. The biasing effect of group-level analysis was demonstrated on multiple occasions in this thesis. Given the frequent use of group-level analyses, each of these findings challenges important existing ideas in the research literature.

First, the exponential function generally provided superior fit in comparison to the power function at the individual-level, yet the reverse was true at the group-level. This finding goes against Newell and Rosenbloom's (1981) proposed Power Law of Practice. It supports the findings of Heathcote et al. (2000), extending them by studying not only cognitive tasks but also a task with cognitive and perceptual-motor components.

Second, the group-level smoothed over the few discontinuities that did occur at the individual-level. The infrequency of discontinuities at the individual-level goes against the suggestions by Haider and Frensch (2002) that discontinuities were under-identified largely due to group-level analyses. However, it did take individual-level analyses and active modelling of discontinuities in order to actually assess the presence of discontinuities at the individual-level.

Third, as with task performance, similar biasing effects were observed when modelling the relationship between practice and strategy sophistication at the group-level. Strategy sophistication often increased fairly abruptly at the individual level while at the group-level it followed a continuous, increasing, decelerating pattern.

Taken together, these findings have a number of important implications. Theories, measurement, and modelling of the relationships between practice, strategy use, and performance need to proceed at the individual-level if psychologically meaningful conclusions are to be drawn. In the present thesis a reconciliation of discontinuous strategy shifts and continuous learning functions was provided that was only possible due to the individual-level perspective adopted.

5.2.2 Trial-Level Strategy Measurement

Another theme in this thesis was the importance of trial-level measurement of strategy use at the individual-level. The use of text editing enabled recording of key presses which facilitated extraction of strategy measurement. This enabled actual measurement and modelling of the relationship between practice and strategy use.

This contrasts with many previous studies which did not or were not able to measure strategy for each individual on each trial. For example, Haider and Frensch (2002) inferred strategy use from average performance in blocks and behaviour on a transfer task. While not all tasks have strategies that manifest in observable behaviour, tasks that do have observable strategies represent a good starting point for researching the relationship between practice and strategy use.

5.2.3 Individual Differences

A third general theme was that of individual differences. On the criterion side, the thesis provided a more nuanced explanation of individual differences in performance. On the predictor side, models that emphasise prior experience, ability, and skill were supported.

Importantly, no one function can be used to describe relationships between practice, strategy use, and performance. In addition to the common finding that parameters of functions vary between individuals, the actual functional form differed between individuals. This variation was even more prominent in the relationship between practice and strategy sophistication than it was between practice and performance. These two levels of randomness—between individuals and over time—are essential to developing models of skill acquisition.

5.2.4 Continuity and Discontinuity

The theme of continuity and discontinuity has flowed throughout this thesis. Analyses highlighted the importance of adopting a formal modelling procedure that assesses the degree of fit of discontinuous models. Such analyses were also predicated on the previously mentioned individual-level of measurement.

Measurement, modelling, and individual-level focus have not co-occurred in previous research when modelling the learning curves. There are several possible explanations for this. First, many learning curves appear to be relatively well-explained by continuous models. Second, the literature on discontinuity analysis is relatively disconnected from psychology largely existing in statistics and econometrics. Third, many tasks are unlikely to have discontinuities. A contribution of this thesis has been to highlight how individual-level discontinuities can be modelled.

Instructed strategy shift, not surprisingly, resulted in discontinuities to both strategy sophistication and performance in ways distinct from self-initiated strategy shift. External interventions represent a discontinuity in the environment, which helps to explain the discontinuous effect on the individual. In contrast, without external intervention, the environment tends to be fairly constant in a typical skill acquisition study. This typically leads to big shifts occurring early and with reduced frequency over time. In addition, natural processes of strategy selection appear to interact with top-down processes leading to minimal immediate performance effects of self-initiated strategy shifts. Similarly self-initiated strategy changes are more often gradual, possibly caused by both local reinforcement processes and the potentially implicit method of strategy entry into the repertoire.

5.3 Limitations

5.3.1 Overview

This section focuses on some of the limitations of the empirical research presented in this thesis. For the most part the limitations focus on the degree to which the findings can be generalised to other samples, tasks, and contexts.

5.3.2 Sample

First, participants in the present research were mostly young adult university students. This sample can be contrasted with children (e.g., Siegler, 2006) and older adults (e.g., Hertzog, Touron, & Hines, 2007; Rogers, Fisk, & Hertzog, 1994; Rogers et al., 2000), two groups that have both received substantial research interest. Also, while the sample varied in prior experience on the text editing task, most participants had a reasonable level of prior experience.

Given the selective nature of the sample, it is unclear how well the findings would generalise to children, older adults, and those with less prior experience. One possibility is that the findings would largely be the same except that performance would be poorer, use of sophisticated strategy would be less, and achievement of minimal task requirements would occur less often. Alternatively, quite different dynamics might emerge. In particular, the inability to use a keyboard or having no prior experience with text editing would create a wide range of new challenges which this thesis has not aimed to address. It would be interesting if future research were to adopt the same design and modelling paradigms as used in this thesis and applied it to these other populations.

5.3.3 Task

The second limitation relates to the consistent use of keyboard based text editing task in all three studies. The task was chosen to enable generalisation of findings to other repetitive tasks that contain cognitive, perceptual, and motor components along with the potential for a simple–sophisticated strategy shift, such as air-traffic control tasks, and many other computerised tasks. Nonetheless, future research could examine the extent to which present findings generalise to other tasks. In particular, it may be that the complexity of real-world tasks reduces further the potential for discontinuities to emerge.

Future research could also take a broader perspective on strategy use. While the present thesis operationalised strategy use based on key logs, future research could examine several other strategic elements related to the text editing. First, algorithm-retrieval shift could be examined. This could include memorising editing changes, memorising aspects of the edited text, and memorising shortcut keys. Second, eye gaze could be studied, which could include the efficient allocation of visual attention to aspects of the editing screen (for an example, see F. J. Lee & Anderson, 2001). Finally, the learning orientation of participants could be examined. For example, a learner may decide strategically to learn and apply advanced shortcut keys even if it means a temporary decline in performance. Another learner may not be willing to accept temporary declines in performance.

5.3.4 Context

The studies presented in this thesis all involved a monitored laboratory setting with performance feedback. While this is common for psychological studies of learning curves, it does raise issues regarding generalisation. It may be that learning processes are simply sped up in an experimental skill acquisition study. However, in real-world settings discretionary effort, breaks in practice, and the unstructured nature of instructions are likely to change the dynamics of learning. The main reason for the choice of the lab based experimental method was to focus on the particular learning processes of interest and thereby remove consideration of issues such as discretionary effort, which are certainly important in the real-world. Nonetheless, caution should be exercised in generalising findings to work or other real-world settings. It may be, as Charman and Howes (2003) suggested, that participants in lab settings switch to better strategies more quickly because they are less concerned with higher-order goals. Likewise, the lack of prediction by personality variables may be related to the minimal capacity to use discretionary effort.

A second aspect of the setting is the duration and concentration of practice. For the studies in this thesis, total practice time was short relative to some real-world contexts. Also, in real-world settings, practice is often less concentrated. Factors of motivation, fatigue, and forgetting are likely to be more important. In particular in the real-world, learning tasks such as typing or using complex software can occur over months and years. It is likely that on these more complex tasks that discontinuities are likely to occur with much less frequency.

5.4 Conclusion

The process by which humans get faster at completing a task with practice has been and remains an exciting area of research. This process is part of the amazing capacity of humans to adapt to their environment. Describing the functional form of the relationship between practice and performance, and articulating the processes that underlie both task performance and learning, remains an active interest to researchers in psychology. This thesis aimed to contribute to this understanding in several ways, by taking an individual-level perspective, by testing discontinuous models, by examining individual differences, and by comparing self-initiated and instructed strategy shifts. A means of reconciling discontinuous models of strategy shift with continuous models of the learning curve was proposed. These theoretical, modelling, measurement, and design ideas will hopefully be useful for future researchers looking at new tasks, samples, and contexts.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*, 3–27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*, 288–318.

Ackerman, P. L. (1989). Within-task intercorrelations of skilled performance: Implications for predicting individual differences? (a comment on Henry & Hulin, 1987). *Journal of Applied Psychology*, *74 (2)*, 360–364.

Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 883–901.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, *77*, 598–614.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitve and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131(4)*, 567–589.

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied*, *1(4)*, 270–304.

Ackerman, P. L., & Woltz, D. J. (1994). Determinants of learning and performance in an associative memory/substitution task: Task constraints, individual differences, volition, and motivation. *Journal of Educational Psychology*, *86(4)*, 487–515.

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention and transfer of human motor skills. *Psychological Bulletin*, *101*, 41–74.

Allison, R. B. (1960). *Learning parameters and human abilities (Office of Naval Research Techinical Report).* Princeton, NJ: Educational Testing Service.

Andersen, T., Carstensen, J., Hernndez-Garca, E., & Duarte, C. M. (2009). Ecological thresholds and regime shifts: Approaches to identification. *Trends in Ecology & Evolution*, *24*(1), 49–57.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–406.

Anderson, J. R. (1987). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* NJ: Lawrence Erlbaum.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (2000). *Cognitive psychology and its implications (5th ed.).* New York: Worth Publishers.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111(4)*, 1036–1060.

Anderson, J. R., Matessa, M., & Legiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human Computer Interaction*, *12*, 439–462.

Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, *61*, 821–856.

Anglim, J. (2000a). Simple, 2-Choice, 4-Choice Reaction Time Test [Computer software manual].

Anglim, J. (2000b). Text editing task for jeromy anglim's honours thesis, supervised by janice langan-fox (1.0 ed.) [Computer software manual].

Anglim, J., & Waters, L. (2007). Practical tips on how to conduct a sophisticated online psychological experiment. In *43rd APS Annual Conference, 23-27 September 2008.* Hobart, Tasmania, Australia..

Armstrong, K. (2000). *Cognitive appraisals, metacognition and emotion during skill acquisition.* Unpublished doctoral dissertation, University of Melbourne, School of Behavioural Science.

Bahrick, H. P., Fitts, P. M., & Briggs, G. E. (1957). Learning curves: Facts or artifacts? *Psychological Bulletin*, *54*(3), 256–268.

Barrick, M., & Mount, M. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26.

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications.* NY: John Wiley & Sons, Inc.

Beem, A. L. (1995). A program for fitting two-phase segmented-curve models with an unknown change point, with an application to the analysis of strategy shifts in a cognitive task. *Behavior Research Methods, Instruments and Computers*, *27*(3), 392–399.

Beier, M. E., & Ackerman, P. L. (2005). Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment. *Psychology and Aging*, *20*, 341–355.

Bhavnani, S. K., & John, B. E. (2000). The strategic use of complex computer systems. *Human-Computer Interaction*, *15*(2), 107–137.

Blessing, S. B., & Anderson, J. R. (1996). How people learn to skip steps. *Journal of Experimental Psychology-Learning Memory and Cognition*, *22*(3), 576–597.

Bolker, B. M. (2008). *Ecological models and data in R.* New Jersey: Princeton Univ Press.

Brown, R., Durbin, J., & Evans, J. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, *37*(2), 149–192.

Bryan, W. L., & Harter, N. (1897). *Studies in the physiology and psychology of the telegraphic language.* New York: Macmillan.

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

Burke, M. (1984). Validity generalization: A review and critique of the correlation model. *Personnel Psychology*, *37*(1), 93–115.

Card, S. K., Moran, T. P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, *23*(7), 396–410.

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

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Carroll, J. M., & Rosson, M. B. (1987). *Paradox of the active user.* Cambridge, MA, USA: MIT Press.

Carroll, L. (1865). *Alice's adventures in wonderland.* England: Macmillan.

Cary, M., & Reder, L. M. (2002). Metacognition in strategy selection: Giving consciousness too much credit. In M. Izaute, P. Chambres, & P. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 63–78). New York: Kluwer.

Charman, S. C., & Howes, A. (2003). The adaptive user: An investigation into the cognitive and task constraints on the generation of new methods. *Journal of Experimental Psychology: Applied*, *9*(4), 236–248.

Chu, C., Stinchcombe, M., & White, H. (1996). Monitoring structural change. *Econometrica: Journal of the Econometric Society*, *64*(5), 1045–1065.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155–159.

Cook, R., Kay, J., Ryan, G., & Thomas, R. (1995). A toolkit for appraising the long-term usability of a text editor. *Software Quality Journal*, *4*(2), 131–154.

Crossman, E. R. F. (1959). A theory of the acquisition of speed skill. *Ergonomics*, *2*, 153–166.

Crowley, K., Shrager, J., & Siegler, R. (1997). Strategy discovery as a competitive negotiation between metacognitive and associative mechanisms. *Developmental Review*, *17*, 462–489.

Crowley, K., & Siegler, R. (1999). Explanation and generalization in young children's strategy learning. *Child Development*, *70*, 304–316.

Cudeck, R., & Harring, J. R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology*, *58*, 615–637.

De Jong, J. R. (1957). The effects of increasing skill on cycle time and its consequences for time standards. *Ergonomics*, *1*(1), 51–60.

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science*, *9(1)*, 1–7.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*, 735–808.

Digman, J. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41* (1), 417–440.

Donchin, E. (1995). Video games as research tools: The space fortress game. *Behavior Research Methods Instruments and Computers*, *27*, 217–217.

Ecker, U., Lewandowsky, S., Oberauer, K., Chee, A., & Ecker, U. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *36* (1), 170–189.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Services.

Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making*, *21* (5), 575–597.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*, 725–747.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Review*, *53*, 134–140.

Farina, A., & Wheaton, G. (1971). *Development of a taxonomy of human performance: The task characteristics approach to performance prediction.* Defense Technical Information Center.

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

Fleishman, E. A. (1972). On the relation between abilities, learning, memory and human performance. *American Psychologist*, *27*, 1017–1032.

Fleishman, E. A., & Fruchter, B. (1960). Factor structure and predictability of successive stages of learning morse code. *Journal of*

Applied Psychology, *44*(2), 97–101.

Frensch, P., Lindenberger, U., & Kray, J. (1999). Imposing structure on an unstructured environment: Ontogenetic changes in the ability to form rules of behavior under conditions of low environmental predictability. In A. D. Friederici & R. Menzel (Eds.), *Learning: Rule extraction and representation* (pp. 139–164). Berlin: Walter De Gruyter Inc.

Fry, E. B., Kress, J. E., & Fountoukidis, D. L. (1993). *Reading teachers book of lists* (3rd ed.). Center for Applied Research in Education. Available from `http://www.duboislc.org/EducationWatch/First100Words.html`

Gilbreth, F. B. (1911). *Motion study: A method for increasing the efficiency of the workman.* New York: Van Nostrand.

Goldberg, L. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. D. Fruyt, & F. Ostendorf (Eds.), *Personality psychology in europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied*, *6*(4), 322–335.

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

Gray, W. D., Sims, C. R., Fu, W. T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*, 461–482.

Haider, H., & Frensch, P. A. (1999). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied*, *5(2)*, 129–151.

Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmieri (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28(2)*, 392–406.

Hansen, B. (2001). The new econometrics of structural change: Dating breaks in US labor productivity. *Journal of Economic perspectives*, *15*(4), 117–128.

Harvey, L., & Rousseau, R. (1995). Development of text-editing skill: From semantic and syntactic mappings to procedures. *Human-Computer Interaction*, *10*(4), 345–400.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review*, *7(2)*, 185–207.

Hertzog, C., Cooper, B. P., & Fisk, A. D. (1996). Aging and individual differences in the development of skilled memory search performance. *Psychology and Aging*, *11*, 497–520.

Hertzog, C., Touron, D. R., & Hines, J. C. (2007). Does a time-monitoring deficit influence older adults' delayed retrieval shift during skill acquisition? *Psychlogy and Aging*, *22*(3), 607–624.

Hirsch, W. Z. (1952). Manufacturing progress functions. *Review of*

Economics and Statistics, *34*(2), 134–155.

Huet, S., Bouvier, A., Poursat, M. A., & Jolivet, E. (2004). *Statistical tools for nonlinear regression.* New York: Springer.

Hunter, J., & Schmidt, F. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, *124*(2), 262–274.

Ippel, M., & Beem, A. (1987). A theory of antagonistic strategies. *Learning and Instruction: European Research in an International Context*, *1*, 111–121.

John, B. E. (2003). Information processing and skilled behavior. In J. M. Carroll (Ed.), *HCI models, theories, and frameworks: Toward a multidisciplinary science* (pp. 55–101). San Francisco, LA: Morgan Kaufmann.

John, B. E., & Lallement, Y. (1997). Strategy use while learning to perform the Kanfer-Ackerman Air Traffic Controller Task. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society: August 7–10, 1997, Stanford University* (pp. 337–342). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*, 657–690.

Keil, C. T., & Cortina, J. M. (2001). Degradation of validity over time: A test and extension of Ackerman's model. *Psychological Bulletin*, *127*, 673–697.

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

Kelley, K. (2009). The average rate of change for continuous time models. *Behavior Research Methods*, *41*(2), 268–278.

Khodadadi, A., & Asgharian, M. (2008). Change-point problem and regression: An annotated bibliography. *COBRA Preprint Series*, 44.

Kolers, P. A. (1975). Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(6), 689–701.

Kolers, P. A., & Duchnicky, R. (1985). Discontinuity in cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 655–674.

Konz, S. A. (1987). *Work design: Industrial ergonomics.* Scottsdale, AZ: Publishing Horizons.

Kramer, A., Strayer, D., & Buckely, J. (1990). Development and transfer of automatic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 505–522.

Kramer, W., Ploberger, W., & Alt, R. (1988). Testing for structural change in dynamic models. *Econometrica: Journal of the Econometric Society*, *56*(6), 1355–1369.

Kyllonen, P., Lohman, D., & Woltz, D. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology*, *76*(6), 1325–1345.

Lacey, S. (2007). *Fundamental processes of skill acquisition for basic choice-reaction tasks.* Unpublished doctoral dissertation, University of Michigan.

Lane, N. E. (1987). *Skill acquisition rates and patterns: Issues and training implications.* New York: Springer-Verlag.

Lang, J. W. B., & Bliese, P. D. (2009). General mental ability and two types of adaptation to unforeseen change: applying discontinuous growth models to the task-change paradigm. *Journal of Applied Psychology*, *94*(2), 411–428.

Langan-Fox, J., Armstrong, K., Balvin, N., & Anglim, J. (2002). Process in skill acquisition: Motivation, interruptions, memory, affective states and metacognition. *Australian Psychologist*, *37(2)*, 104–117.

Lee, F. J., & Anderson, J. R. (2001). Does learning a complex task have to be complex?: A study in learning decomposition. *Cognitive Psychology*, *42(3)*, 267–316.

Lee, F. J., Anderson, J. R., & Matessa, M. P. (1995). Components of dynamic skill acquisition. In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 506–511). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin and Review*, *12*(4), 605–621.

Leisch, F. (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. In W. Härdle & B. Rönz (Eds.), *Compstat 2002 — proceedings in computational statistics* (pp. 575–580). Heidelberg, Germany: Physica Verlag.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*, 83–97.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology*, *18(5)*, 883–914.

Lovett, M. C. (2005). A strategy-based interpretation of stroop. *Cognitive Science*, *29*, 493–524.

Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology*, *31(2)*, 168–217.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1997). Modeling individual differences in a digit working memory task. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual conference* *of the Cognitive Science Society* (pp. 460–465). Mahwah, NJ.

Luchins, A. S. (1942). Mechanisms in problem solving: The effect of Einstellung. *Psychological Monographs*, *54*, 1–110.

Luwel, K., Beem, A. L., Onghena, P., & Verschaffel, L. (2001). Using segmented linear regression models with unknown change points to analyze strategy shifts in cognitive tasks. *Behavior Research Methods, Instruments, and Computers*, *33*, 470–478.

Luwel, K., Verschaffel, L., Onghena, P., & De Corte, E. (2003). Strategic aspects of numerosity judgment: The effect of task characteristics. *Experimental Psychology*, *51(1)*, 63–75.

MacKay, D. G. (1982). The problem flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, *89*, 483-506.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*, 107–127.

Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A re-examination of the leanring curve. *Psychological Bulletin*, *85*(6), 1256–1274.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of performance. *Journal of Applied Psychology*, *73*, 327–330.

McGeoch, J. A. (1927). The acquisition of skill. *Psychological Bulletin*, *24*(8), 437–466.

Michaelis, L., & Menten, M. (1913). Die kinetik der invertinwirkung. *Biochemische Zeitschrift*, *49*, 333-369.

Moore, M., & Dutton, P. (1978). Training needs analysis: Review and critique. *Academy of Management Review*, *3*, 532–545.

Morrison, R., & Brantner, T. (1992). What enhances or inhibits learning a new job? A basic career issue. *Journal of Applied Psychology*, *77*(6), 926–940.

Myung, I. J., Kim, C., & Pitt, M. A. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. *Memory and Cognition*, *28*(5), 832–840.

Neisser, U. (1963). Decision time without reaction time: Experiments in visual scanning. *American Journal of Psychology*, *76*(3), 376–385.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101.

Neves, C. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57–84). Hillsdale, NJ: Erlbaum.

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

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 324–354.

Palmeri, T. J. (1999). Theories of automaticity and the power law of practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25(2)*, 543–551.

Patrick, J. (1992). *Training: Research and practice.* Academic Press London.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552.

Payne, S. J., Squibb, H. R., & Howes, A. (1990). The nature of device models: The yoked state space hypothesis and some experiments with text editors. *Human-Computer Interaction*, *5*(4), 415–444.

Ploberger, W., & Kramer, W. (1992). The CUSUM test with OLS residuals. *Econometrica: Journal of the Econometric Society*, *60*(2), 271–285.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance*. Thousand Oaks, CA: Sage.

R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org` (ISBN 3-900051-07-0)

Rabbitt, P., Banerji, N., & Szymanski, A. (1989). Space Fortress as an IQ test? Predictions of learning and of practised performance in a complex interactive video-game. *Acta Psychologica*, *71*(1-3), 243–257.

Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, *52*, 501–525.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.

Reder, L. M. (1988). Strategic control of retrieval strategies. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 227–259). New York: Academic Press.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? familiarity with question terms, not with the answer. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(3), 435–451.

Reed, H. B., & Zinszer, H. A. (1943). The occurrence of plateaus in telegraphy. *Journal of Experimenta1 Psychology*, *33*, 130–135.

Revelle, W. (2010). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Available from `http://CRAN.R-project.org/package=psych` (R package version 1.0-88)

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, *126*, 288–310.

Rickard, T. C. (1999). A CMPL alternative account of practice effects in numerosity judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 532–542.

Rickard, T. C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30 (1)*, 65–82.

Ritter, F., & Schooler, L. (2002). The learning curve. In *International encyclopedia of the social and behavioral sciences* (pp. 8602–8605). Amsterdam: Pergamon.

Ritz, C., & Streibig, J. C. (2008). *Nonlinear regression with R.* New York: Springer.

Robertson, S. P., & Black, J. B. (1983). Planning units in text editing behavior. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 217–221). New York: ACM.

Robinson, K. M. (2001). The validity of verbal reports in children's subtraction. *Journal of Educational Psychology*, *93(1)*, 211–222.

Rogers, W. A., Fisk, A. D., & Hertzog, C. (1994). Do ability-performance relationships differentiate age and practice effects in visual search? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 710–738.

Rogers, W. A., Hertzog, C., & Fisk, A. D. (2000). An individual differences analysis of ability and strategy influences: Age-related differences in associative learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 359–394.

Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2000). Acquisition of intellectual and perceptual-motor skills. *Annual Review of*

Psychology, *52*, 453–470.

Rossett, A. (1987). *Training needs assessment*. New Jersey: Educational Technology Publications.

Sackett, P., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology*, *59*, 419-450.

Sackett, P., Schmitt, N., Ellingson, J., & Kabin, M. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, *56*(4), 302–318.

Sackett, P., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, *73*(3), 482–486.

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, *73*, 482–486.

Sarkar, D. (2010). lattice: Lattice graphics [Computer software manual]. Available from `http://CRAN.R-project.org/package=lattice` (R package version 0.18-8)

Schmidt, F., & Hunter, J. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*(5), 529–540.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological Review*, *84*, 127–190.

Schunn, C. D., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General*, *130(1)*, 59–76.

Schunn, C. D., Reder, L. M., & Nhouyvanisvong, A. (1997). To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *23*(1), 3–29.

Seber, G. A. F., & Wild, C. J. (2003). *Nonlinear regression* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Seibel, R. (1963). Discrimination reaction time for 1,023-alternative task. *Journal of Experimental Psychology*, *66*, 215–226.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*(3), 250–264.

Siegler, R. S. (1988a). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development*, *59*(4), 833–851.

Siegler, R. S. (1988b). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *17(3)*, 258–275.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S. (2006). Microgenetic analyses of learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology* (Vol. 2, pp. 464–510). Hoboken, NJ: Wiley.

Siegler, R. S., & Crowley, K. (1992). Microgenetic methods revisited. *American Psychologist*, *47*, 1241–1243.

Siegler, R. S., & Crowley, K. (1996). The microgenetic method. In L. Smith (Ed.), *Critical readings on Piaget* (pp. 606–620). London: Routledge.

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, *126*(1), 71–92.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive* *competence: New approaches to process modeling* (pp. 31–76). New York: Academic Press.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*, 377–397.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Singley, M., & Anderson, J. (1985). The transfer of text-editing skill. *International Journal of Man-Machine Studies*, *22*(4), 403–423.

Singley, M., & Anderson, J. (1987). A keystroke analysis of learning and transfer in text editing. *Human-Computer Interaction*, *3*(3), 223–274.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Snoddy, G. S. (1926). Learning and stability. *Journal of Applied Psychology*, *10*, 1–36.

Sobel, M. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 290–312.

Speelman, C. P., & Kirsner, K. (2005). *Beyond the learning curve: The construction of mind.* Oxford, UK: Oxford University Press.

Staw, B. (1981). The escalation of commitment to a failing course of action. *Academy of Management Review*, *6*, 577–587.

Taatgen, N. A. (2001). A model of individual differences in learning air traffic control. In E. M. Altmann, A. Cleeremans, C. D. Schunn, & W. D. Gray (Eds.), *Fourth International Conference on Cognitive Modeling* (pp. 211–216). Mahwah, NJ: Lawrence Erlbaum.

Tett, R., Jackson, D., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*(4), 703–742.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, *8*, 247–261.

Touron, D. R. (2002). Does what we know affect how we learn? Adult age differences in the effects of item knowledge and rule knowledge on skill development (Doctoral dissertation, Syracuse University). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *62(9-B)*, 4250.

Touron, D. R., & Hertzog, C. (2004). Strategy shift affordance and strategy choice in young and older adults. *Memory and Cognition*, *32*, 298–310.

Touron, D. R., Hoyer, W. J., & Cerella, J. (2004). Age-related differences in the component processes of cognitive skill learning. *Psychology and Aging*, *19(4)*, 565–580.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual review of psychology*, *47*(1), 513–539.

Verschaffel, L., De Corte, E., Lamote, C., & Dherdt, N. (1998). The acquisition and use of an adaptive strategy for estimating numerosity. *European Journal of Psychology of Education*, *13*(3), 347–370.

Voelkle, M. C., Wittmann, W. W., & Ackerman, P. L. (2006). Abilities and skill acquisition: A latent growth curve approach. *Learning and Individual Differences*, *16*, 303–319.

Walsh, M. M., & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive Psychology*, *58*, 416–440.

Yechiam, E., Erev, I., & Parush, A. (2004). Easy first steps and their implication to the use of a mouse-based and a script-based strategy. *Journal of Experimental Psychology: Applied*, *10(2)*, 89–96.

Yechiam, E., Erev, I., Yehene, V., & Gopher, D. (2003). Melioration and the transition from touch-typing training to everyday use. *Human Factors*, *45*(4), 671–684.

Zeileis, A., Kleiber, C., Krmer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics and Data Analysis*, *44*(1-2), 109–123.

Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). strucchange: An r package for testing for structural change in linear regression models. *Journal of Statistical Software*, *7*(2), 1-38. Available from `http://www.jstatsoft.org/v07/i02/`

Appendix A

Study 1: Additional Materials

A.1 Overview

This Appendix provides assorted additional material related to the introduction, method, and results of Study 1.

A.2 Materials

A.2.1 Study 1 Prior Experience Scale

Question text, response options, and assigned weight for each item on the scale are shown below. Weights were used in the calculation of prior experience. Weights were not displayed to participants.

```
Q1) How often do you use word processing programs,
    for example Microsoft Word?
  8 = Frequently (nearly ever day);
  5 = Often (once or twice a week);
  2 = Occasionally (once or twice a month);
  0 = Rarely or Never

Q2) Would you say you are experienced
```

```
    and/or skilled at using a keyboard?
  5 = YES, very experienced and skilled;
  3 = YES, moderately experienced and skilled;
  0 = NO, not at all experienced or skilled

Q3) Do you feel eyestrain and have vision problems
    from looking at a computer screen?
  0 = Yes, usually after a few minutes;
  1 = Yes, but only after a long period of time;
  2 = NO, never

Q4) Do you ever have any recurring problems with your vision?
  0 = Yes, my vision is substantially impaired;
  3 = Yes, but my vision is corrected by visual aides;
  4 = NO, never

Q5) Do you have any physiological (physical) problems that
    impair your motor control (the movement) in your arms or wrists?
  5 = NO;
  2 = YES --- moderate impairment;
  0 = YES --- severe impairment
```

For each of the following items participants were asked to rate their level of competence and familiarity using the following scale: 3 = Highly competent; 2 = Somewhat competent; 1 = Familiar with but not experienced; 0 = Never heard of it.

```
Q6) Short cut keys
Q7) Cut, copy, paste
Q8) Formatting paragraphs (e.g., using tabs, indenting)
Q9) Outline view or master document
```

`Q10) Macros`

The following questions required participants to provide an answer to each question. Each item was scored 3 = correct and 0 = incorrect.

```
Q11) What is the Windows short cut key for copy?
Q12) What is the Windows short cut key for paste?
Q13) What is the Windows short cut key for undo?
Q14) Which of the following keys is used to highlight
     text with the keyboard?  (please circle)
     i) Alt  ii) Control (Ctrl)  iii) Shift
Q15) Where does the cursor go if you press the ‘home’ key?
Q16) Where does the cursor go if you press
     the ‘ctrl’ and ‘end’ keys?
Q17) What does holding ‘ctrl’ and ‘shift’
     and pressing the right arrow button, accomplish?
```

Answers were scored correct if the following answers were provided: Q11: `Ctrl+X`; Q12: `Ctrl+V`; Q13: `Ctrl+Z`; Q14: `Shift`; Q15: Start of line; Q16: End of document; Q17: one of (a) Select word to the right or (b) unselect word from left of selection.

A.3 Text Editing Task

A.3.1 Additional Screenshots of Computer Interface

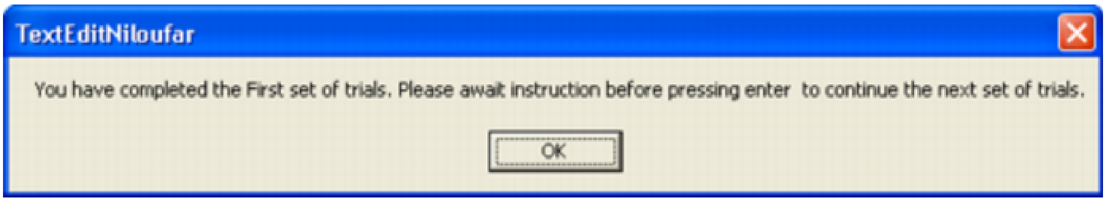

Figure A.1: Study 1 text editing task dialog box displayed at the end of a set of trials.

Figure A.2: Study 1 text editing task dialog box displayed at the end of all trials on the text editing task.

A.3.2 Text Editing Instructions

The following instructions were given to participants to read:

(Adapted from Armstrong (2000))

The task you are now going to complete is a text-editing task. This task is a relevant skill in many work settings. For example, text editing is used in administration and office settings. It is also becoming increasingly important in managerial roles where professional reports are commonly written with the aid of a personal computer. Publishing is also an area where this skill is vital. Text editing may also be a useful predictor of job performance in many white-collar jobs. Today you will be given a chance to practice this skill and will receive feedback about your performance on this task.

The task stimuli will consist of a computer text-editing task. This will entail making corrections in the text file from a marked manuscript provided.

You will sit in front of a computer terminal with a keyboard for input and a video display terminal for output. You are to make each of the marked modifications in the text file to produce an updated file. Do not change anything else in the document and complete the corrections as quickly and accurately as possible. Do not spend much time checking what you have completed.

You will be provided with feedback on your speed and accuracy after each 1 minute trial. The first trial will be for practice and then you will have 54 trials to complete divided into 18 trials of approximately 20 minutes, with breaks in between. When you are ready to start the trial press F2 and when you have completed each trial press F3 to move on to the next trial. Once you have reached the end of the 18 trials please take a 5 minute break.

In order to make the changes to the file on the computer you will need to use several different text-editing keys. An outline of all possible text-editing functions is included on the following page.

PLEASE DO NOT USE THE MOUSE!

Once you have read the instructions and understood what is required from you for this task, you will be directed to open the file on the computer named 'Text Editing' and then upon instruction you can commence the task.

If you have any questions please do not hesitate to ask the researcher

A.3.3 Accuracy Calculation

Accuracy was measured as a score out of 12. One accuracy point was awarded for the presence of each of the following strings.

(i) "We know"

(ii) ", no languages or"

(iii) "languages or cultures"

(iv) "Self-"

(v) "-knowledge"

(vi) "199"

(vii) "1994"

(viii) "and selfhood"

(ix) "selfhood."

(x) "of identity"

(xi) "of identity and"

(xii) "of identity and selfhood"

A.4 Study 1 Data Cleaning

Study 1 revealed many of the challenges that can arise in experimental skill acquisition research. Specifically, the aim of measurement was to obtain an accurate measure of a participants' performance on a given trial assuming they were applying reasonable effort. Extensive exploratory analysis of the resulting data revealed many problematic trials. These problematic trials occurred more frequently for some participants than others. Thus, from a research perspective some participants were more problematic than others. The following section outlines how problematic data was identified and how it was cleaned.

After a thorough analysis of key logs, task completion time, and accuracy data, the following procedure was adopted to clean the data. A small number of participants had corrupted or missing key log data, who were thus excluded from analysis. For participants with key log data, each trial was classified into one of seven categories (see Table A.1). As a percentage of all raw trials, 0.73% were Cheat, 61.31% were Accurate, 21.46% were Sloppy, 3.88% were Slow, 5.74% were Skip, and 4.72% were Avoid. There were 6 participants who had one or more trials classified as a cheat trial.

Table A.1: Study 1 Trial Type Classification for Validity Purposes.

	Type	**Definition**	**Cheat**	**Valid**	**Adjust**	**Penalty**
1.	Cheat	Pasted clipboard from previous trial	TRUE	FALSE	NA	100
2.	Accurate	Not cheat and Accuracy = 12	FALSE	TRUE	FALSE	0
3.	Sloppy	Not cheat and 9 <= Accuracy < 12	FALSE	TRUE	TRUE	1
4.	Slow	Trial timed out and 5 <= Accuracy < 9	FALSE	TRUE	TRUE	2
5.	Sleepy	Trial timed out and Accuracy < 5	FALSE	FALSE	NA	4
6.	Skip	Trial ended manually and 5 <= Accuracy < 9	FALSE	FALSE	NA	4
7.	Avoid	Trial ended manually and Accuracy < 5	FALSE	FALSE	NA	8

Note. Accuracy is out of 12.

Each trial type was given a penalty rating based on the degree to which the trial type reflected undesirable participant behaviour. These penalty points were then summed for each participant over the 54 trials of practice. Participants with more than 39.5 penalty points were removed from the cleaned dataset. This specific threshold was chosen for two reasons. First, the distribution of penalty points appeared to be a mixture distribution reflecting a main normally distributed group of participants who were presumed to have adequate skill and be applying reasonable effort, and a second skewed distribution derived from problematic participants. The chosen threshold discriminated these two distributions well. Second, actual examination of sequences of participant trial codes after the threshold started to appear problematic in an absolute sense, based on the frequency and type of problematic trials.

In addition to using trial types for assigning penalty points, trials types were also classified as either valid or invalid. Invalid trials were removed from the cleaned dataset. The result was that most retained participants still had one or two trials removed. As a percentage of all trials for retained participants, none were Cheat, 81.28% were Accurate, 15.29% were Sloppy, 1.47% were Slow, 1.23% were Skip, and 0.47% were Avoid.

Finally, trial completion time was adjusted for accuracy. Accuracy ratings were based on 12 checks done on the task. Accuracy of 12 out of 12 (i.e., 100%) had a multiple of 1.0, which meant that the completion time for the trial was unaltered. Table A.2 shows the multiples applied for less than perfect accuracy. The multiple aimed to approximate the time it would have taken the participant to complete the trial with perfect accuracy.

Fortunately, through careful data cleaning, meaningful data could be extracted from the majority of cases in Study 1. However, the challenges encountered in Study 1 provided several experimental design lessons that were incorporated into Studies 2 and 3. Trials labelled as cheating in the form of copy and pasting completed text from previous to subsequent trials was prevented in study 2 and 3 by clearing the clipboard after

Table A.2: Study 1 Task Completion Time Multiplier Associated with a Given Trial Accuracy out of 12.

Accuracy	RT Multiplier
12	1.00
11	1.05
10	1.10
9	1.15
8	1.20
7	1.25
6	1.30
5	1.40

each trial. Sleepy trials were subsequently largely avoided through the use of a system whereby the absence of a key press for five seconds or more paused the timer. Thus, participants could not engage in avoidant behaviour. Skipping was prevented in Study 2 and 3, by making end of trials contingent on accurately completing the required edits rather than allowing participants to end the trial themselves. Several issues with accuracy were subsequently resolved by designing the text editing task to only have one edit per trial and incorporating informative messages related to common accuracy issues such as leaving extra spaces in the edited text.

Appendix B

Study 2: Additional Materials

B.1 Materials

B.1.1 IPIP Personality Test

Big 5 personality was measured using the IPIP personality test (L. R. Goldberg et al., 2006). The details of each item are shown in Table B.1. The *Reverse* column indicates whether the item was negatively worded and needed to be reversed (i.e., $6 - x$, where x is the score on the item) before calculating total scores. The test aimed to measure the Big 5 personality factors: Extraversion (EX), Conscientiousness (C), Emotional Stability (ES), Openness (O), and Agreeableness (A).

Table B.1: Study 2 IPIP Big 5 Personality Measure

	Item Text	Reverse	Factor
1	I am the life of the party.	1	EX
2	I insult people.	-1	A
3	I am always prepared.	1	C
4	I get stressed out easily.	-1	ES
5	I have a rich vocabulary.	1	O
6	I often feel uncomfortable around others.	-1	EX
7	I am interested in people.	1	A

Table continues on next page

	Item Text	Reverse	Factor
8	I leave my belongings around.	-1	C
9	I am relaxed most of the time.	1	ES
10	I have difficulty understanding abstract ideas.	-1	O
11	I feel comfortable around people.	1	EX
12	I am not interested in other people's problems.	-1	A
13	I pay attention to details.	1	C
14	I worry about things.	-1	ES
15	I have a vivid imagination.	1	O
16	I keep in the background.	-1	EX
17	I sympathize with others' feelings.	1	A
18	I make a mess of things.	-1	C
19	I seldom feel blue.	1	ES
20	I am not interested in abstract ideas.	-1	O
21	I start conversations.	1	EX
22	I feel little concern for others.	-1	A
23	I get chores done right away.	1	C
24	I am easily disturbed.	-1	ES
25	I have excellent ideas.	1	O
26	I have little to say.	-1	EX
27	I have a soft heart.	1	A
28	I often forget to put things back in their proper place.	-1	C
29	I am not easily bothered by things.	1	ES
30	I do not have a good imagination.	-1	O
31	I talk to a lot of different people at parties.	1	EX
32	I am not really interested in others.	-1	A
33	I like order.	1	C
34	I get upset easily.	-1	ES
35	I am quick to understand things.	1	O
36	I don't like to draw attention to myself.	-1	EX
37	I take time out for others.	1	A
38	I shirk my duties.	-1	C
39	I rarely get irritated.	1	ES
40	I try to avoid complex people.	-1	O
41	I don't mind being the center of attention.	1	EX
42	I am hard to get to know.	-1	A
43	I follow a schedule.	1	C
44	I change my mood a lot.	-1	ES
45	I use difficult words.	1	O
46	I am quiet around strangers.	-1	EX
47	I feel others' emotions.	1	A
48	I neglect my duties.	-1	C

Table continues on next page

	Item Text	Reverse	Factor
49	I seldom get mad.	1	ES
50	I have difficulty imagining things.	-1	O
51	I make friends easily.	1	EX
52	I am indifferent to the feelings of others.	-1	A
53	I am exacting in my work.	1	C
54	I have frequent mood swings.	-1	ES
55	I spend time reflecting on things.	1	O
56	I find it difficult to approach others.	-1	EX
57	I make people feel at ease.	1	A
58	I waste my time.	-1	C
59	I get irritated easily.	-1	ES
60	I avoid difficult reading material.	-1	O
61	I take charge.	1	EX
62	I inquire about others' well-being.	1	A
63	I do things according to a plan.	1	C
64	I often feel blue.	-1	ES
65	I am full of ideas.	1	O
66	I don't talk a lot.	-1	EX
67	I know how to comfort others.	1	A
68	I do things in a half-way manner.	-1	C
69	I get angry easily.	-1	ES
70	I will not probe deeply into a subject.	-1	O
71	I know how to captivate people.	1	EX
72	I love children.	1	A
73	I continue until everything is perfect.	1	C
74	I panic easily.	-1	ES
75	I carry the conversation to a higher level.	1	O
76	I bottle up my feelings.	-1	EX
77	I am on good terms with nearly everyone.	1	A
78	I find it difficult to get down to work.	-1	C
79	I feel threatened easily.	-1	ES
80	I catch on to things quickly.	1	O
81	I feel at ease with people.	1	EX
82	I have a good word for everyone.	1	A
83	I make plans and stick to them.	1	C
84	I get overwhelmed by emotions.	-1	ES
85	I can handle a lot of information.	1	O
86	I am a very private person.	-1	EX
87	I show my gratitude.	1	A
88	I leave a mess in my room.	-1	C
89	I take offense easily.	-1	ES
90	I am good at many things.	1	O

Table continues on next page

	Item Text	Reverse	Factor
91	I wait for others to lead the way.	-1	EX
92	I think of others first.	1	A
93	I love order and regularity.	1	C
94	I get caught up in my problems.	-1	ES
95	I love to read challenging material.	1	O
96	I am skilled in handling social situations.	1	EX
97	I love to help others.	1	A
98	I like to tidy up.	1	C
99	I grumble about things.	-1	ES
100	I love to think up new ways of doing things.	1	O

B.1.2 Typing Test

The typing test consisted of three one minute trials. Before the task started, the following initial instructions were displayed on the screen:

> TYPING TEST: The following task assesses your typing speed. The task involves three trials. Each trial involves typing as much of a paragraph of text as you can in one minute. At the end of one minute the task will end. Your performance will be the number of words that you accurately type in the paragraph minus the number of words that you inaccurately type. Try to type as quickly and accurately as you can.

On each trial, the following message was displayed at the bottom of the screen: "YOU HAVE 60 SECONDS TO TYPE THE TEXT IN GREEN IN THE BOX BELOW:".

At the end of each trial, the screen changed to an all white background with the following text: "Press [ENTER] when you are ready for the next trial"

The typing test paragraphs were taken from articles on Wikipedia. The paragraphs were selected, and subsequently cleaned, so that they contained only words, numbers, and common punctuation.

Paragraph 1: `http://en.wikipedia.org/wiki/Psychology`

```
Psychology describes and attempts to explain
consciousness, behaviour, and social interaction.
Empirical psychology is primarily devoted to
describing human experience and behaviour as it
actually occurs.  Various schools of thought have
argued for a particular model to be used as a
guiding theory by which all, or the majority, of
human behaviour can be explained.  The popularity
of these has waxed and waned over time.  Some
psychologists may think of themselves as adherents
to a particular school of thought and reject the
others, although most consider each as an approach
to understanding the mind, and not necessarily
as mutually exclusive theories.  Psychology
encompasses a vast domain, and includes many
different approaches to the study of mental
processes and behaviour.
```

Paragraph 2: `http://en.wikipedia.org/wiki/Space`

```
The concept of space has been of interest for
philosophers and scientists for much of human
history.  The term is used somewhat differently
in different fields of study.  Hence, it is
difficult to provide an uncontroversial and
clear definition outside of specific defined
contexts.  Disagreement also exists on whether
space itself can be measured or is part of the
measuring system.  Science considers space to
be a fundamental quantity.  Thus an operational
definition is used in which the procedure of
measurement of space intervals and the units of
measurement are defined.  The way in which space
is perceived is an area which psychologists first
began to study in the middle of the 19th century,
and it is now thought by those concerned with such
studies to be a distinct branch within psychology.
```

Paragraph 3: `http://en.wikipedia.org/wiki/Peru`

```
Peru is a representative democratic republic
divided into 25 regions.  Its geography varies
from the arid plains of the Pacific coast to the
peaks of the Andes mountains and the tropical
forests of the Amazon Basin.  Its main economic
activities include agriculture, fishing, mining,
and manufacturing of products such as textiles.
The main spoken language is Spanish, although
a significant number of Peruvians speak native
languages.  This mixture of cultural traditions
has resulted in a wide diversity of expressions
in fields such as art, cuisine, literature, and
music.  Peruvian economic policy has varied widely
over the past decades.  Recent economic growth
has been fuelled by macroeconomic stability,
improved terms of trade, and rising investment
and consumption.  Peru's main exports are copper,
gold, zinc, textiles, and fish meal.
```

B.2 Text Editing Standardisation

The standardisation of task completion time on the text editing task can be expressed mathematically as follows.

$$z_{ij} = \frac{y_{ij} - \bar{y}_j}{\mathrm{sd}(y_j)}$$

where y_{ij} is the raw task completion time for the ith observation for the jth trialcode, $\bar{y}_j$ is sample mean raw task completion time for the jth trial code, $\mathrm{sd}(y_j)$ is the sample standard deviation for the jth trial code, and thus, z_{ij} is a z-score of raw task completion time within trialcodes. Then, adjusted task completion time, $\hat{y}_{ij}$, was calculated as

$$\hat{y}_{ij} = \bar{y}_{..} + \mathrm{sd}(y_{ij}) z_{ij}$$

where $\bar{y}_{..}$ is the grand mean of all observations, and $\mathrm{sd}(y_{ij})$ is the standard deviation of all observations

B.3 Text Editing Task Instructions

1. Motivating Introduction: The following initial instructions were given in an attempt to motivate participants to apply effort to the training task.

> I am making the assumption that most of you perform text editing using the mouse. While the mouse can be used to do text editing, this training will focus on performing text editing exclusively with the keyboard. Keyboard-based text editing is typically more efficient than using the mouse once you get familiar using the keys. In particular moving from typing to keyboard based text editing is quicker than moving to the mouse.

> We all spend a lot of time editing documents, writing emails, and entering text using a keyboard. While some of you may be familiar with some of the keys I will discuss, it is unlikely that you are familiar with them all. Text editing is the kind of task that can save you a minute a day, an hour a month, and a day a year. You may gain a whole month of your life in additional productivity, if you become efficient at text editing. Also, by learning how to efficiently edit text, you can focus more on the process of content creation.

2. *Overview of Hand Positioning*: The instructor then explained

> One of the keys to efficient text editing is proper hand positioning. Just as touch typing involves positioning the hands and fingers on the home keys, `ASDF` for the left hand and `JKL;` on the right hand, there is also an efficient positioning for hands and fingers for keyboard based text editing.

The instructor then held up the keyboard to the group of participants with the key faced towards them so they could hear and see where they should place their hands when performing text editing. Participants were then told that

> You should place your left little finger on the `Control` key and your left ring finger on shift which will be used to for modifying and selecting text; Place you left middle finger over the `Z`-key, which can be used to undo text and your left index finger can be used to press, `X`, `C` and `V` for cut, copy and paste, respectively. We will go over each of these keys later. For now, just focus on understanding where to place your fingers.

The instructor then held up the index, middle and ring fingers of the right hand and said,

> Now, place your right hand index, middle and ring fingers on the cursor keys. Thus, your index finger is on the `Left` arrow, your middle finger is on the `Up` or `Down` arrow, and your ring finger is on the `Right` arrow. You use your left index finger to press `Left` and `Backspace`, and `Delete`. You use your middle finger to press `Up`, `Down`, `Home` and `End`. You use your ring finger to press `Right` and you can also press `Page Up` and `Page Down`, although we wont be using the last two keys much in the training.

3. Follow-Along Demonstration of Text Editing Keys: Participants were then asked to open up a passage of text in Microsoft Word ^{TM}. It was the same passage of text as was used in the main Text Editing Task. Participants were told,

> The idea is for you to practice the keys after I read them out so that you can see what each key does. To start with I assume you are all familiar with the idea that `left-right` moves the cursor one character to the left and one character to the right and that `Up` and `Down` moves the cursor up and down lines. `Home` and `End` keys move the cursor to the beginning and end of the line. `Control Left` and `right` moves the cursor forward and back whole words at time. This is particularly useful as often we want to edit whole words. `Control up` and `control` down moves the cursor to the beginning of the previous or next paragraph. `Control Home` and `Control End` takes you the start and end of the document. Often we want to edit text. To do this we typically first have select the text. The main key for selecting text is the `Shift` key. If we hold down the `Shift` key and move the cursor keys `left` and `right` we can select characters. If we press `Shift Up` and `Down` we

> can select lines of text. We can use the `Control Shift` and `Left` or `Right` to select words. Once we've selected text we can use `Control X` to cut and `Control V` to paste or `Control C` to copy and `Control V` to paste. If we wish to undo a change we've just made, we can press `Control Z`. I assume that you are familiar with `Delete` and `backspace`. `Delete` removes characters to the right of the cursor and `backspace` removes characters to the left of the cursor. In addition, `Control Delete` and `Control Backspace` deletes the word in front or behind the cursor.

In between each of these sentences participants were given a few seconds to practice the key combinations that had just been presented. They were also shown the effect of the key combinations on text using an overhead projector.

4. Task Instructions: Participants were then told:

> In the next period you will have the opportunity to practice these short cut keys and acquire the skill of advanced text editing. The task involves a series of trials, which requires you to make a single change to a passage of text. Each trial will end once you have correctly made the change. Each trial automatically times out, if you are unable to complete the trial in 40 seconds. In order to proceed, you need to make the change exactly. In particular, when editing text, make sure that you leave only a single space between words. Leaving no space or leaving two spaces will not constitute a successful edit.

Participants were then instructed to click on a link on their computer to start the task.

B.4 Study 2 Data Cleaning

The aim of the initial data cleaning was to exclude trials that provided invalid data and exclude participants that did not provide sufficient valid trial data. The guiding principle was that a trial should be excluded if it did not yield valid information about the speed with which a person could complete the task. The *first* trial was removed because participants were typically still getting acquainted with the task interface. The *final* trial was removed because of interference caused by ending the experiment. A small number of trials were labelled *broken* because problems arose in determining accuracy.

In contrast to the previously mentioned types of invalid trials the following two types were related to participant behaviour. First, *delayed* trials were defined as those where no key was pressed for a period greater than six seconds. This was typically caused by participants taking a break or engaging in other off-task behaviour. Second, *incomplete* trials were those that timed out after 40 seconds. Probable causes included a lack of skill, making an irrecoverable error, distraction, and taking a break.

This second set of invalid trials were used to allocate *penalty points*. Each trial that fell into this second category was given a penalty point. A participant's penalty score was their percentage of trials with penalty points. After examining the distribution of penalty points and the associated data with various cut-offs, cases were retained with penalty points on fewer than 7.5% of trials. A small number of cases were also excluded because they had less than 27 minutes of on-task performance time.

Finally, an examination of trial types showed that cut-and-paste trials were taking around twice as long to complete as the other three trial types. Even with standardisation such trials would introduce substantial noise into analyses. For this reason, all cut-and-paste trials were excluded. Note that in all the above exclusion cases except the first trial, the excluded trial still counted towards the trial count for modelling and graphing purposes.

Appendix C

Study 3: Additional Materials

C.1 Materials

C.1.1 Prior Experience Questionnaire

The questions and response options for the *Prior Experience Questionnaire* are shown below. The Prior Experience Weighting is shown in brackets for each response option.

```
1. What is the computer Operating System that you use most often?
    1. Apple Mac (0)
    2. Windows (3)
    3. Linux (5)

2. When typing do you look at the keys?
    1. Always (1)
    2. Usually (2)
    3. Sometimes (3)
    4. Almost never (4)
    5. Never (5)
```

3. How often do you use the SHIFT key
   to select text on the computer?
   1. Never (0)
   2. Almost never (1)
   3. Sometimes (2)
   4. Usually (4)
   5. Always (6)

4. What is the status of your vision?
   1. Severe visual impairment (0)
   2. Mild visual impairment (2)
   3. Normal or corrected to normal (3)

5. Do you have normal motor control in your hands?
   1. Mild deficit of motor control in hands (0)
   2. Yes, normal motor control (2)

6. When editing text how often do you use keys such as
   control+c, x, and v for cut, copy, and paste?
   1. Never (0)
   2. Almost Never (1)
   3. Sometimes (2)
   4. Usually (3)
   5. Always (4)

7. How comfortable do you think you would be in
   using only the keyboard to edit text?
   1. Very uncomfortable (0)
   2. Somewhat uncomfortable (1)
   3. Neither comfortable nor uncomfortable (2)

4. Comfortable (3)

5. Very comfortable (4)

8. When you are editing text such as selecting,
copying, pasting, and so on,
what is the main device that you use?

1. Mouse (0)
2. Mouse and keyboard (2)
3. Keyboard (4)

9. How often do you play computer games?

1. Never (0)
2. Almost never (1)
3. Sometimes (2)
4. Often (3)
5. Very often (4)

10. How often do you play computer games where
you use a standard keyboard to play the game?

1. Never (0)
2. Almost never (1)
3. Sometimes (2)
4. Often (3)
5. Very Often (4)

11. Look down at the keyboard in front of you,
does it have the same key layout
as the keyboard you typically use?

1. No, it is different (0)
2. It looks similar (1)

```
    3. It's exactly the same (2)

12. Main computer
    1. Laptop (0)
    2. Desktop (1)

13. Would you describe yourself as a power-user of computers?
    1. Not at all (0)
    2. Probably Not (1)
    3. Possibly (2)
    4. Yes, definitely (3)

14. How often do you do computer programming?
    1. Never (0)
    2. Almost Never (1)
    3. Sometimes (2)
    4. Often (3)
    5. Very Often (4)
```

The following provides a brief rationale for each item in the Prior Experience Questionnaire.

1. The text editing task was conducted on a Microsoft Windows operating system and the text editing keys were standard for Microsoft Windows operating system. Thus, participants with more experience using Microsoft Windows, should be better at the task. While Linux users might be used to different text editing keys, they also tend to be advanced computer users.

2. Touch-typing is generally associated with faster typing. Faster typing leads to faster text editing. In addition, the Text Editing Task displays the required changes on the screen. Thus, if participants do

not need to move their eyes between the screen and the keyboard, they should perform better.

3. Using the `Shift` key for text editing on Microsoft Windows is indicative of a more serious adoption of the keyboard as a means of text editing. This can be contrasted with a participant who would usually use the mouse to select text and may or may not combine a few shortcut keys such as `Ctrl+X`, `Ctrl+C`, and `Ctrl+V`.

4. Lack of vision impairment should make deciphering text on the screen easier.

5. Text editing is a psychomotor task. In particular it requires rather complex coordination of movement of the fingers.

6. `Ctrl+X`, `Ctrl+C`, and `Ctrl+V` are some of the most commonly use shortcut keys. A participant who instead uses the mouse to activate menus to perform these tasks is likely to be less familiar with shortcut keys in general.

7. Self-reported comfort with using just the keyboard to edit text should be associated with skill in doing so.

8. Participants who normally uses just the keyboard to edit text, are likely to be better at using the keyboard.

9. Playing computer games may be indicative of greater exposure to computers. The nature of the experiment is also such that it resembles a computer game in some respects. There is a goal, a computer interface, feedback of results, and the aim is to complete the task as quickly as possible. Knowledge of these conditions, and skill in performing well in such situations may be related to playing computer games.

10. Computer games vary in the degree to which they involve use of the keyboard. Many PC games involve using the keyboard and often have a rather large set of shortcut keys for performing actions. Experience and skill with such games may improve text editing skills. In addition, such experience is likely to be correlated with more experience with computers

in general.

11. Keyboards vary in design. In particular, laptop keyboards typically have a different layout for navigational keys such as `Home`, `End`, `Left`, `Right` than do standard PC keyboards. If the keyboard that a participant typically used is the same as the one in the experiment, this may provide greater transfer of prior skill. Note: This item was not associated with text editing performance. It may be that knowing the difference between two keyboards is indicative of greater familiarity with computers.

12. Because laptops typically have a different arrangement of navigational keys, and because the experiment used standard PC keyboards, it was expected that the task would be easier for participants who did not typically use a laptop.

13. Participants who see themselves as advanced computer users are likely to be more experienced at text editing. Note: during the administration of the experiment, two or three participants, probably with English as a second language, asked "What is a power-user?". Thus, in future it may be better to reword "power" to "advanced". However, another perspective is that someone who does not know what a "power-user" is, is not a power-user.

14. Computer programming is a task performed by people who are advanced computer users. It implies that the participant has many years of experience using computers. Computer programming also encourages the development of advanced text editing skills because of the way that it creates a need to write and edit code efficiently.

C.1.2 IPIP Big 5 Personality Measure

Big 5 personality was measured using a 50 item version of the IPIP personality test (L. R. Goldberg et al., 2006). Items are shown in Table C.1.

The *Reverse* column indicates whether the item needed to be reversed (i.e., $6-x$, where x is the score on the item) before calculating total scores. The scale aimed to measure the Big 5 personality factors: Extraversion (EX), Conscientiousness (C), Emotional Stability (ES), Openness (O), and Agreeableness (A).

The initial instruction page had the following text:

> On the following pages, there are phrases describing people's behaviours. Please use the rating scale below to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Please read each statement carefully, and then circle the response most appropriate.

Each item was displayed on a separate page with the following text displayed on the top of the screen: "How accurately does this statement describes you?"

The response options were displayed as:

```
1=Very Inaccurate
2=Moderately Inaccurate
3=Neither Inaccurate nor Accurate
4=Moderately Accurate
5=Very Accurate
```

Table C.1: Study 3 Items Used in IPIP Big 5 Personality Measure

	Item Text	Reverse	Factor
1	I Am the life of the party	1	EX
2	I Feel little concern for others	-1	A
3	I Am always prepared	1	C
4	I Get stressed out easily	-1	ES
5	I Have a rich vocabulary	1	O
6	I Don't talk a lot	-1	EX
7	I Am interested in people	1	A
8	I Leave my belongings around	-1	C
9	I Am relaxed most of the time	1	ES
10	I Have difficulty understanding abstract ideas	-1	O
11	I Feel comfortable around people	1	EX
12	I Insult people	-1	A
13	I Pay attention to details	1	C
14	I Worry about things	-1	ES
15	I Have a vivid imagination	1	O
16	I Keep in the background	-1	EX
17	I Sympathize with others feelings	1	A
18	I Make a mess of things	-1	C
19	I Seldom feel blue	1	ES
20	I Am not interested in abstract ideas	-1	O
21	I Start conversations	1	EX
22	I Am not interested in other peoples problems	-1	A
23	I Get chores done right away	1	C
24	I Am easily disturbed	-1	ES
25	I Have excellent ideas	1	O
26	I Have little to say	-1	EX
27	I Have a soft heart	1	A
28	I Often forget to put things back in their proper place	-1	C
29	I Get upset easily	-1	ES
30	I Do not have a good imagination	-1	O
31	I Talk to a lot of different people at parties	1	EX
32	I Am not really interested in others	-1	A
33	I Like order	1	C
34	I Change my mood a lot	-1	ES
35	I Am quick to understand things	1	O
36	I Don't like to draw attention to myself	-1	EX
37	I Take time out for others	1	A
38	I Shirk my duties	-1	C

Table continues on next page

	Item Text	Reverse	Factor
39	I Have frequent mood swings	-1	ES
40	I Use difficult words	1	O
41	I Don't mind being the center of attention	1	EX
42	I Feel others emotions	1	A
43	I Follow a schedule	1	C
44	I Get irritated easily	-1	ES
45	I Spend time reflecting on things	1	O
46	I Am quiet around strangers	-1	EX
47	I Make people feel at ease	1	A
48	I Am exacting in my work	1	C
49	I Often feel blue	-1	ES
50	I Am full of ideas	1	O

C.2 Text Editing Task

C.2.1 Text Editing Task Initial Instructions

All instructions for the text editing task were administered via computer. Participants were first required to read a series of instruction pages. The participant could only navigate to the next instruction page after a minimum amount of time had passed. This minimum time was set as the time it took to read aloud the passage at a moderate rate. This was done to maximise the chance that participants read the instructions carefully. The instructions are shown below:

"The following screens present important task instructions. Please read these instructions slowly and carefully. Failure to understand these instructions may make your results invalid. To ensure careful reading it is not possible to go to the next instruction screen until the expected reading time has passed.

You are about to complete an experimental task that lets you practice Windows keyboard-based text editing. Most people who perform text editing use the mouse. Once the skill is acquired keyboard-based text editing is more efficient than mouse-based text editing. There are several reasons for this including: 1) it is quicker to move from typing to keyboard-based editing than it is to move from typing to the mouse; 2) keyboard-based text editing is more precise; 3) pressing a key to edit is quicker than accessing menus, toolbars, and other mouse based editing options. Given the amount of time people spend using computers becoming skilled at text editing can save you substantial time. Such a skill can also help you focus on the broader goals of editing your text, such as improving your writing.

The task uses only the keyboard to edit text. The mouse will not work, and it should not be used. Please place the mouse behind your computer now.

The experiment is made up of 30 blocks. Each block goes for 80 seconds. Each block contains multiple trials. Each trial requires you to delete a passage of text marked in red as quickly as possible. Once you have correctly made the deletion the trial automatically ends. Your performance on each block is based on the number of trials you can accurately complete in 80 seconds. If you stop performing the task for more than 5 seconds, the Trial Timer will pause and will restart when you recommence editing. If you have not successfully completed the edit after 30 seconds of elapsed trial time, the trial will end. If you make a mistake during a trial, you can press CONTROL+Z to undo the change.

This study aims to provide meaningful data for the Data Analysis Task and class discussion. It may also be used for publication purposes to advance knowledge in the area of skill acquisition. Thus, it is important that a certain standard of experimental control exists during the exercise. Specifically:

- Do not talk to other students until the task is complete

- Do not instruct other students on how to complete the task
- Do not look at how other students are completing the task
- Do not use your computer for other purposes such as the internet
- Turn off or put on silent your mobile phone
- If you have a question about the experiment, raise your hand and your tutor will come round
- If you need to have a break during the task, feel free to stretch or go for a short walk, but do not talk to other students
- If you do not want or are unable to continue to do the task, feel free to leave the room and come back at 1 hour and 10 minutes after class commenced (e.g., if your class started at 11:00AM come back at 12:10PM)

Before the task commences you are going to be given some instructions about keyboard-based text editing. Before being introduced to each of the keys, you will be introduced to proper hand positioning.

- Place your left little-finger on the LEFT CONTROL KEY and your left ring-finger on the LEFT SHIFT KEY
- Place your left middle finger over the Z-key (undo) and your left index finger can be used to press X (cut), C (copy), and V (paste)
- Place your right hand index, middle and ring fingers on the arrow keys.
- Use your right index finger to press LEFT, BACKSPACE, and DELETE.
- Use your right middle finger to press UP, DOWN, HOME, and END.

- Use your right ring finger to press RIGHT

In the next section of the instructions you will be introduced to each of the text editing keys. Each instruction screen sets out a couple of text editing keys for you to briefly practice. On each screen try to get a sense of what each key presented does. Look at the effect of the key press on the text editor. Try to use the recommended hand position mentioned previously. Once these key instructions are over the main task will commence. You can proceed to the next screen of key instructions once 12 seconds has passed and you have pressed EACH key introduced FOUR times."

C.2.2 Additional Controls

If a participant did not press a key for five seconds, trial time was paused and the following message was displayed: "No user activity: timer has stopped. If you stopped because you made an error, hold down Control+Z to Undo." This message was incorporated into the text editing software for several reasons: (a) Five seconds was long enough to reflect actual task withdrawal behaviour as opposed to natural delays between key presses. (b) The mechanism discouraged off-task behaviour, because any off-task behaviour added to the total time to complete the study. (c) Prevalence of off-task behaviour was not an aspect of performance that this study aimed to model. Thus, trial pauses were used to minimise the influence of distractions, breaks, and other off-task behaviour on performance measurement. (d) Participants who made an error from which they could not recover often triggered the pause. The error message communicated how to recover from such an error using the Undo functionality.

To prevent participants making the mistake of leaving two spaces where there should be one, the following message was displayed: "You need to remove additional spaces from your text".

C.2.3 Training Condition Instructions

The following instructions were displayed after block 15 to participants in the Training Condition.

> You are now going to receive some additional training in keyboard-based text editing. You will first receive some general theory. Then, you will have some practice. To ensure careful reading it is not possible to go to the next instruction screen until the expected reading time has passed.
>
> Text editing involves three core elements: movement, selection, and manipulation. Speed of text editing can be improved by reducing the time between key presses or by reducing the required number of key presses. Reducing key presses often requires some thought and can involve using more sophisticated key combinations. However, with practice key selection becomes automatic and pressing the keys becomes a smooth action.
>
> Movement involves pressing keys to move the cursor from the current location to the target location. In general, the task can often be broken down into steps: 1) get cursor to paragraph [use control+down/up; or control+home/end]; 2) get cursor to line [use up/down]; 3) get cursor to word [use control+left/right; or home/end]. Sometimes it is quicker to go passed the target location and then come back (e.g.,control+end then control+left).
>
> Deletion can be done by simply pressing delete or backspace. Faster options include control+delete, control+backspace, and select and delete. The select and delete

strategy involves selecting the text and then pressing delete. This is a desirable deletion strategy as the number of words to be deleted increases. It involves four steps. 1) Get to one end of the text to be deleted; 2) hold down shift; 3) move cursor to other end of text; 4) press delete. Steps 1 and 3 apply all the principles of movement. Efficient movement depends on minimising key presses.

On the following screens you will do 6 Training Trials similar to the main task. Your goal is NOT speed. Your goal is to minimise key presses by applying the theory presented on the previous screens. Each Training Trial is made up of three Sub-Trials:

Sub-trial A) This is your first attempt to minimise key presses

Sub-trial B) You are presented with a sequence of keys which will minimise key presses. Your task is to follow them exactly.

Sub-trial C) The aim is to use the keys set out in Sub-trial B without assistance.

To communicate the sequence of keys that you should press, key sequences like this will be presented:

Down, Down, End, Control(Left, Left), Shift(Down, Control(Left, Left, Left, Left)), Delete

The order of the words indicates the sequence that you should press the keys. Brackets are used to indicate that you should continue to hold down the key preceding the bracket. E.g., ‘Control(Left, Left)’ indicates that you should hold down control while you press Left twice. If you have followed the sequence, the text should be successfully deleted.

As you press the key sequence make sure to also observe the effect that the sequence is having on the text editor.

C.2.4 Text Editing Task Problems

Table C.2 shows the database of 500 text editing task problems used to create the text editing trials in Study 3. The *Words* column indicates the number of words. The *Point* column indicates the number of characters into the text editing original text that the *Deletion Text* was inserted.

Table C.2: Study 3 Text Editing Task Items

ID	Words	Point	Deletion Text
1	1	1	boy
2	1	7	school
3	1	13	must
4	1	19	covered
5	1	25	song
6	1	31	city
7	1	37	such
8	1	43	a
9	1	49	now
10	1	55	few
11	1	61	known
12	1	67	paper
13	1	73	wood
14	1	79	night
15	1	85	my
16	1	91	color
17	1	97	really
18	1	103	kind
19	1	109	would
20	1	115	study
21	1	121	air
22	1	127	car
23	1	133	animal
24	1	139	plan
25	1	145	thousands
26	1	151	my
27	1	157	oh
28	1	163	example
29	1	169	something
30	1	175	too

Table continues on next page

ID	Words	Point	Deletion Text
31	1	181	oil
32	1	187	machine
33	1	193	war
34	1	199	stand
35	1	205	only
36	1	211	once
37	1	217	vowel
38	1	223	large
39	1	229	money
40	1	235	notice
41	1	241	pulled
42	1	247	street
43	1	253	street
44	1	259	horse
45	1	265	water
46	1	271	travel
47	1	277	country
48	1	283	heavy
49	1	289	their
50	1	295	ever
51	2	1	am fall
52	2	7	would once
53	2	13	men father
54	2	19	life without
55	2	25	products music
56	2	31	surface remember
57	2	37	five become
58	2	43	upon surface
59	2	49	animal she
60	2	55	room told
61	2	61	life king
62	2	67	went he
63	2	73	try became
64	2	79	change don't
65	2	85	time was
66	2	91	when our
67	2	97	also until
68	2	103	than heavy
69	2	109	side check
70	2	115	body round
71	2	121	learn gave
72	2	127	seen numeral

Table continues on next page

ID	Words	Point	Deletion Text
73	2	133	questions important
74	2	139	animal am
75	2	145	be eat
76	2	151	year took
77	2	157	space understand
78	2	163	line heavy
79	2	169	play number
80	2	175	space up
81	2	181	those hear
82	2	187	run one
83	2	193	an these
84	2	199	we add
85	2	205	during building
86	2	211	write use
87	2	217	sea always
88	2	223	even grow
89	2	229	below ask
90	2	235	write him
91	2	241	both said
92	2	247	hear should
93	2	253	area after
94	2	259	game show
95	2	265	go began
96	2	271	hours space
97	2	277	war young
98	2	283	all special
99	2	289	follow music
100	2	295	their set
101	3	1	find no think
102	3	7	try step find
103	3	13	learn become not
104	3	19	morning reached play
105	3	25	light start off
106	3	31	light birds three
107	3	37	wind map carry
108	3	43	give didn't read
109	3	49	about him black
110	3	55	world should measure
111	3	61	know too object
112	3	67	letter fall less
113	3	73	they way another
114	3	79	call school near

Table continues on next page

ID	Words	Point	Deletion Text
115	3	85	however common soon
116	3	91	together hours voice
117	3	97	large travel wait
118	3	103	got though city
119	3	109	got had no
120	3	115	just second reached
121	3	121	read surface cut
122	3	127	sing our music
123	3	133	house line covered
124	3	139	there friends America
125	3	145	inches many thing
126	3	151	wood reached saw
127	3	157	how ship noun
128	3	163	great noun problem
129	3	169	learn went which
130	3	175	voice bring mile
131	3	181	idea play up
132	3	187	step saw take
133	3	193	learn scientists hours
134	3	199	island my made
135	3	205	an rest grow
136	3	211	world car other
137	3	217	father an use
138	3	223	after how soon
139	3	229	long several fast
140	3	235	until didn't shape
141	3	241	did several less
142	3	247	put with before
143	3	253	deep verb time
144	3	259	full ten eye
145	3	265	language whole family
146	3	271	field week object
147	3	277	cannot write different
148	3	283	circle base into
149	3	289	north plane take
150	3	295	America less mile
151	4	1	ask eye year thousands
152	4	7	her five a small
153	4	13	whole another came didn't
154	4	19	fly remember step knew
155	4	25	early every got this
156	4	31	three vowel land unit

Table continues on next page

ID	Words	Point	Deletion Text
157	4	37	among of inches life
158	4	43	deep carry live understand
159	4	49	he write leave light
160	4	55	start work left stop
161	4	61	than himself family than
162	4	67	horse old into ball
163	4	73	king complete dry may
164	4	79	understand knew several wind
165	4	85	and another but been
166	4	91	not fish leave hand
167	4	97	filled must space stood
168	4	103	even men word gave
169	4	109	hand show them often
170	4	115	whole friends almost base
171	4	121	door got different went
172	4	127	today since gave yes
173	4	133	stand found vowel grow
174	4	139	men piece may the
175	4	145	it's white side any
176	4	151	cut both come way
177	4	157	and horse don't system
178	4	163	since white important green
179	4	169	we own another animal
180	4	175	any as measure along
181	4	181	war all cannot land
182	4	187	once life night four
183	4	193	cried system little her
184	4	199	point come room car
185	4	205	small follow three include
186	4	211	second deep waves it
187	4	217	pattern find front such
188	4	223	minutes follow shape there
189	4	229	material filled four room
190	4	235	equation move kind sing
191	4	241	four check after fire
192	4	247	over dry read your
193	4	253	away size plan run
194	4	259	farm eye well ran
195	4	265	near family are more
196	4	271	children him man me
197	4	277	money be morning became
198	4	283	saw keep air scientists

Table continues on next page

ID	Words	Point	Deletion Text
199	4	289	large by name my
200	4	295	they mean write cold
201	5	1	color feel white give covered
202	5	7	color complete use special too
203	5	13	many fire more told hand
204	5	19	mean almost language place read
205	5	25	song than less minutes your
206	5	31	over near mother down song
207	5	37	wheels from state how father
208	5	43	hand own but around mile
209	5	49	decided base up notice one
210	5	55	finally thing dog became father
211	5	61	body measure rest yet soon
212	5	67	first us explain all river
213	5	73	building vowel large money travel
214	5	79	work shown farm look covered
215	5	85	between green often about country
216	5	91	together sing spell write able
217	5	97	object for any you ago
218	5	103	example king tell fast of
219	5	109	made low make problem two
220	5	115	once remember best street understand
221	5	121	happened slowly six man eat
222	5	127	come gave stood products through
223	5	133	its found body material deep
224	5	139	white even wheels inches hot
225	5	145	of six of plan children
226	5	151	not pulled side never family
227	5	157	family quickly time left boat
228	5	163	warm not more north give
229	5	169	later think great country remember
230	5	175	explain am plan understand most
231	5	181	had will mile home stay
232	5	187	would animal begin would that
233	5	193	dry talk head hard be
234	5	199	plant always carefully animal found
235	5	205	material no most government upon
236	5	211	the may young stay usually
237	5	217	less five different upon left
238	5	223	these wait horse around turn
239	5	229	if away because night fire
240	5	235	check whole until something inches

Table continues on next page

ID	Words	Point	Deletion Text
241	5	241	live story piece by night
242	5	247	paper well home word both
243	5	253	strong go ground people any
244	5	259	story power grow only come
245	5	265	halt use would farm wait
246	5	271	until strong mark six English
247	5	277	your language hard small hot
248	5	283	answer hear though some while
249	5	289	class see ball second plan
250	5	295	during inside letter left page
251	6	1	surface sing no animal seen year
252	6	7	travel time still upon base room
253	6	13	face state building paper only head
254	6	19	those wheels side food too special
255	6	25	same her does fish big full
256	6	31	sure became one I'll close base
257	6	37	people back kind call true a
258	6	43	new with far book young less
259	6	49	song sound say she only another
260	6	55	did filled oh friends its as
261	6	61	in fact too against war this
262	6	67	verb but inside home song land
263	6	73	day bring at birds plane system
264	6	79	space passed king quickly that class
265	6	85	its yet example which old show
266	6	91	should sometimes learn other clear ocean
267	6	97	today color give its new then
268	6	103	step four own not waves have
269	6	109	circle contain music pair hours rest
270	6	115	letter thing boat money room both
271	6	121	stars away are being much way
272	6	127	walk hold plane space number then
273	6	133	new did give three early toward
274	6	139	above decided her oil began pulled
275	6	145	waves inside inches river animal passed
276	6	151	gave part inside your so ball
277	6	157	became group may list girl contain
278	6	163	came keep number vowel read sea
279	6	169	kind ball products sea there add
280	6	175	grow sing oil old seen minutes
281	6	181	eat has object stars above when
282	6	187	map inches all were behind just

Table continues on next page

ID	Words	Point	Deletion Text
283	6	193	look ocean made look land listen
284	6	199	any town among during vowel ago
285	6	205	red me word day example back
286	6	211	known other old south machine and
287	6	217	figure came though sure listen food
288	6	223	children circle low a warm above
289	6	229	inside white year nothing door cannot
290	6	235	once write animal us father include
291	6	241	music heard verb together fact face
292	6	247	each name often they think people
293	6	253	must him explain through work around
294	6	259	mile ago car her water yes
295	6	265	plant covered explain important do building
296	6	271	boat thousands building last hundred they
297	6	277	another when king plane follow thing
298	6	283	answer watch is their letter stand
299	6	289	class try sound behind below room
300	6	295	first play special note America young
301	7	1	set time high top fish less inside
302	7	7	though form deep certain travel even English
303	7	13	man been follow number make made our
304	7	19	best ball become need fine person cut
305	7	25	done among week week little carry side
306	7	31	note always cut game war answer hand
307	7	37	whole these vowel got house mark there
308	7	43	say space great men room she around
309	7	49	way strong were it here friends size
310	7	55	order people river happened put pulled list
311	7	61	he himself another all took mountain oh
312	7	67	near full became great off knew oil
313	7	73	and power country it's halt any began
314	7	79	ago once her front learn black close
315	7	85	reached many passed red understand other listen
316	7	91	today open have mile wait back mean
317	7	97	old study should machine been hundred money
318	7	103	across that other king have explain problem
319	7	109	told finally house early sun boy listen
320	7	115	told behind might state live ball reached
321	7	121	not same white minutes include day grow
322	7	127	questions miss fall made one write quickly
323	7	133	light time life form paper war follow
324	7	139	earth stand until week mark found correct

Table continues on next page

ID	Words	Point	Deletion Text
325	7	145	told week hear common top back read
326	7	151	questions mile put there less I'll part
327	7	157	him order not whole figure saw farm
328	7	163	circle him by I'll school air circle
329	7	169	again we might at look example learn
330	7	175	south high cut mountain look family is
331	7	181	do halt food group course morning spell
332	7	187	second voice covered ocean before never side
333	7	193	around song became important home tree an
334	7	199	boy family cold pulled circle south stand
335	7	205	plant air too understand full front game
336	7	211	yes scientists small ground cried vowel want
337	7	217	dry thought fly heavy don't wood just
338	7	223	think mean whole fish plane sentence both
339	7	229	last note rock above could place slowly
340	7	235	off can stop last any I'll feet
341	7	241	produce vowel sound write book next five
342	7	247	large few children sometimes mountain scientists white
343	7	253	island able like out language any ground
344	7	259	sea cannot complete high us game verb
345	7	265	system toward heard been however city class
346	7	271	or vowel such heavy plant I north
347	7	277	begin hold explain do reached has him
348	7	283	cannot miss why his several idea course
349	7	289	correct inches down line was feet once
350	7	295	king both its go don't bring went
351	8	1	heard has word fact correct until run check
352	8	7	number your king air hand black use every
353	8	13	verb deep true form leave come took better
354	8	19	watch always animal surface what look too base
355	8	25	note building being must below top today easy
356	8	31	fall seem road plan I miss measure over
357	8	37	plane river but order me mountain down find
358	8	43	she sound toward the special me start now
359	8	49	through north two where oh go knew road
360	8	55	men inside three eye color did where watch
361	8	61	hand problem them such carefully should among time
362	8	67	known of answer river it's way the face
363	8	73	has let oil all friends products do after
364	8	79	big spell way mark little here told around

Table continues on next page

ID	Words	Point	Deletion Text
365	8	85	did country group an plant inches became six
366	8	91	man carefully long with sing low birds look
367	8	97	might the always very long among less unit
368	8	103	body life south saw eye heavy stop came
369	8	109	full green front him true inches language check
370	8	115	also way their eye farm set table hear
371	8	121	south town little ever form sun notice first
372	8	127	line below fine if didn't ball there during
373	8	133	often set against pattern every ball down numeral
374	8	139	round against keep measure if sound which mile
375	8	145	night equation these said enough built girl next
376	8	151	plan really ran other many any come farm
377	8	157	understand let water light well person people scientists
378	8	163	air rule far much best cold house I'll
379	8	169	something base town contain farm some two course
380	8	175	night his such knew he young equation material
381	8	181	size five after study your own explain surface
382	8	187	stand together fine under language picture its short
383	8	193	figure always check remember try man fast yet
384	8	199	though name were made sound has water right
385	8	205	front would call is rest pair among under
386	8	211	an food yes area of story near hold
387	8	217	stop were stood head heard in year part
388	8	223	easy close area products sure known however however
389	8	229	slowly children being great better miss night ran
390	8	235	size friends you ran turn oil there upon
391	8	241	your earth brought came might seen about well
392	8	247	almost known or thousands paper island plane English
393	8	253	toward during he piece space shape below is
394	8	259	car need plant fire week face power include
395	8	265	and always five saw never food way food
396	8	271	seem after before land move since problem are
397	8	277	today order she tell birds color almost play
398	8	283	began told too knew any true stood old
399	8	289	mark food car morning start able plane example
400	8	295	become his base across unit then wheels every
401	9	1	hot street example follow most turn live school were

Table continues on next page

ID	Words	Point	Deletion Text
402	9	7	people into ten leave went the yet place work
403	9	13	true live below car because tree heavy cannot short
404	9	19	seen music answer town had ball before clear boy
405	9	25	yes man sentence start any rest whole side number
406	9	31	even men her thing as nothing list after sound
407	9	37	dog above mark look well both get there map
408	9	43	only seem food where mean and will these her
409	9	49	food oh area base seen while tell whole different
410	9	55	said he our listen complete point measure one fall
411	9	61	both road the ball school minutes letter upon ocean
412	9	67	both every covered right war done often city ship
413	9	73	ground do week base remember they those call his
414	9	79	sing keep well heat reached cold list course into
415	9	85	noun king own being young then seem however without
416	9	91	fall an change I me follow easy him may
417	9	97	paper across certain began problem long low stars follow
418	9	103	our English white there were song pair has ocean
419	9	109	pulled my since those later week may live plant
420	9	115	oh top really need people run have find full
421	9	121	only space am around want walk door that minutes
422	9	127	well began how happened known street back building rest
423	9	133	until water voice if mean figure stand use nothing
424	9	139	became south should there open leave verb story add
425	9	145	made carefully find ask because heard just run road
426	9	151	much letter easy back yes see your known wait
427	9	157	three room south dog door before time town during
428	9	163	slowly keep once sometimes didn't next last wait several
429	9	169	him bring toward special being heavy less hear picture
430	9	175	become covered keep happened yes start such check world
431	9	181	first miss set area front special special building ago
432	9	187	I'll their deep didn't our almost same tell watch
433	9	193	six slowly boat town wait near the king sun
434	9	199	told street call special an name step point near

Table continues on next page

ID	Words	Point	Deletion Text
435	9	205	halt building big rock easy hear verb waves been
436	9	211	if away here told what mountain sure what first
437	9	217	pair if night plane back land close but school
438	9	223	didn't building close certain tell measure never told the
439	9	229	must away when friends land on our ship plane
440	9	235	go against out hear idea another strong ocean fire
441	9	241	around book cold through door family usually by only
442	9	247	your than book family group page went don't mark
443	9	253	food war that sometimes ten from up sing problem
444	9	259	box plan family live must its other this material
445	9	265	story hard across clear important inches voice map north
446	9	271	those top cannot food sun three across how also
447	9	277	began end equation numeral being note island dog boy
448	9	283	table waves each usually word did close ocean change
449	9	289	learn then heard fish a heat five music correct
450	9	295	warm they day front really write ground us or
451	10	1	grow form circle put tree about food English better I'll
452	10	7	those so you give short this back equation toward had
453	10	13	street circle hold young will say noun front but way
454	10	19	look check I six problem machine write wind birds passed
455	10	25	both air example products up when have me street story
456	10	31	let stay quickly miss course line yes size class small
457	10	37	above or hard can so complete plant all one who
458	10	43	been father class contain yes around knew world begin ship
459	10	49	line a only first clear nothing heat would also see
460	10	55	bring watch take wood knew their stars came there fine
461	10	61	air very better hold told there wind mile hot whole
462	10	67	sure feel put some enough night want stood sure listen
463	10	73	problem special city grow is questions to in next yet

Table continues on next page

ID	Words	Point	Deletion Text
464	10	79	easy each table four little check feel inches stars see
465	10	85	building just mile more farm heavy week ago very mountain
466	10	91	our about mother side put island part order have not
467	10	97	thousands like shape plant explain hundred then add example walk
468	10	103	sure surface kind turn very turn ten ask deep notice
469	10	109	pair such door scientists week food street around across against
470	10	115	live feel began often plan know power name change mile
471	10	121	pair reached town also group place study war scientists long
472	10	127	fast contain word more talk dark able halt remember mean
473	10	133	government show just were cried men pair try nothing idea
474	10	139	note may think south plan person today deep put door
475	10	145	north there round head still carry how where over made
476	10	151	language covered or and four boat across does try earth
477	10	157	me remember five king that more cannot light was travel
478	10	163	every take right from best short only travel waves horse
479	10	169	first short during always light than field material watch carry
480	10	175	list which many most such whole whole notice word same
481	10	181	tree kind you few heard enough decided more field heavy
482	10	187	part never complete use for big under mark very wait
483	10	193	built yes well north knew head true the them number
484	10	199	will full room notice piece its he animal force went
485	10	205	fact form by tell front line life front help water

Table continues on next page

ID	Words	Point	Deletion Text
486	10	211	music though size easy mile and study early watch head
487	10	217	just wind children right right me space notice object built
488	10	223	side far something equation five do too run happened mark
489	10	229	cannot something off listen fish said it follow yet also
490	10	235	him dry strong pattern begin far stop sure heavy through
491	10	241	cannot scientists Indian ship write those young new material word
492	10	247	carry animal place sentence many begin cried are questions would
493	10	253	birds children gave example cannot be many best order something
494	10	259	room built plan wood hot any enough front easy contain
495	10	265	America base he often because think sometimes found explain yet
496	10	271	rest turn correct us feet in city fact it's hand
497	10	277	children form fall plane year use work help like city
498	10	283	not boat need idea wait piece set list thousands land
499	10	289	correct got done took start side short would need explain
500	10	295	old numeral problem carefully figure all wind fire kind note