FACTORS INFLUENCING THE ACCURACY OF TASK COMPLETION TIME ESTIMATES

by

KEVIN EDWARD THOMAS

A thesis submitted to the University of Plymouth

in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

Faculty of Science

February 2004

Factors Influencing the Accuracy of Task Completion Time Estimates

Kevin Edward Thomas

Abstract

Whilst considerable research has found that people tend to underestimate their task completion times (e.g., Buehler et al., 1994), factors that might influence the accuracy of temporal predictions have received little empirical treatment. The research presented in this thesis identified two distinct factors that mediated time estimation accuracy and bias. One factor was task duration, whereas the other factor was the person's prior experience of the task. There was evidence that having prior experience of performing all or a substantial part of the same task enabled participants to more accurately estimate its duration. Additionally, predictions were more accurate when participants viewed tasks before making time estimates. Contrary to the theory of the planning fallacy (Kahneman & Tversky, 1979), these findings suggest that people do take account of their previous task performance, and use such distributional information to good effect. However, there was evidence of time prediction bias when unrelated tasks were completed beforehand, suggesting that erroneous information about previous task performance was used when making a subsequent estimate.

The directional nature of time estimation bias was also highlighted in the present research. In general, there was some evidence of temporal overestimation on tasks with a duration of up to four or five minutes, whereas participants tended to underestimate their completion times on tasks that took between eight and 16 minutes to complete. These findings indicate that task duration influences the direction in which time estimates are biased (i.e., under or overestimation), with the temporal underestimation indicative of the planning fallacy occurring on tasks of at least eight minutes' duration. The present research has potential implications for task duration estimation in everyday life, and outlines conditions under which prediction bias can be reduced. The present findings are discussed in relation to the theory of the planning fallacy and the potential role of cognitive judgemental heuristics in determining temporal misestimation.

List of Contents Page

List of Contents (Continued) Page

List of Contents (Continued) Page

List of Contents (Continued) Page

Appendices

List of Tables and Figures Page

Chapter Two

Experiment 1

Experiment 2

Experiment 3

Chapter Three

Experiment 4

Acknowledgements

The author wishes to thank the Economic and Social Research Council (ESRC) of the United Kingdom for granting him a postgraduate studentship award, which provided him with the financial means necessary to undertake this doctoral research degree.

The author also wishes to thank his Director of Studies, Professor Stephen Newstead, and his Second Supervisor, Doctor Simon Handley, for their professional guidance, advice and support during the period of this research degree.

Finally, the author wishes to thank his partner, Katrina Waters, for all her unstinting support and companionship throughout the period of this research degree.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This programme of research was financed by a postgraduate studentship award from the Economic and Social Research Council (ESRC) of the United Kingdom (Award Number: R42200034413).

A programme of advanced study was undertaken prior to the commencement of this research programme. This study programme took the form of a one-year full-time MSc. course in Psychological Research Methods at the University of Plymouth (1999 - 2000). During the course of the present research programme, two scientific conferences have been attended at which the author's work was presented. This research programme has thus far yielded the publication of one journal article, a copy of which is presented in Appendix 10 of this thesis (pp. 247 - 284).

Publication:

Thomas, K. E., Newstead, S. E. & Handley, S. J. (2003). Exploring the Time Prediction Process: The Effects of Task Experience and Complexity on Prediction Accuracy. Applied Cognitive Psychology, 17, 6, 655 - 673.

Presentation and Conferences Attended:

Thomas, K. E. (2003). Exploring the Time Prediction Process: The Effects of Task Experience and Complexity on Prediction Accuracy. Paper Presented at the Annual Conference of the Cognitive Section of the British Psychological Society (BPS) at the University of Kent, Canterbury (9 - 11 September 2002).

Thomas, K. E. (2003). Estimating Task Completion Times: Task Experience and Duration Affect Temporal Prediction Bias. Paper Presented at the Biennial Meeting of the Society for Applied Research in Memory and Cognition (SARMAC) at the University of Aberdeen (2 - 6 July 2003).

Signed.........................

Date......13.2.04..........

Chapter One

Introduction

1.1 Overview

The process of estimating how much time an upcoming task will take to complete has been the focus of considerable research (e.g., Byram, 1997; Josephs & Hahn, 1995; Koole & Van't Spijker, 2000). Such research has found that people are generally over-optimistic, that is, they tend to underestimate their task completion times. A closely related and well established cognitive judgement phenomenon is the planning fallacy, which was identified by Kahneman and Tversky (1979).

The planning fallacy is the tendency to make optimistically biased estimates of the duration of a current task despite being aware that previous similar activities took more time than predicted. Kahneman and Tversky suggest that the planning fallacy is a consequence of heuristic information processing whereby information concerning the current task becomes the focus of attention at the expense of information pertaining to previous similar tasks. Whilst there is considerable research evidence of the planning fallacy (Buehler, Griffin & Ross, 2002), there has been little empirical treatment of factors that might influence the accuracy of task completion time estimates.

This thesis focuses on a number of potentially important factors thought to mediate the accuracy of people's estimates of their task completion times. Principally, these factors concern the nature of the task itself, and people's prior experience of the task that is about to be performed. The factors of interest in relation to task experience are pre-exposure to an identical task, and prior experience of performing similar or dissimilar tasks. In relation to the nature of the task, the factors of interest are task complexity and task duration. In addition to these factors, one study (Experiment 3) examines the impact of monetary rewards on time estimation bias. By investigating these potential mediating factors, the present research aims to provide information concerning the reasons for temporal misestimation as well as highlighting conditions under which the accuracy of judgements of task duration can be improved.

Given the lack of research into factors that might mediate time estimation accuracy, it is necessary for a extensive range of literature to be reviewed in this chapter. For example, it is necessary to consider research into the accuracy of judgements without a time element, as such research has identified a number of factors that mediate judgement bias. However, given the focus of this thesis, the bulk of the chapter will be devoted to the subject of temporal rather than non-temporal judgements.

The chapter begins with a section that emphasises the prevalence of temporal misestimation and poor task planning in everyday life. In addition to outlining some explanations for temporal misestimation, this section highlights the adverse consequences associated with poor task planning and inaccurate time estimation. The next section focuses on research into the planning fallacy and time estimation bias, with the aim of demonstrating the extent of empirical support for temporal underestimation. Research reviewed in this section suggests that people tend to be over-optimistic when estimating the duration of tasks as diverse as college coursework assignments, anagrams and income tax forms. The prevalence of poor time estimation is further emphasised in the next section, which focuses on the subject of human time perception. In addition to highlighting the extent of poor time perception, this section will consider the cognitive processes that are involved in judging the length of temporal intervals.

In order to demonstrate that bias is not confined to temporal estimates, the issue of inaccuracy in non-temporal judgements is briefly outlined in the next section. This section highlights the prevalence of biases such as overconfidence and provides some explanations for judgement inaccuracy. Given the prevalence of inaccuracy in temporal and non-temporal judgements, the next section focuses on techniques that are intended to debias estimates of task duration and task performance. Whilst some research concerning non-temporal judgements will be considered in this section, the emphasis will be on the success of strategies that are intended to improve the accuracy of judgements of task duration. Research reviewed in this section will demonstrate that techniques have largely been unsuccessful in debiasing time estimates, whereas debiasing strategies have met with greater success in relation to non-temporal judgements.

In the next section, the cognitive heuristics and biases that contaminate the judgement and decision making process are considered. The research presented in this section demonstrates that using cognitive heuristics not only leads to judgement bias, but also tends to be characterised by a neglect of relevant information such as base rate data. This section will also show that the anchoring and adjustment cognitive heuristics are applicable to judgments of task duration. The final section of the literature review focuses on factors that have been found to influence bias and accuracy in non-temporal judgements. Research presented in this section demonstrates that factors such as task complexity and task experience have been found to affect judgement bias, and could thus be important determinants of time estimation accuracy.

1.2 Background

Estimating how much time an activity will take to complete is one of the most common judgements made by humans. In the course of daily life people are required to estimate the duration of a variety of tasks such as crossing the road safely (Zakay & Block, 1996), arranging a dinner party (Byrne, 1977) and shopping for groceries (Hayes-Roth & Hayes-Roth, 1979). Making accurate estimates of personal task performance is of considerable importance in terms of the enhancement of feelings of self-efficacy and well-being (Armor & Taylor, 2002; Taylor & Brown, 1988), and personal and professional achievement. Similarly, it is necessary to schedule activities effectively if people are to fulfil work and personal obligations.

The existence of a profitable self-help industry (Macan, Shahani, Dipboye & Phillips, 1990) highlights the benefits of effective time management (e.g., stress reduction), which are also well supported empirically (Burt, 1993; Burt & Kemp, 1994; Francis-Smythe & Robertson, 1999; Shahani, Weiner & Streit, 1993). Research also emphasises the prevalence of procrastination (e.g., Lay, 1986; Schouwenberg, 1992), which has been observed in relation to task planning and time estimation (Pychyl, Morin & Salmon, 2000). Hence, it seems that people do not always begin tasks as planned. However, when tasks are commenced, research suggests that estimates of task duration are often over-optimistic (e.g., Koole & Van't Spijker, 2000), with actual task duration exceeding the amount of time predicted. Although judgements of task completion times are an important part of everyday life, it seems that the ability to make accurate temporal estimates is far from universal.

The prevalence of temporal misestimation is highlighted by the late completion of several high profile projects, such as the construction of the Channel Tunnel. The late completion of such projects is not only characterised by massive delays and overspend, but is also likely to have adverse consequences for those organisations involved, such as a loss of profits (Markham, 1997). However, in the workplace, it may be beneficial to underestimate the duration of a project in order to secure a lucrative contract via competitive tendering systems. For example, within the private and public sectors, contracts to undertake construction projects are often awarded on the basis of the lowest cost and the shortest time scale for completion (Markham, 1997).

Although it may be beneficial to underestimate task duration, it is becoming common practice for the legal contracts that accompany large scale organisational projects to contain penalty clauses for late completion. In such instances, the organisation undertaking the project will be penalised financially if the project is not completed on schedule (Grundy & Brown, 2002). Hence, the potential financial benefits of underestimating the duration of large scale organisational projects must be evaluated against the adverse consequences associated with late task completion.

There is empirical evidence of temporal misestimation and poor task planning within organisations (Ariely, Loewenstein & Kahneman, 2000; Das & Teng, 1999; Kahneman & Lovallo, 1993; Klayman & Schoemaker, 1993). Research by Kidd (1970) revealed that professional engineers were over-optimistic about the amount of time needed to overhaul electricity generating equipment. That is, the actual duration of equipment maintenance projects was longer than the amount of time that the engineers had planned for such overhauls. However, such findings are not confined to organisations, as poor task planning has also been observed at the level of the individual. Considerable research has found that plans for task completion are often over-optimistic, that is, people predict that they will finish more tasks than they actually complete within a given period of time (Goldin & Hayes-Roth, 1980; Hayes-Roth & Hayes-Roth, 1979; Smith, 1996).

Research by Hayes-Roth (1980) revealed that the number of errands that participants predicted they would complete within a certain period of time was less than the number that were actually finished. Such over-optimism is attributed to a tendency to focus attention on higher-level aspects of the planning process at the expense of lower-level aspects (Hayes-Roth, 1980). Hayes-Roth suggests that higher-level aspects include deciding the order in which the errands are to be performed, whereas lower-level aspects include factors that might delay task completion. For example, a higher-level aspect of the planning process would be going shopping for groceries followed by a trip to the gymnasium, whereas a lower-level aspect would be the delay caused by traffic congestion en route to the gymnasium. In addition to overestimating the number of errands that could be accomplished, participants in Hayes-Roth's (1980) research tended to underestimate the duration of each errand, suggesting that poor planning was accompanied by temporal misestimation.

Time estimation has been the subject of considerable research (e.g., Byram, (1997; Josephs & Hahn, 1995; McClain, 1983; Zakay, 1989), with studies revealing evidence of poor judgement accuracy. There are two distinct areas of time estimation research, one of which concerns estimates of task completion times whilst the other concerns the perception of temporal intervals. Research into time perception has found that estimates of temporal intervals where no task is performed are generally inaccurate, suggesting that people tend to misperceive the passage of time (Block, 1989; Hawkins & Tedford, 1976; Poynter, 1989; Zakay & Block, 1997; Zakay & Fallach, 1984). For example, Aschoff (1985) found that participants were inaccurate when pressing a buzzer to signal the end of temporal intervals that ranged in duration from a few seconds to two hours.

Research concerning estimates of task completion times has also found evidence of judgement inaccuracy, with many studies revealing that people tend to underestimate task duration. For example, Buehler, Griffin and Ross (1994) found that student participants underestimated the duration of real world tasks such as college coursework assignments. Such research provides support for the existence of the planning fallacy and the prevalence of temporal underestimation. Research into task completion time estimation will now be considered.

1.3 Task Completion Time Estimation

Research in this area has generally focused on the planning fallacy (e.g., Buehler et al., 1994), with an optimistic time prediction bias being observed on a variety of tasks. The planning fallacy was identified by Kahneman and Tversky (1979) in their research into the cognitive heuristics and biases that contaminate experts' judgements. They found that experts such as stockbrokers underestimated their task completion times despite possessing information about the late completion of previous similar activities. Kahneman and Tversky distinguish between two kinds of data that are available to people when planning tasks: information such as previous task performance (i.e., distributional information); and information concerning the task at hand (i.e., singular information).

An aspect of singular information is the amount of work involved in completing a current task, whereas the duration of previous similar activities is an aspect of distributional information. Kahneman and Tversky suggest that the planning fallacy is a consequence of heuristic information processing whereby case-specific or singular information becomes the focus of attention at the expense of distributional information, which is neglected. Hence, the planning fallacy occurs because the current task is treated as a unique event, which is dissociated from previous similar activities (Buehler et al., 1994).

Although Kahneman and Tversky's (1979) research concerned experts' judgements, there is considerable evidence that non-experts also tend to make optimistically biased predictions of task duration (Buehler, Griffin & MacDonald, 1997; Byram, 1997; Griffin & Buehler, 1999; Koole & Van't Spijker, 2000). Using real world tasks that ranged in duration from a number of days to several weeks, Buehler et al. (1994) found that participants tended to underestimate their completion times. For example, student participants underestimated the duration of various college coursework assignments including their final year dissertation. Analyses of verbal protocols revealed that participants tended to construct detailed mental scenarios of how a current task would proceed without impediments, suggesting that time predictions were based on singular information.

Buehler et al. suggest that people tend to neglect information about past experiences because the process of predicting and planning naturally involves looking forward. However, Buehler et al. also found that temporal underestimation was reduced when participants were instructed to identify similarities between the current task and previous similar activities. They suggest that this focus of attention manipulation forced participants to take account of distributional information (e.g., previous task duration), which was incorporated into their time estimates, leading to a reduction in over-optimism.

Temporal underestimation has also been observed on laboratory tasks that are of shorter duration (i.e., less than two hours) than the real world tasks used by Buehler et al. (1994). Research by Josephs and Hahn (1995) revealed that student participants tended to underestimate the duration of tasks such as solving anagrams and proof-reading a manuscript. Likewise, Byram (1997) found evidence of temporal underestimation on laboratory tasks as diverse as building self-assembly furniture and folding paper (i.e., origami).

The role of motivational incentives was also investigated by Byram, who found that temporal underestimation was exacerbated when monetary rewards were contingent on the speed of task completion. That is, estimates were more over-optimistic among participants who were offered a financial bonus for completing an origami task within a pre-specified period of time. However, Byram also found that time estimates were lower in the speed incentive condition, whereas actual task completion times were not. Byram explains this finding by suggesting that the monetary incentive motivated participants to engage in wishful thinking (Price, 2000).

The interplay between cognition and motivation was investigated by Buehler et al. (1997), who also found that monetary incentives influenced time estimation bias. For example, individuals who expected to receive an income tax refund underestimated the duration of completing their income tax form to a greater extent than people who did not anticipate a refund. In a laboratory-based study, Buehler et al. found that participants' motivation for completing an anagram task influenced the direction in which their time estimates were biased. That is, over-optimism was evident when monetary incentives were contingent on the speed of task completion (speed incentive), whereas judgements were pessimistically biased when incentives were dependent on the accuracy of temporal estimates (accuracy incentive).

Consistent with the notion of motivated reasoning (Kunda, 1990), Buehler et al. suggest that the performance-contingent incentives affected the kind of reasoning strategy that participants adopted. Analyses of verbal protocols revealed that the speed incentive motivated participants to use a goal-oriented reasoning strategy in which attention was focused on information concerning successful task completion. For example, the number of vowels in each anagram became the focus of attention at the expense of other relevant information such as the person's ability to solve word puzzles. Conversely, Buehler et al. found that participants in the accuracy incentive condition used a more deliberative reasoning strategy in which information such as potential impediments to successful task completion was considered and incorporated into temporal estimates.

Interestingly, whilst Buehler et al. found that over-optimism was eliminated in the accuracy incentive condition, time estimation bias was not reduced in absolute terms. That is, the magnitude of the pessimistic bias associated with the accuracy incentive condition was similar to the degree of over-optimism observed in the speed incentive condition. Although Buehler et al. offer no explanation for this finding, their research indicates that motivational incentives can influence the direction, if not the magnitude, of time estimation bias.

Given Buehler et al.'s (1997) research, it seems that motivational incentives for accurately estimating task duration can eliminate the temporal underestimation indicative of the planning fallacy. However, there is considerable research evidence of the planning fallacy in the temporal judgements of experts and non-experts, suggesting that it is a robust cognitive phenomenon, which is applicable to various long (Buehler et al., 1994) and short duration tasks (Byram, 1997). Thus, it seems that experts and non-experts alike tend to be inaccurate when making predictions of task duration. In a similar vein, there is a sizeable body of research which suggests that people are poor at judging or perceiving the passage of time (e.g., Curton & Lordahl, 1974; McClain, 1979; Poynter, 1989). That is, people are generally inaccurate when judging the length of temporal intervals. Research into human time perception will now be considered.

1.4 Time Perception

The major areas of time perception that have been studied are the psychophysical aspects of temporal stimuli (Allan, 1979) and the cognitive processes involved in estimating temporal intervals (Block & Zakay, 1997; Kawamura, 2000; Zakay, Nitzan & Glicksohn, 1983). A number of contextual factors have also been studied including time estimation paradigm (Zakay, 1989) and the method of estimation (McConchie & Rutschmann, 1971).

The most widely used methods of time estimation are the verbal estimation of an impending or just-completed interval, producing an upcoming interval, or reproducing a just-completed interval. The production and reproduction estimation methods involve experimental participants pressing a buzzer for pre-specified periods of time. There is evidence that time perception is equally poor with each of the three time estimation methods. McConchie and Rutschmann (1971) found that the extent to which participants misestimated the duration of temporal intervals was similar across all estimation methods. Hence, it seems that poor time perception is evident regardless of the manner in which people estimate the length of temporal intervals.

Research into the type of time estimation paradigm has generally focused on differences in the length of estimates given at the end of an interval as a function of receiving some prior warning (Block & Zakay, 1994). Two distinct estimation paradigms have been employed: prospective and retrospective. In the prospective paradigm, participants receive advance warning that an estimate will be required at the end of a temporal interval, whereas no such prior warning is given in the retrospective paradigm. Research by Zakay (1989) revealed that the way in which temporal information is cognitively processed differed according to the presence or absence of an advance warning about a subsequent time estimate. Zakay claims that people possess temporal and non-temporal information processing mechanisms, which compete for cognitive resources.

Zakay (1989) presents a model of time perception that highlights the importance of the presence or absence of an advance warning of a subsequent estimate in explaining the way in which temporal information is cognitively processed. This model suggests that, in the retrospective paradigm, greater cognitive resources are devoted to processing non-temporal information because, during the temporal interval, people are unaware that they will have to estimate time subsequently. Conversely, in the prospective estimation paradigm, greater cognitive resources are devoted to processing temporal information because the prior warning of having to give an estimate subsequently serves as a cue to monitor the passage of time.

Zakay's model also highlights the impact of the type of time estimation paradigm on the kind of information that is held in working memory. On the basis of a series of experiments, Zakay suggests that, in the prospective paradigm, the advance warning means that temporal information is retained in working memory and is used as a basis for estimating time subsequently. Conversely, in the retrospective paradigm, the absence of any prior warning means that temporal information is not the focus of attention and is thus ignored. Zakay suggests that retrospective time estimates are based on non-temporal information such as the amount of work involved in completing a task, which is retained in working memory.

Zakay also found that information processing load interacted with the length of prospective temporal estimates. That is, compared to temporal intervals where no task was performed, prospective estimates were found to be shorter when a proof-reading task was performed during an interval. Zakay suggests that greater cognitive resources are devoted to processing non-temporal information when a task is performed during an interval, leaving fewer resources available to monitor the passage of time. Thus, concurrent task performance means that fewer cognitive resources are available to process temporal information, which results in shorter prospective estimates.

Research concerning the presence or absence of an advance warning about a subsequent time estimate has highlighted the cognitive processes involved in judging temporal intervals (Block & Zakay, 1996). Zakay's (1989) research indicates that performing a task during a temporal interval leaves fewer cognitive resources available to monitor the passage of time, which affects the length of prospective estimates. Extrapolating from the findings of Zakay, it could be that task complexity influences estimates of task duration. That is, given that greater cognitive resources are needed to perform complex tasks (Maynard & Hakel, 1997), fewer attentional resources should be available to monitor the passage of time. Hence, estimates given at the end of a complex task should be based on non-temporal information such as the amount of work involved in task completion.

Research into time perception has demonstrated that people are poor at estimating the duration of temporal intervals. Poor time perception is evident when a task is performed during a temporal interval (Hicks & Brundige, 1974), or when task performance does not occur (Curton & Lordahl, 1974; Michon, 1965). Consistent with research into estimates of task duration (e.g., Byram, 1997), time perception research suggests that people tend to make inaccurate temporal judgements. Given that experts and non-experts make biased estimates of their own task completion times, it is unsurprising that there is considerable evidence of bias in non-temporal judgements of task performance (e.g., Einhorn & Hogarth, 1978). Research in this domain will now be considered in order to provide some explanations for non-temporal judgement bias, and highlight the prevalence of overconfidence.

1.5 Bias in Non-Temporal Judgements

There is evidence of bias or inaccuracy in non-temporal judgements of performance on a variety of tasks (Pulford & Colman, 1997). Considerable research has found that people tend to display an overconfidence bias, that is, judgements of personal task performance are in general greater than actual task performance (Gilovich, Kerr & Medvec, 1993; Sniezek, Paese & Switzer, 1990; Trafimow & Sniezek, 1994). For example, Lichtenstein, Fischhoff and Phillips (1982) report that the number of general knowledge test questions that participants answered correctly was less than the number that they judged to be correct. Judgement overconfidence has been observed on tasks as diverse as answering general knowledge questions (Wright & Wisudha, 1982), memorising text (Winman, 1999) and predicting future events (Griffin, Dunning & Ross, 1990).

This overconfidence bias has been attributed to inadequate cognitive processing of task-related or relevant information, that is, people often use sub-optimal information processing strategies when making judgements of personal task performance (Koriat, Lichtenstein & Fischhoff, 1980). Support for this suggestion comes from Sniezek et al. (1990) who found that overconfidence was reduced when attention was focused on information that was relevant to the task at hand. Sniezek et al. manipulated participants' focus of attention by providing differing numbers of cues on a multiple choice test and found that judgement overconfidence was greater when attention was focused on only one possible answer. Sniezek et al. suggest that focusing attention on all possible answers meant that participants engaged in more thorough information processing, which resulted in greater judgement accuracy.

There is also evidence of judgement inaccuracy in real world settings such as business and commerce (Lawrence & O'Connor, 2000; Loffler, 1998). Research by Cooper, Woo and Dunkelberg (1988) revealed that entrepreneurs were over-optimistic when predicting the chances of success for a new business venture. Cooper et al. suggest that, because entrepreneurs devote considerable financial resources to a new business venture, they are motivated to base their plans on the potential benefits of business success whilst neglecting information such as the high incidence of bankruptcy among new businesses (Busenitz & Murphy, 1996).

In a similar vein, Baron (1998) found that entrepreneurs were over-optimistic when making various business-related decisions including how much work could be completed within a certain period of time. Baron suggests that entrepreneurs are prone to cognitive biases such as overconfidence because they make professional judgements under conditions of uncertainty. For example, when forecasting sales figures for a novel product, entrepreneurs have little or no task-related information upon which the base their judgement.

Judgement inaccuracy is also apparent in the domain of expert judgement where medical decision making has been the focus of considerable research (Elstein, Shulman & Sprafka, 1990; Harries, Evans, Dennis & Dean, 1996). In general, such research has revealed that physicians fail to take account of all relevant information when making professional judgements (Wigton, 1996). For example, Speroff, Connors and Dawson (1990) found that judgements of levels of cardiac functioning differed from estimates that were derived from a statistical modelling procedure. Speroff et al. suggest that these judgements were based on information derived from medical records and physical examinations at the expense of information derived from specialist medical testing procedures, which was overlooked. The findings of such research indicate that judgement inaccuracy occurs because experts such as physicians and entrepreneurs fail to consider all information that is relevant to the task at hand.

There is also evidence that non-experts (i.e., experimental participants) make inaccurate judgements because they too neglect information that is relevant to the task at hand (Sniezek et al., 1990). Hence, it seems that experts and non-experts are often inaccurate when making non-temporal judgements of task performance. Indeed, given the findings of research into the planning fallacy (e.g., Buehler et al., 1994), it seems that people's temporal judgements also tend to be biased. In view of the prevalence of inaccuracy in temporal and non-temporal judgements, it is unsurprising that a number of techniques intended to improve judgement accuracy have been proposed. The success of debiasing techniques in relation to temporal judgements will now be considered. In addition, some techniques that have been successful in debiasing non-temporal judgements of task performance will be specified.

1.6 Debiasing Techniques

Techniques intended to improve the accuracy of task performance judgements have been the focus of considerable research (Balzer, Doherty & O'Connor, 1989; Mumpower & Stewart, 1996; Shanteau, 1992; Shanteau & Stewart, 1992). In the domain of expert judgement, a number of strategies intended to reduce overconfidence in temporal and non-temporal judgements were specified by Kahneman and Tversky (1979). These procedures are based largely on the principles of statistical prediction such as regression towards the mean. In essence, these procedures entail experts accessing or being provided with information about previous forecasts, which is then used as a basis for a current task prediction. Consistent with their research into the planning fallacy, Kahneman and Tversky suggest that using such distributional information leads to more accurate judgements because experts are able to relate the current task to previous similar tasks rather than treat it as a unique event.

The use of base rate data such as personal performance on previous similar tasks has been found to improve the accuracy of judgements of future life events such as getting married and obtaining employment (e.g., Osberg & Shrauger, 1986; Vallone, Griffin, Lin & Ross, 1990). Similarly, using information about previous task performance has met with some success in reducing temporal misestimation (Buehler et al., 2002). As previously mentioned, Buehler et al. (1994) found that over-optimism was attenuated among participants who were instructed to link the current task to their performance on previous similar activities. Specifically, prior to estimating the duration of an upcoming college assignment, student participants were required to describe in detail a plausible scenario of task completion. Buehler et al. found that time estimates were less optimistically biased among participants who generated mental scenarios in which elements that were common to previous and present tasks were identified.

The impact of information about previous similar tasks on time estimation bias was also studied by Byram (1997), who found that over-optimism was not attenuated when participants acquired task experience before predicting the duration of an origami task. However, as the task experience manipulation involved viewing the task instructions on two occasions only, it might have been insufficient as a method for providing task-related information. That is, as participants did not actually perform the origami task beforehand, they would not have possessed information about their previous task performance.

The effects of prior task experience were also studied by Josephs and Hahn (1995), who found that time estimates made after 10 minutes of task performance were less optimistically biased than estimates given beforehand. Josephs and Hahn also found that the accuracy of on-line time estimates interacted with actual task performance, as over-optimism was only reduced among participants who had completed more of a proof-reading task in 10 minutes. Josephs and Hahn suggest that time prediction accuracy was greater among the faster readers because these participants possessed greater task-related information, which was used as a basis for their on-line estimate. The findings of Buehler et al. (1994) and Josephs and Hahn suggest that using information about personal performance on previous tasks can lead to time estimates that are less optimistically biased, and thus more accurate.

Instructing people to generate mental scenarios of task performance is another debiasing technique that has been studied in relation to temporal judgements. This technique essentially involves constructing mental scenarios of task completion that are less over-optimistic (Buehler et al., 2002). Research by Newby-Clark et al. (2000) used a mental scenario generation technique where participants were required to construct optimistic, realistic and pessimistic scenarios about the completion of various real world tasks such as college coursework assignments.

The pessimistic scenario manipulation used by Newby-Clark et al. required participants to think about and document potential pitfalls that could delay task completion such as computer failure. Newby-Clark et al. found that pessimistic scenarios were rated as less plausible than optimistic ones when manipulated between-groups or as a repeated-measure. Specifically, participants reported that pessimistic scenarios were unlikely to actually occur. Pessimistic mental scenarios were also found to be no more accurate than optimistic scenarios, that is, the magnitude of temporal underestimation did not differ according to the type of mental scenario generated. Newby-Clark et al. also found that participants tended to focus attention on optimistic mental scenarios, suggesting that time estimates were based on best-case scenarios, which led to optimistically biased predictions.

A similar debiasing technique was employed by Byram (1997), who instructed participants to think about and document potential pitfalls that might impede the completion of a self-assembly furniture task. However, Byram found that listing factors such as a broken task component did not reduce temporal underestimation. Whilst these findings suggest that thinking about potential pitfalls and setbacks does not reduce temporal underestimation, one study (Taylor & Armor, 1997, as cited in Taylor, Pham, Rifkin & Armor, 1998) has found that mentally simulating the task completion process can attenuate the planning fallacy.

The research reported by Taylor et al. (1998) revealed that nearly half of all participants who received training in a mental simulation technique finished a college coursework assignment as predicted. This debiasing technique entailed thinking about the process of task completion, and carefully visualising all parts of the task. Participants were also required to rehearse this technique several times before task performance commenced. In contrast, fewer than 15 percent of participants who did not use this mental simulation technique completed a college coursework assignment as predicted.

Taylor et al. suggest that mentally simulating the task completion process is beneficial in terms of goal attainment because it gives people the opportunity to consider how potential obstacles to successful task performance can be overcome. For example, the potential obstacle of not being able to obtain a recommended textbook from the college library can be overcome by ordering the book from the library well before the deadline for coursework submission (Taylor et al., 1998). However, as the majority of participants in the mental simulation condition underestimated task duration, this technique reduced but did not eliminate the planning fallacy.

The findings reported by Taylor et al. and those of Byram (1997) and Newby-Clark et al. (2000) provide rather equivocal evidence concerning the effectiveness of such techniques in debiasing time estimates. The limited success of mental scenario generation techniques contrasts with other research where thinking about alternative task outcomes has been effective in improving the accuracy of non-temporal judgements such as the prediction of future life events (Griffin et al., 1990).

The generation of possible alternative outcomes has been found to improve the accuracy of judgements on various tasks including general knowledge questions (Koriat et al., 1980), clinical diagnosis (Arkes, Faust, Guilmette & Hart, 1988), future career prospects (Hoch, 1985) and life events (Dougherty, Gettys & Thomas, 1997; Hoch, 1984; Vallone et al., 1990). These debiasing techniques involve participants having to think about, and often list, reasons contradicting an impending judgement or a judgement that had just been made. The ensuing reduction in overconfidence and improvement in judgement accuracy has principally been attributed to greater cognitive processing of task-related information, which occurs when alternative outcomes are considered and incorporated into judgements (Sniezek et al., 1990).

Research by Koriat et al. (1980) revealed that overconfidence in judgements of general knowledge test questions was reduced among participants who had to choose between multiple-choice answers compared to those who had to generate their own answer. Koriat et al. also found that, relative to participants who produced a single answer each to general knowledge test question, judgement overconfidence was reduced among those individuals who had to generate two answers to each question. Koriat et al. suggest that considering alternative answers eliminates the tendency to seek out information supporting a chosen answer and selectively neglect contradictory evidence. Likewise, Arkes et al. (1988) claim that having to list reasons favouring different medical diagnoses meant that neuropsychologists cognitively processed task-related information in greater depth, which resulted in less overconfidence and greater judgement accuracy.

Whilst there is evidence that generating alternative task outcomes attenuates overconfidence in judgements of future events (Fischhoff & MacGregor, 1982; Koehler, 1991), generating mental scenarios of task completion has met with limited success in relation to time estimation bias (Buehler et al., 2002). It has been suggested that individuals have more of a vested interest in accurately estimating the duration of college coursework assignments than answering general knowledge test questions (Newby-Clark et al., 2000). Hence, the personal relevance of the task could be responsible for the differential success of such techniques. Although this explanation seems intuitively plausible, there is evidence that judgements concerning personally-relevant tasks are less biased when such techniques are used.

Research by Hoch (1985) revealed that generating alternative task outcomes was successful in improving the accuracy of judgements of participants' future career prospects. This finding suggests that the personal relevance of a task cannot explain the differential success of such debiasing techniques. Support for this suggestion derives from the widespread and successful use of multiple scenario generation techniques within business settings (Kuhn & Sniezek, 1996; Schnaars & Topol, 1987; Schoemaker, 1993), where decision makers often have a vested interest in the outcome of judgement tasks (Cooper et al., 1988). Whilst the reasons for the differential success of mental scenario generation techniques are not clear, there is evidence that such debiasing strategies are relatively ineffective in relation to judgements of task duration (Byram, 1997).

Other debiasing techniques have been employed, but have in general been unsuccessful in reducing the temporal underestimation indicative of the planning fallacy. Although research suggests that aggregate or frequentist non-temporal judgements are less biased than probabilistic ones (Griffin & Tversky, 1992; Reeves & Lockhart, 1993; Sniezek & Buckley, 1991; Tversky & Kahneman, 1983), Griffin and Buehler (1999) found that frequency-based time predictions were no less optimistically biased than predictions made in a probabilistic format.

Using a variety of everyday tasks such as shopping, Griffin and Buehler's participants were required to predict the duration of each task and then either estimate how many of the tasks would be completed on time, or judge the probability of each task being completed within the time predicted. Griffin and Buehler found evidence of temporal underestimation regardless of the judgement format. They suggest that, because the tasks were self-selected, participants ignored the statistical nature of the experimental manipulation and focused attention instead on information concerning successful task completion.

Task decomposition has been found to improve the accuracy of non-temporal judgements on tasks such as quantity estimation (MacGregor & Armstrong, 1994; MacGregor, Lichtenstein & Slovic, 1988), and is widely used as an organisational decision making aid (Klayman & Schoemaker, 1993). In relation to temporal judgements, this debiasing technique was employed by Byram (1997) who divided a furniture assembly task into three discrete components according to information presented in a step-by-step instruction booklet. Participants had to predict the duration of each of the components and then provide an estimate of the duration of the whole task. Byram found that this technique was unsuccessful in reducing judgement bias, as predictions for each of the three task components were no less optimistically biased than estimates concerning overall task duration. Whilst Byram offers little in the way of an explanation for this finding, it seems that this task decomposition technique was unsuccessful in reducing time estimation bias.

As can be seen above, research has employed several techniques intended to debias temporal and non-temporal judgements. There is evidence that considering information about previous task performance can improve the accuracy of time predictions (Buehler et al., 1994) and non-temporal judgements (Vallone et al., 1990). Consistent with the work of Kahneman and Tversky (1979), it thus seems that taking account of such distributional information can help to overcome the planning fallacy.

Whilst the use of such distributional information has been shown to be an effective technique for debiasing time estimates, other strategies that have been employed have not in general been successful in improving prediction accuracy. For example, making probabilistic time predictions does not seem to reduce over-optimism (Griffin & Buehler, 1999), a finding that contrasts with other research where this technique has been successful in reducing overconfidence in non-temporal judgements (e.g., Reeves & Lockhart, 1993). Likewise, techniques such as task decomposition and generating alternative task outcomes have met with considerable success in debiasing non-temporal judgements (e.g., Koriat et al., 1980), whereas similar techniques have not been successful in reducing time estimation bias (Byram, 1997; Newby-Clark et al., 2000). Thus, it seems that the majority of techniques employed have been unsuccessful in debiasing estimates of task duration.

Since the planning fallacy is due to an over reliance on singular information and the neglect of distributional information (Kahneman & Tversky, 1979), heuristic information processing is presumably used when estimating task duration (Buehler et al., 2002). That is, people tend not to take account of all pertinent information available to them, and so base their time estimates on incomplete data. Given this suggestion, time estimation bias may well be influenced by judgemental heuristics such as anchoring and adjustment. Since heuristic information processing tends to result in judgement bias (Gilovich, Griffin & Kahneman, 2002), it is important to understand how cognitive heuristics operate in order to fully explore the time estimation process. To this end, research into judgemental heuristics and biases will now be considered.

1.7 Judgemental Heuristics and Biases

Considerable research suggests that people tend to use cognitive heuristics or mental shortcuts when making judgements under conditions of uncertainty (Chapman & Johnson, 1994; 1999; 2002; Jacowitz & Kahneman, 1995; Nisbett & Ross, 1980). In essence, judgements under conditions of uncertainty entail an element of intuition and speculation. For example, a condition of uncertainty exists when a publisher predicts the future sales figures for a new book (Kahneman & Tversky, 1979). In an extensive review of the early work in this domain, Kahneman, Slovic and Tversky (1982) found that judgements which are based on cognitive heuristics tend to result in error or bias. That is, judgements based on heuristic information processing often fail to meet the standards of normative theories such as probability calculus (Tversky & Kahneman, 2002), which results in biases such as overconfidence and the neglect of base rate data.

Cognitive heuristics have been found to influence judgements of performance on tasks as diverse as forecasting financial growth (Kahneman & Lovallo, 1993) and editing text (Switzer & Sniezek, 1991). Judgemental heuristics have also been observed in a variety of settings such as business and commerce, and several academic disciplines including law, medicine and economics (Gilovich et al., 2002). Kahneman et al. (1982) suggest that cognitive heuristics are essentially information processing strategies, which enable people to make judgements in a fast and efficient manner. However, the heuristic nature of such information processing means that judgements are contaminated by error or bias because they are based on limited amounts of information (Gilovich & Griffin, 2002).

Kahneman and Tversky (1973) identified three major cognitive heuristics that influence judgements made under conditions of uncertainty: representativeness; availability; anchoring and adjustment.

The representativeness heuristic refers to the fact that judgements are often based on the degree of similarity between a criterion such as a personality characteristic and membership of a target category such as a vocational occupation (Tversky & Kahneman, 1982b). Kahneman and Tversky (1982b) found that the use of the representativeness heuristic was characterised by the neglect of prior probabilities such as population base rate data, which led to judgement bias.

Kahneman and Tversky (1973) found that, when participants took account of base rate data, representativeness was not used and judgements were consequently less biased. Although the representativeness heuristic has not been studied in relation to time estimation, it is characterised by a neglect of base rate data, which if utilised leads to greater judgement accuracy. Thus, there are similarities between this cognitive heuristic and the planning fallacy in that both are characterised by a neglect of information such as base rate data, which leads to biased judgements.

The availability heuristic occurs because judgements about a target event are often based on the ease with which information about the event is brought to mind. For example, an individual might judge their chances of contracting a particular illness using readily available information such as the number of family members who have suffered from the illness (Tversky & Kahneman, 2002). Kahneman and Tversky (1982a; 1982b) suggest that the ease with which event-consistent information is brought to mind stems from memory retrieval and mental simulation processes. Consistent with this suggestion, Tversky and Kahneman (1973) found that participants who had just read a newspaper article about a house fire tended to overestimate the probability of their involvement in such an incident.

Tversky and Kahneman suggest that such overestimation occurred because information concerning the target event was primed by reading the newspaper article and could easily be brought to mind. Moreover, they suggest that participants mentally simulated the target event and recalled event-consistent information. Such information was then incorporated into participants' judgements at the expense of probabilistic information (e.g., the likelihood of being involved in a house fire), which was neglected.

Whilst the availability heuristic has not been studied in relation to time estimation, it has been suggested as a reason for temporal underestimation. That is, when estimating task duration, people focus attention on information about the current task because it is readily available to them, a strategy that results in an optimistic bias (Byram, 1997). Indeed, there is evidence that over-optimism occurs because time predictions are based on information about how a task will proceed without impediments (Buehler et al., 1994; Newby-Clark et al., 2000). Given this evidence, people may well rely on singular information when estimating task duration because such case-specific data can readily be brought to mind.

The anchoring and adjustment heuristics refer to the undue influence of prior information such as a numerical value on subsequent judgements of a target stimulus or event (George, Duffy & Ahuja, 2000). The use of these heuristics results in judgements being based or anchored on prior information about the target stimulus or event with insufficient adjustment for other relevant information about the target stimulus or event (Carlson, 1990; Schwarz & Wyer, 1985).

A simple but elegant study of the anchoring and adjustment heuristics was conducted by Tversky and Kahneman (1974). This study revealed that estimates of the percentage of African countries in the United Nations were influenced by irrelevant numerical values, which were ostensibly derived at random by spinning a 'wheel of fortune'. Prior to giving a percentage estimate, participants received either a high (65) or a low (10) numerical value from spinning the 'wheel of fortune'. Tversky and Kahneman found that estimates were lower in the low anchor condition, whereas estimates were higher among participants who received the high numerical value. Tversky and Kahneman suggest that judgements were insufficiently adjusted from the numerical values, which thus served as anchors.

The anchoring and adjustment heuristics have been observed on a variety of tasks (Kruger, 1999), and have been used when information that is relevant to the judgement task serves as an anchor. For example, Northcraft and Neale (1987) found that estate agents tended to base their estimates of property prices on previously-received information concerning property prices. Northcraft and Neale suggest that the anchoring and adjustment heuristics are not confined to artificial laboratory tasks, but also lead to biased judgements in real world settings.

In contrast to the standard anchoring paradigm (Chapman & Johnson, 2002), there is evidence that being presented with numerical information is not a pre-requisite for judgement bias to occur as a result of anchoring and adjustment. A series of studies by Epley and Gilovich (2002) revealed that the anchoring and adjustment heuristics were used by participants who had to generate their own numerical answers to general knowledge questions. Epley and Gilovich suggest that bias was a consequence of judgements (i.e., answers) being insufficiently adjusted from self-generated numerical anchor values, which concerned incorrect information about the target questions.

The applicability of the anchoring and adjustment heuristics to judgements of task duration was highlighted by Buehler, MacDonald and Griffin (1994, as cited in Buehler, Griffin & Ross, 1995). Buehler et al. (1995) report that estimates of the duration of a college computer-based assignment were anchored on either a long or a short temporal interval, which was supposedly chosen at random. In actual fact, prior to estimating task duration, student participants were assigned a low (4) or high (17) number by the researchers, which concerned the number of days in which the assignment had to be completed. Buehler et al. (1995) report that time estimates were insufficiently adjusted from these numerical values, with participants who were assigned the low value giving lower predictions than those who were assigned the high value. However, there was no difference in prediction accuracy or the direction of bias (i.e., optimistic or pessimistic) as a function of anchoring, with both groups underestimating task duration to a similar extent.

Since the planning fallacy has been attributed to heuristic information processing (Kahneman & Tversky, 1979), cognitive heuristics can provide potentially important information concerning the reasons for temporal misestimation. For example, as suggested by Byram (1997), the over reliance on singular information when estimating task duration could be explained by the use of the availability heuristic. Moreover, there is evidence that the use of the anchoring and adjustment heuristics leads to biased estimates of task duration (Buehler et al., 1995).

Although it is not known whether the representativeness heuristic is applicable to time estimation, judgements made using this strategy are characterised by a neglect of base rate data, which results in bias. However, when people take account of base rate data, there is evidence that the representativeness heuristic is not used and judgement bias is reduced (Kahneman & Tversky, 1973). Given that the neglect of distributional information is a key determinant of the planning fallacy (Buehler et al., 1994), it seems appropriate to highlight instances where such information has been used to good effect, that is, to reduce judgement bias.

Whilst the use of task-related information has been successful in reducing bias in judgements of task duration (Buehler et al., 2002), few other factors that influence the accuracy of time estimates have been identified. However, a number of factors that mediate bias in non-temporal judgements of task performance have been studied (e.g., McMackin & Slovic, 2000; Payne, Johnson & Bettman, 1993). Given the focus of this thesis, it is necessary to consider such research in order to identify factors that might also mediate the accuracy of temporal judgements. To this end, research concerning factors that influence bias in non-temporal judgements will now be discussed.

1.8 Factors That Influence Bias in Non-Temporal Judgements

Task complexity is one factor that has been found to influence judgement bias (Bénabou & Tirole, 2002; Juslin, Winman & Olsson, 2000; Soll, 1996; Suantak, Bolger & Ferrell, 1996). Considerable research has found that judgement confidence differs according to the level of difficulty of tasks as diverse as answering general knowledge test questions (Bornstein & Zickafoose, 1999; Peterson & Pitz, 1988; Pulford & Colman, 1997; Wright & Wisudha, 1982), memorising text (Winman, 1999) and predicting future events (Fischhoff & MacGregor, 1982). Consistent with the phenomenon known as the difficulty effect (Griffin & Tversky, 1992), such research indicates that task complexity is positively related to judgement confidence. That is, judgements of performance on difficult tasks tend to be overconfident, whereas underconfidence prevails on easier tasks.

Empirical support for the impact of task difficulty on judgement bias comes from Suantak et al. (1996). Suantak et al. categorised several multiple-choice general knowledge test questions according to their level of complexity and asked participants to estimate their degree of confidence in the chosen answers being correct. They found evidence of overconfidence on questions classified as being difficult, whereas underconfidence occurred on test items classified as being easy.

Suantak et al. provide an anchoring and adjustment explanation for this hard-easy effect by suggesting that judgement confidence is insufficiently adjusted according to changes in the difficulty of tasks (or test questions). That is, overconfidence occurs when task complexity increases because judgements are anchored on the perceived lower level of difficulty of the previous task, with insufficient adjustment for the greater complexity of the next task. Conversely, insufficient adjustment leads to underconfidence when task complexity decreases because judgements are anchored on the perceived greater level of difficulty of the previous task.

Performance feedback is another factor that has been found to mediate judgement bias (Bornstein & Zickafoose, 1999; Hoch & Loewenstein, 1989; Sharp, Cutler & Penrod, 1988), with different types of feedback being studied. For example, self-feedback gained through prior task experience has been found to reduce overconfidence on tasks including judging other people's behaviour (Paese & Sniezek, 1991) and answering general knowledge test questions (Pulford & Colman, 1997).

Research has also revealed that outcome feedback can reduce overconfidence thus improving judgement accuracy (Radhakrishnan, Arrow & Sniezek, 1996). Radhakrishnan et al. found that predictions of examination performance were more accurate once the results of previous examinations were available to participants. They propose a model of performance self-evaluation, which integrates cognitive factors such as a biased search of memory, and temporal factors such as the time lapse between task performance and the receipt of feedback. They suggest that outcome feedback will only reduce judgement bias if it is objective and unambiguous, and is presented shortly after a task has been performed.

There is also evidence that temporal factors influence task performance (Ariely & Loewenstein, 2000; Ariely & Zakay, 2001; Das, 1991; Liberman & Trope, 1998; Trope & Liberman, 2000) and judgement bias (Bernsden & van der Pligt, 2001; Savitsky, Medvec, Charlton & Gilovich, 1998). Research by Shepperd, Oullette and Fernandez (1996) revealed that students' judgements of their future career prospects went from being over-optimistic to pessimistic as the time of college graduation drew nearer. Similarly, using tasks such as memory recall and college examinations, Gilovich et al. (1993) found that the temporal interval between task prediction and task performance mediated judgement bias. That is, judgements were overconfident when task performance occurred at a later date, whereas an underconfidence bias prevailed when task performance was imminent.

Gilovich et al. propose that, as task performance approaches, people think about factors that might impede successful task completion such as inadequate examination revision, which results in pessimistically biased judgements. Gilovich et al. also suggest that such defensive pessimism (Norem & Cantor, 1986a; 1986b) is a coping strategy, which is employed to maintain self-esteem, and is exacerbated by the prospect of one's performance being objectively evaluated. Consistent with this suggestion, research has revealed that judgement confidence is reduced when people feel accountable for their decisions (Lerner & Tetlock, 1999; Tetlock, 1985; Tetlock & Boettger, 1994). Hence, judgements could be pessimistically biased when task performance is imminent because people do not wish to feel and appear incompetent, or be held accountable for their poor performance.

The impact of task experience on judgement accuracy and actual task performance has been the focus of considerable research (e.g., Littlepage, Robison & Reddington, 1997; Nembhard, 2000). Various forms of task experience have been studied including task practice or pre-exposure (Johnson & Kanfer, 1992; King, Zechmeister & Shaughnessy, 1980), expertise (Chi, Glaser & Rees, 1982; Lesgold, Glaser, Rubinson, Klopfer, Feltovich & Wang, 1988) and expert-novice differences (Abernathy, Neal & Koning, 1994; Ste-Marie, 1999). In general, task experience has been found to be positively related to judgement accuracy, with experience leading to greater accuracy. Research by Smith and Kida (1991) revealed that, relative to novices, judgements made by experts were less susceptible to cognitive heuristics such as representativeness and biases such as the neglect of base rate data. Smith and Kida suggest that experts rely less on cognitive heuristics than novices because they possess greater task-related information, which forms the basis of judgements that are more accurate.

Evidence of a link between task experience and the anchoring and adjustment heuristics comes from Mussweiler and Strack (2000). They found that anchoring effects were attenuated among participants who had experience of judgement tasks such as general knowledge tests. Mussweiler and Strack suggest that when task experience is minimal or absent, judgements are anchored on irrelevant information such as a numerical value because of a lack of task-related information. Similar effects were observed by Wilson, Houston, Etling and Brekke (1996), who found a relationship between the level of prior task-related knowledge and reliance on numerical anchor values. For example, judgements made by participants who possessed greater knowledge of the number of countries in the United Nations were less influenced by pre-specified numerical anchor values than participants with less knowledge of the judgement task. Hence, such findings indicate that uncertainty about a task increases the reliance on cognitive heuristics, which in turn leads to judgement bias.

The research reviewed in this section highlights the fact that several factors have been found to mediate bias in non-temporal judgements. It has been shown that different types of feedback about task performance can reduce overconfidence in judgements (e.g., Radhakrishnan et al., 1996). Research has also revealed that temporal factors such as the interval between predicted and actual task performance can affect the direction in which judgements are biased (Gilovich et al., 1993; Shepperd et al., 1996). The direction of judgement bias has also been found to differ according to the level of difficulty of various tasks, with overconfidence being evident on more complex tasks and underconfidence prevailing on simpler ones (Suantak et al., 1996). Whilst it is not known whether the difficulty effect (Griffin & Tversky, 1992) is applicable to judgements of task duration, it could be that task complexity influences the length of estimates made at the end of temporal intervals (Zakay, 1989). In addition to emphasising the cognitive processes involved in time estimation, the work of Zakay suggests that task complexity might influence the length of temporal judgements.

The research reviewed in this section also demonstrates that task experience is a factor that mediates bias in non-temporal judgements. There is evidence that task experience reduces the reliance on cognitive heuristics, thus improving judgement accuracy (Smith & Kida, 1991). Given that the planning fallacy occurs because distributional information is overlooked (Ross & Buehler, 2001), it is interesting that having some experience of a task has been found to reduce temporal underestimation (Josephs & Hahn, 1995). This finding suggests that people not only consider information about their previous task performance, but can also use it to make time estimates that are less optimistically biased. Whilst factors such as task experience and task complexity have been found to influence bias in non-temporal judgements, it is notable that mediating factors have received little empirical treatment in relation to time estimation. This issue is addressed in the present programme of research, where a number of factors that might influence the accuracy of people's task completion time estimates are investigated.

1.9 Research Programme

This research programme focuses on a number of potentially important factors that might mediate the accuracy of people's time predictions. Principally, the effects of task complexity, duration and prior experience on time estimation accuracy are examined. In addition, the issues of monetary incentives and the order in which similar and dissimilar tasks are performed are addressed. The research programme comprises a total of nine experiments. The tasks used in all these studies are well structured, that is, they consist of well defined components and must be completed sequentially. In order to accurately measure task duration, tasks were performed within a laboratory environment in all studies.

Since task experience has been found to improve the accuracy of non-temporal judgements (Smith & Kida, 1991), it may also lead to a reduction in time estimation bias. Given the findings of Smith and Kida, it could be that people with task experience do not engage in heuristic information processing when judging task duration. For example, task experience may provide people with some kind of distributional information, which could be incorporated into their time estimates. On the basis of Buehler et al.'s (1994) research, time predictions should be more accurate when information about personal performance on previous tasks is taken into account when estimating task duration.

Given the role of distributional information in determining the planning fallacy (Kahneman & Tversky, 1979), it is notable that the issue of task experience has received little empirical treatment in relation to time estimation bias. Whilst Josephs and Hahn (1995) found that on-line time estimates were more accurate than those made in advance of task performance, previewing task instructions does not seem to reduce time prediction bias (Byram, 1997). However, tasks on which participants had no prior experience of performing have often been employed in previous research. For example, the college thesis used by Buehler et al. (1994) was a task on which participants were unlikely to possess distributional information when predicting its duration. That is, Buehler et al.'s student participants were unlikely to have completed a college thesis prior to the experiment. In such circumstances, the person can only rely on information about the current task when making a time prediction, a strategy that would be expected to result in temporal underestimation (Kahneman & Tversky, 1979).

In order to address the issue of task experience, it is thus appropriate for research to employ tasks on which participants have or can acquire some prior experience. To this end, participants in the present studies performed similar tasks in succession, or experienced the same task before making a time estimate. Such methods ensured that participants had some kind of information about the upcoming task when they estimated its duration.

Whilst temporal underestimation has been observed on tasks as diverse as self-assembly furniture (Byram, 1997), college assignments (Koole & Van't Spijker, 2000), anagrams and income tax forms (Buehler et al., 1997), the issue of task complexity has not been studied in this domain. Given that bias in non-temporal judgements differs according to task complexity (Bornstein & Zickafoose, 1999; Peterson & Pitz, 1988; Suantak, et al., 1996), it could be that task difficulty also mediates time estimation bias. Hence, the difficulty effect (Griffin & Tversky, 1992) may well generalise to judgements of task duration. That is, time predictions might be over-optimistic or overconfident on more complex tasks, whereas a pessimistic bias or underconfidence may prevail on simpler tasks. The existence of the difficulty effect provides a sound rationale for studying the impact of task complexity on time estimation bias. In the present research, the issue of task complexity was addressed by using different versions of the same task, which differed in cognitive complexity.

If research is to accurately measure the extent of time estimation bias, then tasks on which objective measures of duration can be obtained should be employed. Whilst it is laudable that previous research has used real world tasks such as college theses (Buehler et al., 1994) and income tax forms (Buehler et al., 1997), many studies have used self-report measures of task duration. For example, Newby-Clark et al. (2000) contacted participants by telephone to ascertain when they had completed a college coursework assignment.

The use of such self-report measures meant that there was little or no verification of actual task duration. Moreover, as the self-report measures used previously were often obtained some time after task completion, it is possible that participants might have misremembered their completion times. Consistent with this suggestion, there is evidence that retrospective self-reports are characterised by the inaccurate recall of information (Ross, 1989). Hence, in research where task duration was self-reported (e.g., Koole & Van't Spijker, 2000), participants may well have incorrectly recalled their completion times. In order to accurately measure participants' task completion times, laboratory-based tasks were employed throughout the present research programme.

Given the variety of tasks used in previous research (e.g., Buehler et al., 1994; 1997), it is surprising that many of them do not lend themselves to the study of time estimation and task planning. That is, many of the tasks used previously do not consist of sequential components, thus making it difficult for participants to construct a plan before making a time prediction (Hayes-Roth, 1980). Although some research has employed tasks with well defined components that can be performed sequentially (Byram, 1997), several other studies have used tasks that are less well structured. For example, Buehler et al. (1997) used an anagram task, but research by Kelley and Jacoby (1996) suggests that such tasks are ill structured and solved using the intuitive mode of cognition associated with insight problems. Hence, participants in Buehler et al.'s study may not have perceived the components necessary for successful completion when predicting task duration. If, as seems likely with certain tasks, planning is problematic, it is not clear what information people use as a basis for estimating their completion times. In order to facilitate task planning, tasks that are well structured were employed in the present research programme.

The Tower of Hanoi is a classic example of a well structured laboratory-based task, which has been the subject of considerable research into planning processes (Davies, 2000a; 2000b; Goel & Grafman, 1995; Simon, 1975). This task comprises a number of different-sized disks, which must be moved across three vertical pegs to complete a specified pattern using only permitted moves. The rules of the task specify that larger disks must not be placed on top of smaller ones, and that disks can only be moved one at a time. The Tower of Hanoi lends itself to the study of task planning because successful completion involves reaching a goal by performing a series of actions (Karat, 1982). The Tower of Hanoi is also well suited to the manipulation of task complexity because the number of disks that must be moved varies between different versions of the task (Davies, 2000a). Given that the Tower of Hanoi has been used to study task planning, and is a suitable tool for manipulating task complexity, it was employed in the present research programme.

The research presented in the next chapter of this thesis comprises three studies, which investigate a number of factors that might mediate the accuracy of time estimates on the Tower of Hanoi task. Given the findings of Josephs and Hahn (1995), the issue of task experience was the focus of Experiment 1. Using the three-disk version of the Tower of Hanoi, the impact of task practice and pre-exposure on time estimation bias were examined. The issue of task complexity was addressed in Experiment 2, where the more cognitively complex four-disk version was employed alongside the three-disk task. In this study, task experience was also manipulated by allowing some participants to be exposed to the three-disk and four-disk tasks beforehand. In the light of Buehler et al.'s (1997) research, Experiment 3 included a manipulation of monetary incentives in order to further examine the link between motivated reasoning (Kunda, 1990) and time estimation bias.

Chapter Two

The Effects of Prior Experience, Task Complexity and Motivational Incentives on Time Estimation Bias

2.1 Overview

The research presented in this chapter comprises three experiments, which investigate a number of factors that might influence time estimation accuracy. Since performance on previous similar tasks is an aspect of distributional information (Kahneman & Tversky, 1979), task experience may well be an important determinant of time estimation bias. Indeed, it has been shown that time predictions are less optimistically biased when the similarities between current and previous tasks are recognised (Buehler et al., 1994). Likewise, temporal underestimation has been found to be reduced when people have experience of performing a task (Josephs & Hahn, 1995). These findings indicate that time estimation bias is attenuated when people possess some kind of prior task experience. Hence, it seems that such distributional information is not only incorporated into task duration judgements, but can also be used to good effect, that is, to improve prediction accuracy.

The issue of task experience was further addressed in the present research, where different types of task experience were experimentally manipulated. Experiment 1 focused on the issue of prior experience of the three-disk Tower of Hanoi task, whilst prior experience of the more complex four-disk version was explored in Experiment 2. Using two different versions of the Tower of Hanoi meant that the issue of task complexity was also addressed in Experiment 2.

Another aim of the present research was to examine the impact of motivational incentives on time estimation bias. There is evidence that temporal underestimation is exacerbated when monetary rewards are contingent on the speed of task completion (Byram, 1997), suggesting that motivational incentives can induce wishful thinking (Price, 2000). It has also been demonstrated that monetary incentives are linked to the direction in which task duration estimates are biased (Buehler et al., 1997). That is, an optimistic time prediction bias occurs when rewards are dependent on estimation accuracy, whereas a pessimistic bias is evident when rewards are contingent on the speed of task completion.

Given that these findings have been observed on an anagram task (Buehler et al, 1997), it could be that motivational incentives are an important determinant of time estimation bias on laboratory tasks. For example, when faced with a relatively unfamiliar task, the prospect of receiving a monetary reward could motivate participants to engage in wishful thinking, which results in an optimistic prediction bias. The issue of motivational incentives was addressed in Experiment 3, which aimed to confirm the findings of Buehler et al. (1997) concerning the link between motivated reasoning (Kunda, 1990) and the direction of time prediction bias.

2.2 General Introduction to Experiments 1, 2 and 3

Whilst temporal underestimation has been observed on laboratory tasks as diverse as anagrams (Buehler et al., 1997; Josephs & Hahn, 1995) and self-assembly furniture (Byram, 1997), some of the tasks used previously do not comprise well defined components, meaning that task planning may be problematic (Hayes-Roth, 1980). For example, there is evidence that anagram and word puzzle tasks are ill structured and are solved using the intuitive mode of cognition associated with insight problems (Kelley & Jacoby, 1996). Hence, people might not perceive the components necessary for task completion when predicting the duration of such tasks. Although it is not known whether time estimation accuracy differs according to the type of task that is about to be performed, it is appropriate for research to employ tasks on which planning is possible. In order to facilitate task planning, the present research used a well structured task, that is, a task with well defined components.

The Tower of Hanoi is a laboratory-based task that has been the subject of considerable research into planning and cognitive processes (Davies, 2000a; 2000b; Goel & Grafman, 1995; Kotovsky, Hayes & Simon, 1985; Simon, 1975). In its most common form, the task involves moving a number of disks across three vertical pegs in a specified manner. Disk movement is constrained by the rules of the task, which state that disks must be moved individually and cannot be placed on a peg when it is covered by a smaller disk.

A closely related task that has been used in research into planning and cognitive processes is the Tower of London (e.g., Phillips, Wynn, Gilhooly, Della Sala & Logie, 1999; Phillips, Wynn, McPherson & Gilhooly, 2001; Ward & Allport, 1997). Like the Tower of Hanoi, this task involves transferring different numbers of disks between three vertical pegs. Such tasks lend themselves to the study of planning because successful completion involves reaching a goal by performing a series of steps. For example, on the Tower of Hanoi, some kind of planning is likely to be necessary in order to move the disks in accordance with the rules of the task (Karat, 1982). Since the Tower of Hanoi task has been the subject of considerable research into task planning (e.g., Davies, 2000a; Sohn & Gaudiot, 1992), it was employed in the present studies.

The Tower of Hanoi task is also well suited to the manipulation of task complexity, as different versions vary in the number of subgoals that must be created. Subgoals are established whenever disks cannot be placed on the destination peg because such a move would violate the rules of the task (Goel & Grafman, 1995). Thus, as the number of disks increases, task completion involves establishing a greater number of subgoals (Spitz, Minsky & Bessellieu, 1984). Given that there are only three vertical pegs, and disk movement is constrained by the rules of the task, it is necessary to establish more subgoals as the number of disks increases (Simon, 1975). Similarly, as the number of disks increases, people's ability to mentally represent the disk moves necessary for task completion is likely to be constrained by working memory limitations (Davies, 2000a). These findings suggest that different versions of the Tower of Hanoi task differ in cognitive complexity. Thus, it seems appropriate to use this task when studying the impact of task complexity on time estimation bias.

On the basis of previous research (e.g., Byram, 1997; Koole & Van't Spijker, 2000), it could be that time estimates on the Tower of Hanoi will be optimistically biased, but this is not yet known. Indeed, Lesgold (1988) found that, on problems with well defined initial and end states, people were able to form a mental representation of the problem structure and the actions needed for task completion. Thus, on well structured tasks such as the Tower of Hanoi, time estimates may be more accurate because people have a mental representation of what the task entails when predicting task duration. Conversely, on less well structured tasks such as anagrams, time estimates might be more biased as people are unable to construct a mental representation beforehand. A key aim of the present research was to investigate the direction and extent of time estimation bias on the Tower of Hanoi task.

Given the findings of previous research (e.g., Buehler et al., 1997), motivational incentives may well be an important determinant of time estimation bias. Motivational incentives in the form of monetary rewards have been linked to an increase in temporal underestimation on tasks as diverse as origami (Byram, 1997) and income tax forms

(Buehler et al., 1997). It has been proposed that monetary incentives can motivate people to engage in wishful thinking when predicting task duration, which results in greater over-optimism (Byram, 1997). Monetary incentives have also been found to affect whether time estimates are optimistically or pessimistically biased. Buehler et al. (1997) found that, when incentives were dependent on prediction accuracy, a deliberative reasoning strategy was employed, which led to a pessimistic judgement bias. Conversely, when motivational incentives were contingent on the speed of task completion, a goal-directed reasoning strategy was adopted, which resulted in an optimistic judgement bias.

Motivational incentives have also been found to influence the accuracy of non-temporal judgements of task performance (Ashford, 1989; Henry, 1994; Sniezek & Buckley, 1991; Stone & Ziebart, 1995; Wright & Anderson, 1989; Wright & Aboul-Ezz, 1988). Research by Henry and Sniezek (1993) revealed that predictions were over-optimistic when participants were offered a monetary incentive for completing a general knowledge task. Henry and Sniezek suggest that, in the presence of incentives, judgements of performance are distorted to represent the desired level of achievement rather than the level that is attained. Similarly, Henry (1994) suggests that motivational incentives such as money may encourage a sense of unrealistic optimism leading to an overestimation of personal task performance ability. That is, when incentives are offered people tend to predict a higher level of task performance in order to attain a goal that has a favourable outcome.

These findings indicate that motivational incentives can lead to over-optimism in time estimates (e.g., Byram, 1997) and other judgements of task performance (e.g., Henry, 1994). There is also evidence that monetary rewards can motivate people to engage in different reasoning strategies, which in turn affects the direction in which their time estimates are biased (Buehler et al., 1997). Given such findings, a key aim of the present research was to ascertain whether monetary incentives influenced time estimation bias on the Tower of Hanoi task. This issue was addressed in Experiment 3, where monetary incentives that were contingent on either the speed of task completion or the accuracy of time estimates were offered.

Given that distributional information is a important component of the planning fallacy (Buehler et al., 2002), it is notable that the issue of task experience has received little empirical treatment in relation to time estimation. Whilst the neglect of distributional information is a possible cause of the planning fallacy (Kahneman & Tversky, 1979), there is evidence that over-optimism is attenuated when base rate data are incorporated into temporal judgements. Specifically, Buehler et al. (1994) found that a focus of attention manipulation led to information about previous task performance being considered and used to make predictions that were more accurate. Buehler et al.'s research indicates that, when estimating the duration of a current task, people can draw upon their experience of performing similar tasks previously.

A different type of task experience was studied by Josephs and Hahn (1995), who found that time estimates given on-line were less optimistically biased than those made prior to task performance. Josephs and Hahn attribute this finding to the fact that participants possessed greater task-related information when making an on-line time estimate. Whilst this finding suggests that prior task experience can reduce prediction bias, research by Byram (1997) indicates that previewing task instructions does not reduce temporal underestimation. Such equivocal findings contrast with other research where prior task experience or task familiarity has been found improve the accuracy of non-temporal judgements of task performance (e.g., Johnson & Kanfer, 1992).

Research by King et al. (1980) revealed that predictions concerning the recall of memory test items were more accurate when participants were given practice trials of a memory-monitoring task. That is, reading word lists beforehand resulted in greater accuracy when predicting the subsequent recall of the previously-presented words. In a similar vein, prior task experience has been found to improve the accuracy of judgements of task performance by reducing people's reliance on the anchoring and adjustment (Jacowitz & Kahneman, 1995; Mussweiler & Strack, 2000; Wilson et al., 1996) and representativeness cognitive heuristics (Smith & Kida, 1991). These findings suggest that having some experience of a task enables people to make more accurate non-temporal judgements of their task performance.

In the present studies, different forms of task experience were investigated in order to confirm some of the findings of previous research. Given that time predictions are more accurate when the similarities between previous and current tasks are recognised (Buehler et al., 1994), it could be that prior performance of a similar task leads to a reduction in bias on a subsequent task. In order to address this issue, the order in which different versions of the Tower of Hanoi were performed was manipulated in Experiments 2 and 3. As different versions of this task share the same structure (Kotovsky et al., 1985), it was anticipated that performing the first task would provide participants with some kind of information that was relevant to the second task. That is, participants would recognise the similarities between the just-completed task and the upcoming task. By employing different versions of the Tower of Hanoi, the impact of prior performance of a more or less complex version of the same task on time prediction bias was also investigated.

The issue of prior experience of the same task was also addressed in the present research. Following the work of Byram (1997), the impact of previewing a task and its instructions on time estimation bias was further investigated in Experiments 1 and 2. Given the findings of Josephs and Hahn (1995), the impact of task practice on time estimation bias was also examined in Experiment 1. In this study, prior task experience was manipulated by having some participants perform the three-disk Tower of Hanoi beforehand, whilst others studied the task and its instructions before estimating the duration of a target trial.

2.3 Experiment 1

This study investigated whether time predictions were optimistically biased on a well structured laboratory task, the three-disk Tower of Hanoi. Since over-optimism has been observed on several laboratory tasks (Byram, 1997; Buehler et al., 1997; Josephs & Hahn, 1995), it was hypothesised that there would be a general underestimation of task duration. That is, predictions were expected to be shorter than completion times on the three-disk task. In order to confirm some of the findings of previous research (e.g., Smith & Kida, 1991), the issue of prior experience of the same task was also addressed.

Task experience was manipulated by allowing some participants to mentally plan or actively practise performing the three-disk Tower of Hanoi beforehand. Participants in the mental planning condition were able to view the task apparatus and an instruction sheet before estimating task duration. These participants were also asked to mentally plan the task completion process during the pre-exposure period. Given that mentally simulating the task completion process has been found to attenuate the planning fallacy (Armor & Taylor, 1997, as cited in Taylor et al., 1998), it was anticipated that the present mental planning method would result in less temporal underestimation on the three-disk task.

Consistent with the findings of Josephs and Hahn (1995), it was hypothesised that time estimation accuracy would be greater when the same task had been performed beforehand. Hence, temporal underestimation was expected to be reduced among participants who actively practised performing the three-disk task beforehand. There was a total of three levels of task experience: active practice; mental planning; and no experience. Participants in the no experience condition performed an unrelated distracter task (i.e., a word association checklist) before estimating the duration of the three-disk task.

2.4 Method

2.4.1 Participants

Sixty (53 female and 7 male) students at the University of Plymouth participated voluntarily in partial fulfilment of a psychology course requirement. No biographical information other than gender was recorded.

2.4.2 Materials

The three-disk version of the Tower of Hanoi was used. This task consists of three wooden disks of different sizes with a hole in the middle of each one, and a flat rectangular wooden board containing three equidistantly-spaced vertical pegs of equal length. In the starting position, the three disks were stacked in descending order of size on the left-hand peg. The aim of the task was to transfer the disks to the same position on the right-hand peg by moving one disk at a time, but without placing a larger disk on top of a smaller disk. A word association checklist (Berkowitz & Troccoli, 1990) comprising 40 words was also used. A digital stopwatch was used to measure participants' task completion times.

2.4.3 Design and Procedure

A 2 (time: estimated vs. actual task duration) x 3 (task experience: active practice vs. mental planning vs. no experience) mixed factorial design was used. The time factor was a repeated-measure, with participants giving a time estimate and producing an actual task completion time. The task experience factor was manipulated between groups, with participants being randomly assigned to one of the three equal-sized conditions.

Participants were tested individually. Once in the laboratory, they sat at a table and read the briefing sheet, which explained the experimental rationale. Participants then read and signed the participation consent form. No participants withdrew from the experiment or withdrew their data following participation. Participants were then asked to remove their watches and place them out of sight. The task instructions were then presented, and participants were given 20 seconds to read them. The researcher then placed the task apparatus on the table in front of participants and demonstrated the permissible and non-permissible Tower of Hanoi disk moves. There then followed a two-minute period that was different for the three task experience conditions.

Before the task pre-exposure period began, participants in the mental planning condition were informed that they had a short period of time to think about how to complete the task. However, these participants were instructed not to move any disks during the planning period. Prior to the pre-exposure period, participants in the active practice condition were informed that they had a short period of time to practise performing the task. All participants in this condition completed the task at least once during the practice period. After the Tower of Hanoi task apparatus was placed out of sight, participants in the no experience condition were presented with the word association checklist. These participants were informed about the nature of this task and were then given two minutes to perform it.

At the end of the two-minute period, the Tower of Hanoi task apparatus was placed on the table in front of all participants and they were asked to estimate how much time it would take to complete. Participants then began performing the task and the stopwatch was activated at this point. Once the task was completed, the researcher stopped the stopwatch and manually recorded participants' task completion times. Participants were then debriefed about the experimental rationale. Once the debriefing sheet had been read, the researcher answered any questions posed by the participants. Each testing session lasted approximately 15 minutes.

2.5 Results

Basic descriptive statistics of estimated and actual duration per task experience condition are presented in Table 2.1 below. All test statistics used in Experiment 1 are contained in Appendix 1.

	Task Experience Condition		
	Active Practice	Mental Planning	No Experience
Time Estimate	M = 23.90 SD = 11.09 Mdn = 22.50 N = 20	M = 19.20 SD = 14.78 Mdn = 15.00 N = 20	M = 61.65 SD = 42.77 Mdn = 60.00 N = 20
Completion Time	M = 14.80 SD = 56.41 Mdn = 13.00 N = 20	M = 12.25 SD = 5.48 Mdn = 10.00 N = 20	M = 21.70 SD = 14.15 Mdn = 17.50 N = 20

Table 2.1

Time estimates and completion times per task experience condition (in seconds)

As Table 2.1 shows, estimates exceeded completion times in all task experience conditions, suggesting that participants overestimated the duration of the three-disk Tower of Hanoi. Hence, there was no evidence of the temporal underestimation indicative of the planning fallacy on this task. The standard deviations in Table 2.1 indicated that there was considerable variability within the data. Indeed, the Levene test statistics from the estimated and actual completion time data were significant ($p < .05$), suggesting that the parametric statistical assumption of homogeneity of variance had not been met. The majority of histograms concerning these data were positively skewed, and the majority of normality test statistics were significant ($ps < .05$). These findings indicated that the parametric assumption of normality of distributions was also violated.

The prediction and completion time data were thus subjected to a logarithmic transformation, which improved the normality of distributions. The majority of histograms were normally distributed, and the majority of normality test statistics were non-significant ($ps > .05$). Homogeneity of variance also improved as a result of this transformation, as the Levene test statistic from the completion time data was not significant ($p > .05$). Since the logarithmic transformation improved homogeneity of variance and the normality of distributions, these data were statistically analysed. For ease of interpretation, all means reported in the text in all experiments in this thesis pertain to the untransformed data. A bar graph of predicted and actual completion times per task experience condition is presented in Figure 2.1 below.

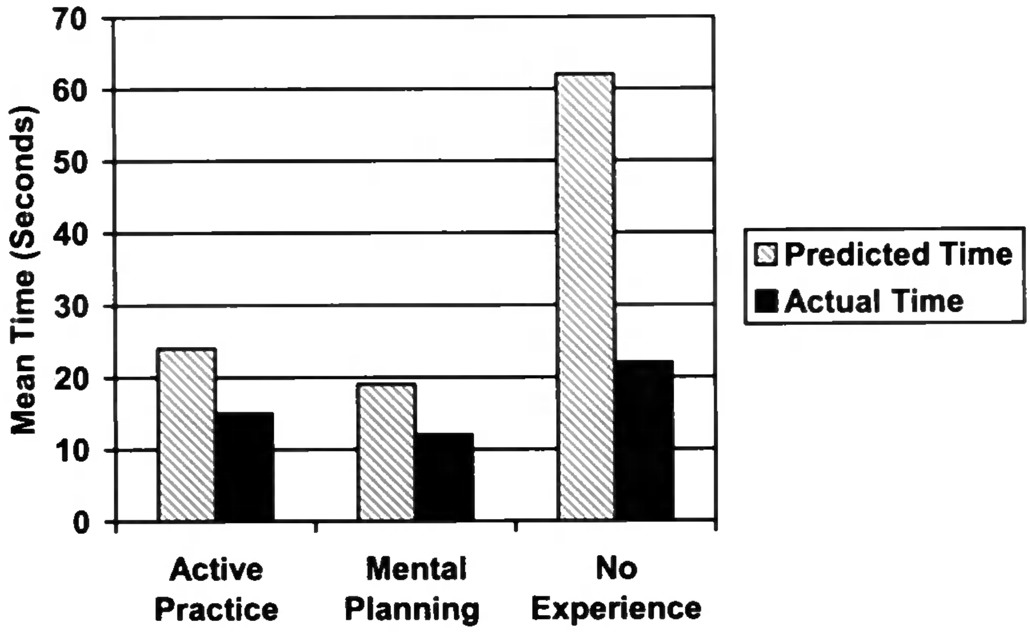

Figure 2.1

Bar graph of predicted and actual completion times per task experience condition

In order to determine how many participants predicted the duration of the three-disk task using whole minutes, the frequency distribution of time estimates was calculated. A histogram of time estimates is presented below in Figure 2.2.

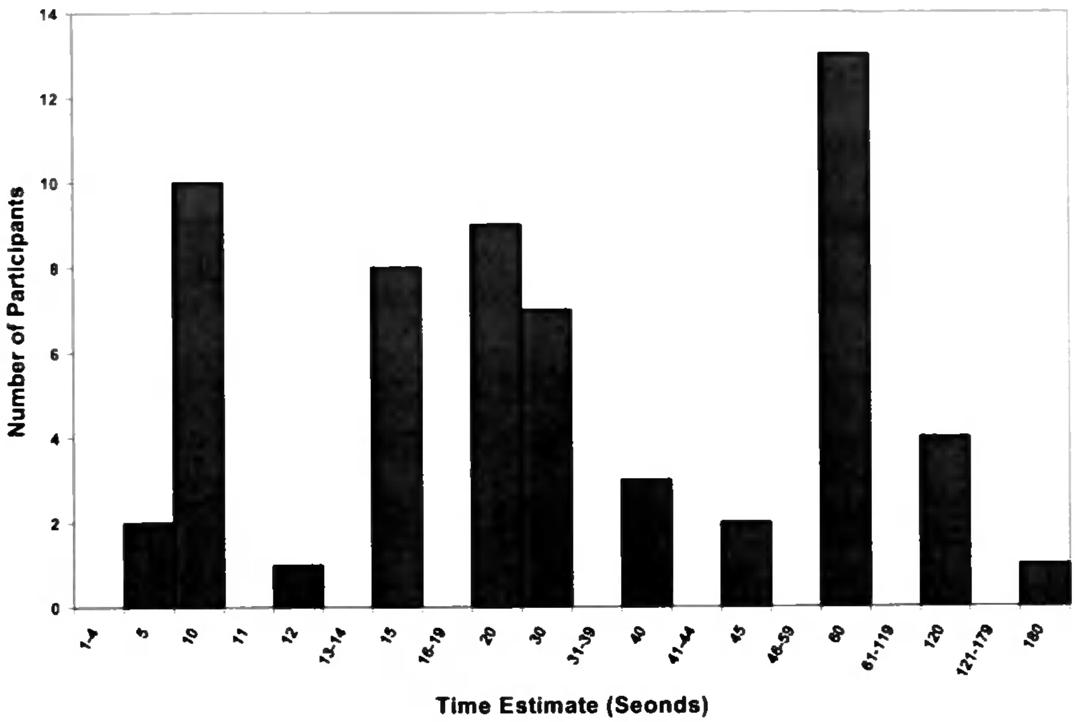

Figure 2.2

Histogram of time estimates

The log-transformed data were subjected to a 2 (time) x 3 (task experience) mixed design factorial analysis of variance (ANOVA). The overestimation of task duration was confirmed by a main effect on the time factor, $F(1,57) = 38.37$, $MSE = .24$, $p < .001$, which revealed that estimates exceeded completion times (Ms = 34.92 and 16.25 seconds, respectively). There was also a main effect of task experience, $F(2,57) = 13.11$, $MSE = .51$, $p < .001$, with overall time being longer in the no experience condition. Scheffé pairwise comparisons revealed significant mean differences between the no experience condition and the other conditions ($ps < .05$). No other condition mean differences were significant ($ps > .10$).

There was also a significant interaction, $F(2,57) = 5.39$, $MSE = .24$, $p < .01$, which revealed that the difference between estimated and actual duration was greatest in the no experience condition (see Figure 2.1 above). Planned comparisons (LSD t-tests, all two-tailed) revealed a significant difference between estimated and actual duration in the no experience and active practice conditions ($ps < .05$). However, there was no significant difference between estimated and actual duration in the mental planning condition ($p > .10$). This finding suggests that time prediction accuracy was greatest among participants who viewed the task and its instructions beforehand.

2.6 Discussion

There was no evidence that time predictions on the three-disk version of the Tower of Hanoi task were optimistically biased. This finding contrasts with research supporting the planning fallacy (e.g., Buehler et al., 1994; Newby-Clark et al., 2000), which has found evidence of an optimistic time prediction bias on various other tasks. In the present study, a pessimistic time prediction bias was observed on the three-disk task in all prior experience conditions. A plausible explanation for the absence of temporal underestimation is that a relatively simple task was used here.

The three-disk version of the Tower of Hanoi task can be correctly completed by making a minimum of only seven disk moves, and took approximately 20 seconds to finish. Hence, it was not only a simple task, but was also of much shorter duration than the laboratory tasks used previously. For example, the anagram task used by Buehler et al. (1997) took about eight minutes to complete, whereas the duration of the laboratory tasks used by Byram (1997) ranged from 10 minutes to over one hour. Given that such tasks took more time, they presumably comprised more components and completion may have required greater cognitive effort. Thus, participants may have been over-optimistic when estimating the duration of such tasks because of cognitive processing limitations.

Empirical support for this notion comes from Sniezek et al. (1990), who found that judgement overconfidence was reduced when participants were encouraged to consider all possible alternative answers to complex multiple choice questions, thus increasing the amount of information being cognitively processed. Sniezek et al. suggest that judgement overconfidence is positively related to task complexity, and occurs because people are unable or unwilling to devote sufficient cognitive resources to predicting their performance on complex tasks.

Such research lends credence to the notion that the greater complexity of the tasks used in previous studies (e.g., Byram, 1997) could be responsible for the temporal underestimation that was observed in such research. That is, on tasks that are more complex than the three-disk Tower of Hanoi, people may not be able to mentally represent the task completion process at the outset. Similarly, the existence of the difficulty effect (Griffin & Tversky, 1992) highlights the role of task complexity in determining bias in non-temporal judgements. If an optimistic judgement bias is only evident on more complex tasks than the present one, then the difficulty effect might be applicable to predictions of task duration. In order to ascertain whether over-optimism occurs on a more complex task, the four-disk version of the Tower of Hanoi task was employed in Experiments 2 and 3.

The frequency distribution of time estimates revealed that almost one-third of the participants used whole minutes when predicting task duration. For example, 13 individuals estimated that the three-disk task would take them one minute to complete. Given that the task itself took about 20 seconds to finish, temporal overestimation would be expected as a consequence of using longer units of time such a one, two or three minutes. A more fine-grained analysis of the issue of whole minutes and longer temporal units being used to judge task duration will be discussed later in this thesis.

Despite the lack of temporal underestimation, time prediction accuracy was found to be greater when participants had some prior task experience. There was evidence that actively practising or mentally planning how to complete the three-disk task resulted in time estimates that were less pessimistically biased. That is, relative to participants who performed the word association task beforehand, participants who had some experience of the three-disk task overestimated their completion times to a lesser extent. Presumably, prior task experience gave these participants the opportunity to mentally represent the correct solution of the task during the pre-exposure period (Lesgold, 1988), which enabled them to estimate its duration more accurately. Moreover, the pre-exposure period might have given these participants time to consider factors that could delay optimal task performance such as incorrectly moving a disk.

Given that the difference between actual and estimated duration was lowest in the mental planning condition, thinking about the task completion process seemed to improve time prediction accuracy. This finding is broadly in line with the research of Armor and Taylor (1997, as cited in Taylor et al., 1998), which revealed that mentally simulating the steps involved in writing an essay reduced the planning fallacy. Consistent with the work of Josephs and Hahn (1995), the present research also emphasises the role of prior task performance in improving time prediction accuracy. That is, temporal overestimation was reduced among participants who performed the three-disk task at least once during the pre-exposure period. This finding suggests that practising the task beforehand enabled participants to acquire pertinent information such as the number of disk moves, which was used to good effect.

Two potentially important findings have emerged from this study. Firstly, there was no evidence of temporal underestimation on the three-disk version of the Tower of Hanoi task. In fact, a pessimistic time prediction bias was evident on this simple well structured task. Secondly, prior task experience resulted in greater time prediction accuracy. That is, the pessimistic time prediction bias was greatest among participants who performed an unrelated task before the target trial. Since time estimation accuracy was greatest in the mental planning condition, this method of task experience was used again in Experiment 2. This study sought to ascertain whether prior experience attenuated time prediction bias on the more complex four-disk version of the Tower of Hanoi task. The impact of task complexity on time prediction bias was also investigated in Experiment 2, where the three-disk and four-disk tasks were performed consecutively.

2.7 Experiment 2

The aims of this study were twofold: to investigate the impact of task complexity on time prediction bias; and to further address the issue of prior task experience. Given the findings of Experiment 1, it was decided to ascertain whether time prediction accuracy was greater when participants mentally planned how to complete the three-disk and four-disk versions of the Tower of Hanoi task. It was suggested that the pre-exposure period in Experiment 1 enabled participants to mentally represent the disk moves needed to complete the three-disk task, which resulted in greater time prediction accuracy.

Whilst this explanation seems feasible given the relative simplicity of the three-disk task, it is not known whether any such mental representation will be possible on the more complex four-disk task. Thus, it is not clear whether the mental planning method from Experiment 1 will reduce time estimation bias on the four-disk task. Indeed, it has been suggested that cognitive processing limitations might preclude mental preplanning as the cognitive complexity of the Tower of Hanoi task increases (Davies, 2000a).

Research by Davies (2000a) revealed that participants did not preplan their disk moves on the five-disk version of the Tower of Hanoi task. Davies suggests that people engage in concurrent planning on this task (and more complex versions of the Tower of Hanoi task) because the number of disk moves required for completion exceeds working memory capacity. In a similar vein, Phillips et al. (1999) found that less time was devoted to mental preplanning on the Tower of London task when memory tasks such as random number generation were performed simultaneously. In other research, Phillips et al. (2001) found that participants could mentally preplan a maximum of seven disk moves on this task. Phillips et al. (2001) suggest that on-line planning strategies place fewer demands on working memory processes and may thus be used on this task.

Given the structural similarity of the Tower of London and Tower of Hanoi tasks (Welsh, Satterlee-Cartmell & Stine, 1999), the findings of Phillips et al. (1999; 2001) may well be applicable to both tasks. Indeed, the work of Davies (2000a) indicates that mental preplanning is unlikely to occur on versions of the Tower of Hanoi that require greater cognitive resources. It could be that people are unable to mentally represent the disk moves on tasks that are more complex than the three-disk version of this task because of working memory limitations.

On the basis of previous research (e.g., Phillips et al., 1999), people may use on-line planning strategies as task complexity increases. Extrapolating from such research, mental preplanning might not reduce time estimation bias on versions of the Tower of Hanoi that are more complex than the three-disk task. That is, when the number of disk moves exceeds working memory capacity, a mental representation of the task completion process may not be possible. Hence, people might be unable to accurately predict the duration of more complex versions of this task because of a lack of task-related information.

In the present study, task experience was manipulated by allowing some participants to preview each task and its instructions for two minutes, whereas others performed a distracter task beforehand. Consistent with the findings of Experiment 1, it was anticipated that time estimates on the three-disk task would be more accurate in the mental planning condition. Whilst mental preplanning is not evident on the five-disk version of the Tower of Hanoi task (Davies, 2000a), it is not known whether it will occur on the less complex four-disk task. Given the findings of Davies, no prediction was made about the impact of prior task experience on time estimation bias on the four-disk task. In order to determine whether prior experience of performing a more or less complex similar task influenced time prediction bias, the order in which the three-disk and four-disk tasks were presented was also manipulated.

By employing the three-disk and four-disk tasks, the impact of task complexity on time prediction bias was examined. Given that task complexity has been found to mediate bias in non-temporal judgements (e.g., Soll, 1996; Suantak et al., 1996), time estimation bias might also differ according to task complexity. Since non-temporal judgements on complex tasks tend to be optimistically biased (Griffin & Tversky, 1992), time estimates might be pessimistically biased only on simple tasks such as the three-disk version of the Tower of Hanoi.

Pilot testing revealed that there was more than a threefold increase in duration on the four-disk task relative to the three-disk task (Ms = 70.56 and 20.25 seconds, respectively). This finding suggests that different versions of the Tower of Hanoi task differ in duration as well as cognitive complexity. Given the findings of Experiment 1, it was hypothesised that temporal overestimation would be evident on the three-disk task. However, as non-temporal judgements are over-optimistic on complex tasks (Suantak et al., 1996), it was anticipated that temporal underestimation would be evident on the four-disk task.

2.8 Method

2.8.1 Participants

Sixty-six (53 female and 13 male) students at the University of Plymouth participated voluntarily. Forty-seven participants took part in partial fulfilment of a psychology course requirement whilst the remainder were paid £1.50 each. No biographical information other than gender was recorded.

2.8.2 Materials

A wooden Tower of Hanoi apparatus containing four different-sized disks was used. The largest disk was removed to form the three-disk task. The rules of both tasks were identical to those which applied to the three-disk task from Experiment 1. Two word association checklists (Berkowitz & Troccoli, 1990) each comprising a different set of 40 words were used. A digital stopwatch was used to measure task duration.

2.8.3 Design and Procedure

A 2 (time: estimated vs. actual) x 2 (task experience: mental planning vs. no experience) x 2 (task: three-disk vs. four-disk) x 2 (task performance order: three-disk first vs. four-disk first) mixed factorial design was used. The time and task factors were repeated-measures. The order in which the tasks were performed was fully counterbalanced. The task performance order and task experience factors were manipulated between groups, with participants being randomly assigned to one of four conditions.

Except for the following differences, the procedure was identical to that of Experiment 1. The two-minute mental planning procedure was applicable to both the three-disk and four-disk tasks. After being shown the relevant task apparatus and reading the instructions, participants in the no experience condition performed one of the two word association tasks for two minutes. The other word association task was performed for two minutes before the second experimental trial. Once each two-minute period had elapsed, the relevant Tower of Hanoi apparatus and its instruction sheet was placed in front of all participants, who were then asked to give a time estimate before performing the task. This procedure was applicable to both experimental trials.

2.9 Results

Basic descriptive statistics of estimated and actual duration per task performance order and prior experience conditions on the three-disk and four-disk tasks are presented below in Tables 2.2 and 2.3, respectively. All test statistics used in Experiment 2 are contained in Appendix 2.

Task Performance Order		Task Experience Condition	
		Mental Planning	No Experience
Three-disk First	Time Estimate	M = 25.94 Mdn = 15.00 SD = 29.20 N = 33	M = 36.50 Mdn = 40.00 SD = 22.34 N = 33
	Completion time	M = 14.94 Mdn = 11.00 SD =11.17 N = 33	M = 18.38 Mdn = 16.50 SD =8.16 N = 33
Three-disk Second	Time Estimate	M = 20.19 Mdn = 17.50 SD = 11.79 N = 33	M = 48.76 Mdn = 50.00 SD = 33.96 N = 33
	Completion Time	M = 9.31 Mdn = 8.50 SD = 2.39 N = 33	M = 14.47 Mdn = 13.00 SD = 5.72 N = 33

Table 2.2

Time estimates and completion times on the three-disk task per performance order and experience condition (in seconds)

Task Performance Order		Task Experience Condition	
		Mental Planning	No Experience
Four-disk First	Time Estimate	M = 60.63 Mdn = 40.00 SD = 48.30 N = 33	M = 142.06 Mdn = 150.00 SD = 79.71 N = 33
	Completion Time	M = 83.44 Mdn = 75.00 SD = 50.49 N = 33	M = 90.94 Mdn = 75.00 SD = 55.87 N = 33
Four-disk Second	Time Estimate	M = 50.00 Mdn = 30.00 SD = 40.89 N = 33	M = 57.75 Mdn = 60.00 SD = 42.84 N = 33
	Completion Time	M = 53.59 Mdn = 51.00 SD = 31.85 N = 33	M = 61.94 Mdn = 63.00 SD = 25.03 N = 33

Table 2.3

Time estimates and completion times on the four-disk task per performance order and experience condition (in seconds)

Table 2.2 shows that estimates exceeded completion times in all cells on the three-disk task. This finding is consistent with that of Experiment 1, where there was evidence of a pessimistic prediction bias on this task. However, there was some evidence of an optimistic prediction bias on the four-disk task, as completion times exceeded estimates in three of the four cells in Table 2.3. Bar graphs of predicted and actual completion times on the three-disk and four-disk tasks are presented below in Figures 2.3 and 2.4, respectively.

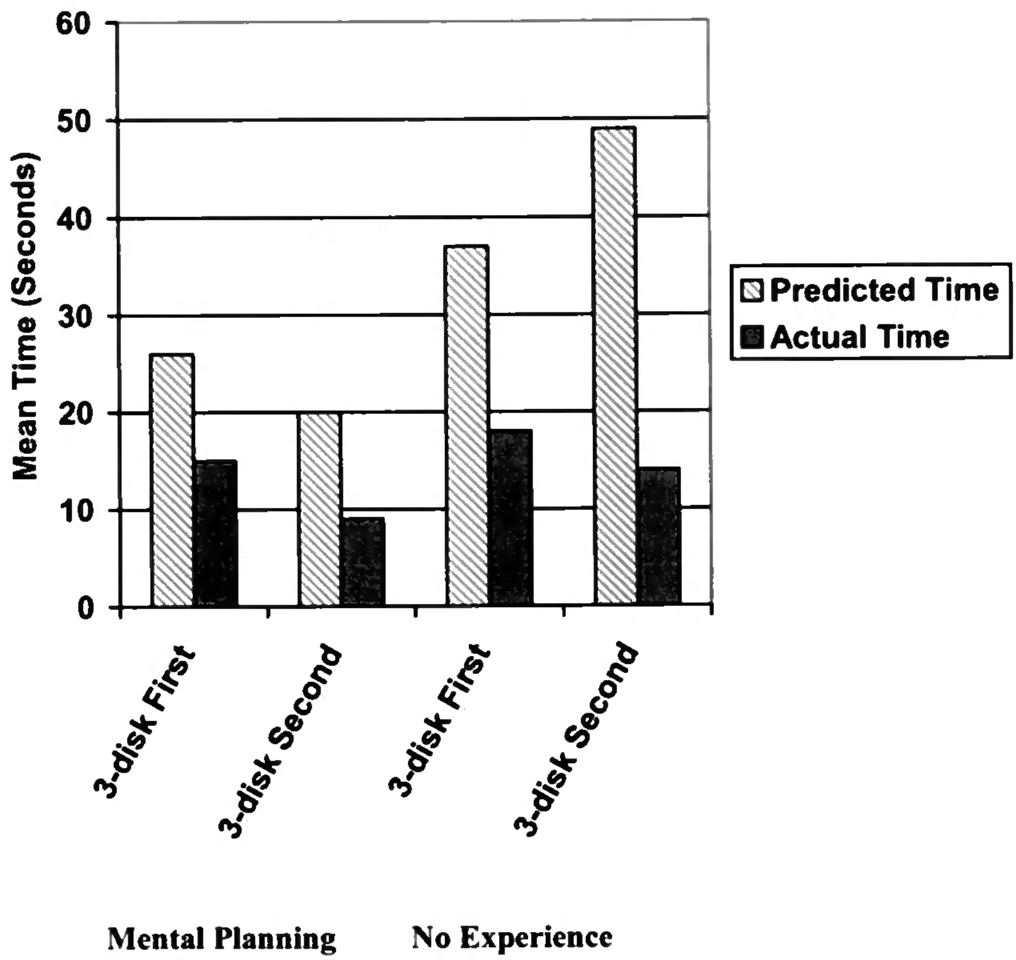

Figure 2.3

Bar graph of predicted and actual completion times per task performance order and task experience condition (data from the three-disk task)

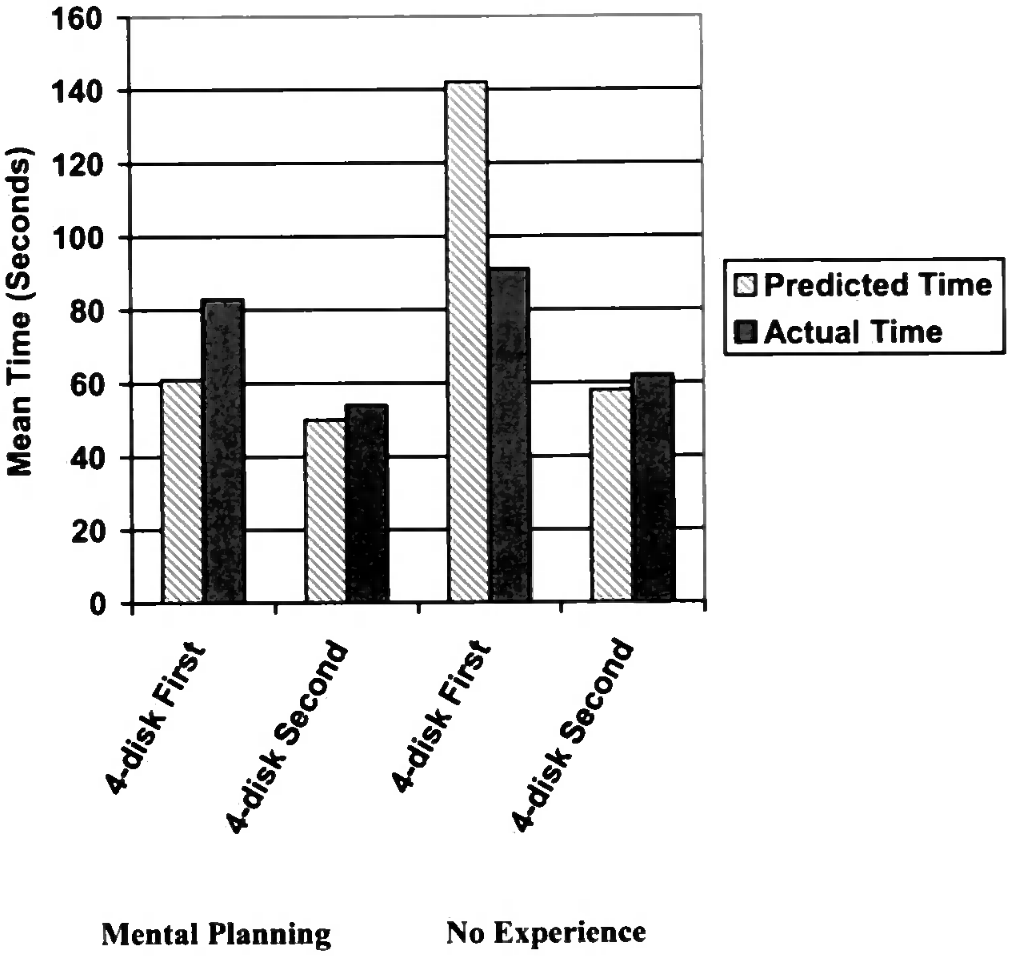

Figure 2.4

Bar graph of predicted and actual completion times per task performance order and task experience condition (data from the four-disk task)

In order to determine how many participants predicted task duration using whole minutes, frequency distributions of time estimates were calculated. Histograms of time estimates on the three-disk and four-disk tasks (regardless of performance order and prior experience condition) are presented below in Figures 2.5 and 2.6, respectively.

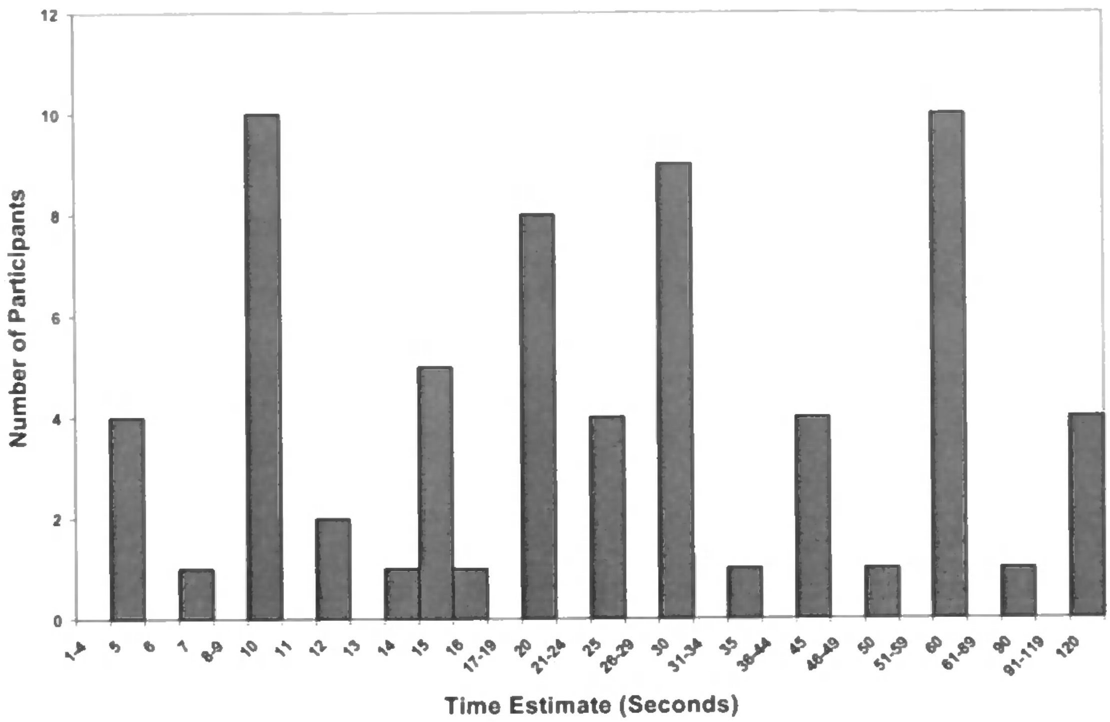

Figure 2.5

Histogram of time estimates on the three-disk task

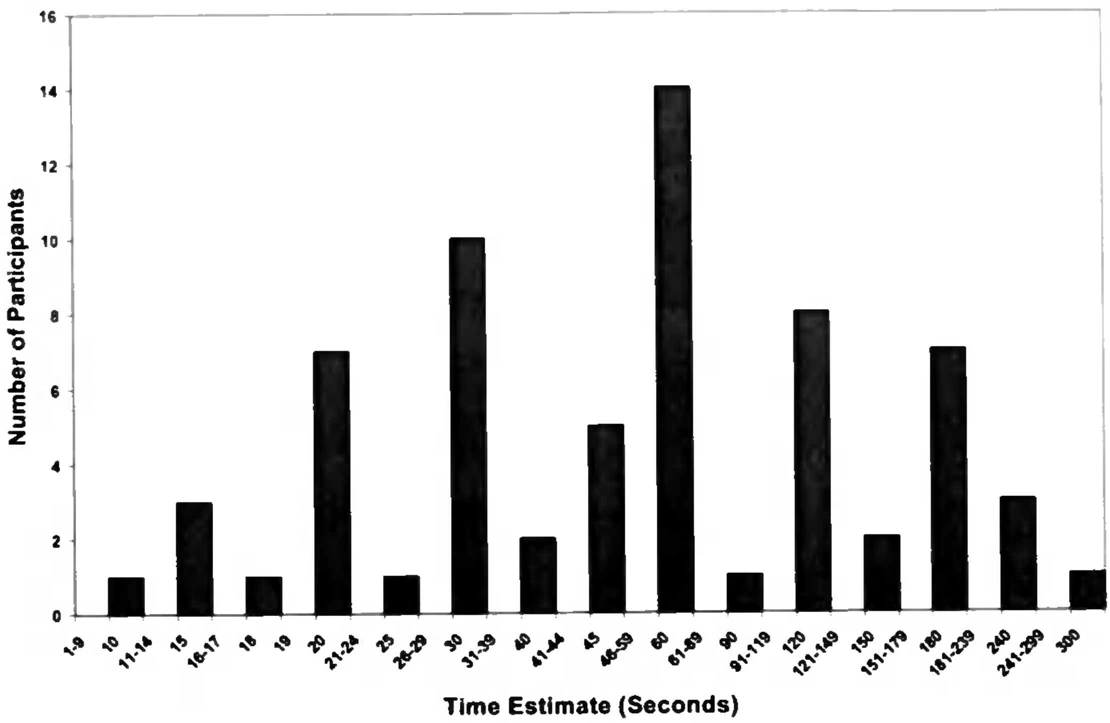

Figure 2.6

Histogram of time estimates on the four-disk task

For the same reasons that were specified in Experiment 1, the estimated and actual completion time data were subjected to a logarithmic transformation before being statistically analysed. The log-transformed data from the first trial were analysed in order to address the issues of prior experience and task complexity. A 2 (task) x 2 (time) x 2 (experience) mixed design ANOVA produced a main effect of time, $F(1,62) = 5.27$, $MSE = .30$, $p < .05$, with estimates exceeding completion times (Ms = 66.28 and 51.91 seconds, respectively). This finding suggests that there was general temporal overestimation on the first trial. The main effect of task experience was also significant, $F(1,62) = 10.48$, $MSE = .65$, $p < .01$, with overall time being longer in the no experience condition. There was also a main effect of task, $F(1,62) = 96.31$, $MSE = .65$, $p < .001$, which revealed that overall time was longer on the four-disk task.

The interaction between time and experience was significant, $F(1,62) = 7.40$, $MSE = .30$, $p < .01$. This revealed that, across both tasks, the difference between estimated and actual duration was greater in the no experience condition. The time by task interaction was also significant, $F(1,62) = 5.53$, $MSE = .30$, $p < .05$ (see Figure 2.7 below). This revealed that estimates exceeded completion times to a greater extent on the three-disk task (Ms = 31.22 and 16.66 seconds, respectively) than on the four-disk version (Ms = 101.35 and 87.19 seconds, respectively). Planned comparisons (LSD t-tests, both two-tailed) revealed that the difference between estimated and actual duration was significant on the three-disk ($p < .05$), but not the four-disk task ($p > .10$). This finding suggests that time predictions were less pessimistically biased on the more complex task.

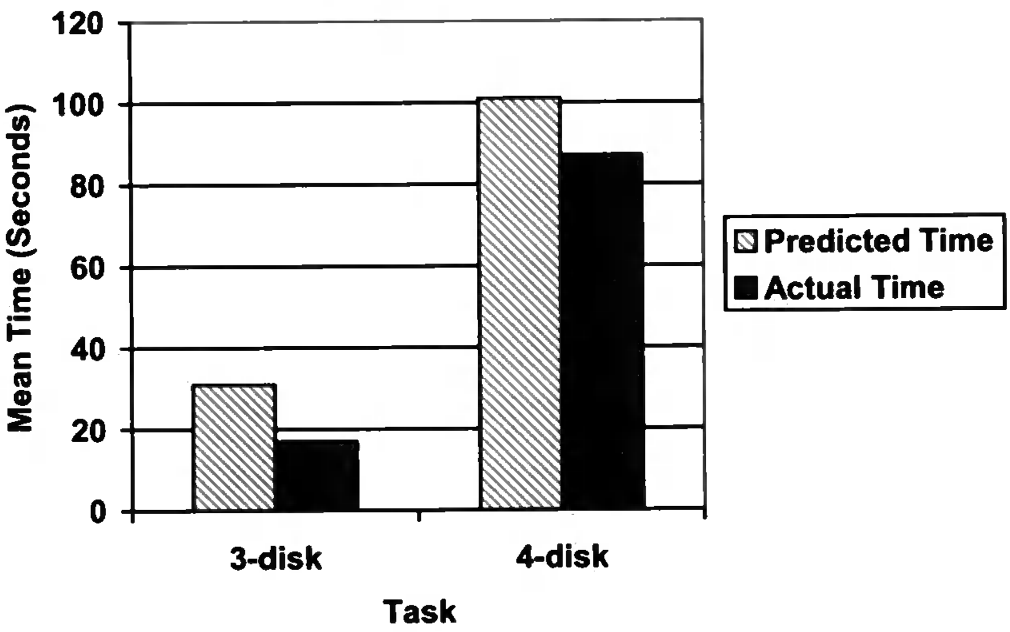

Figure 2.7

Bar graph depicting the interaction between the time and task factors (data from the first task)

There was also a significant interaction between the task, time and experience factors, $F(1,62) = 4.22$, $MSE = .30$, $p < .05$ (see Figure 2.8 below). This revealed that the direction of prediction bias differed according to prior experience on the four-disk task only. That is, temporal underestimation was evident in the mental planning condition, whereas overestimation occurred in the no experience condition. In contrast, temporal overestimation was evident on the three-disk task regardless of prior experience. Planned comparisons (LSD t-tests, all two-tailed) revealed that the difference between the estimated and actual duration of each task was significant in the no experience condition ($ps < .05$), but not in the mental planning condition ($ps > .10$). This finding indicates that time predictions were more accurate among participants who previewed each task and its instructions. The ANOVA produced no other significant main effects or interactions ($Fs < 2$, $ps > .10$).

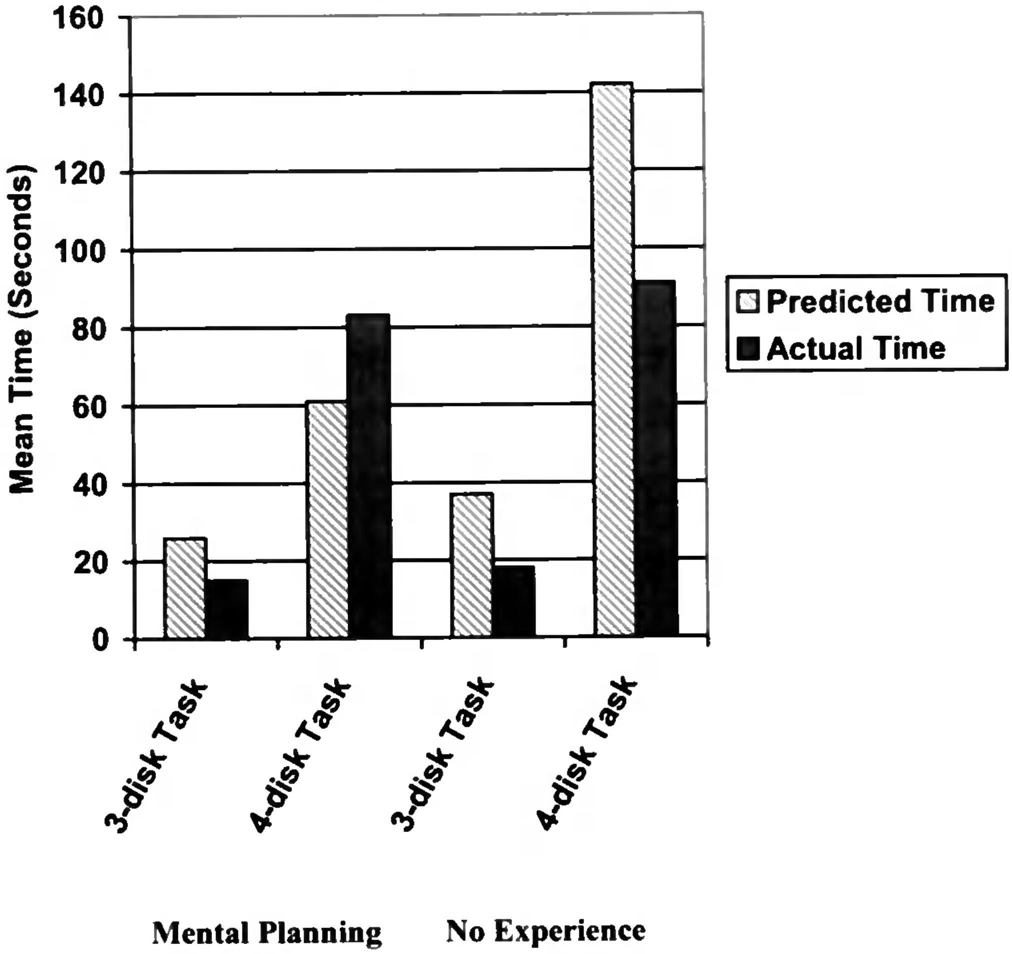

Figure 2.8

Bar graph depicting the interaction between the time, task and experience factors (data from the first task)

In order to ascertain whether prior performance of a different version of the Tower of Hanoi affected prediction bias, the log-transformed data from the second task were analysed. A 2 (time) x 2 (task) x 2 (experience) mixed design ANOVA produced a main effect of time, $F(1,62) = 12.63$, $MSE = .32$, $p < .001$, with estimates exceeding completion times (Ms = 44.18 and 34.83 seconds, respectively). Thus, general temporal overestimation was evident on the second trial. The main effect of task was also significant, $F(1,62) = 95.97$, $MSE = .36$, $p < .001$, with overall time being longer on the four-disk task. There was also a significant main effect of experience, $F(1,62) = 15.32$, $MSE = .36$, $p < .001$, which revealed that overall time was longer in the no experience condition.

The task by experience interaction approached significance, $F(1,62) = 3.14$, $MSE = .36$, $p < .09$. The interaction between time and task was significant, $F(1,62) = 25.41$, $MSE = .32$, $p < .001$ (see Figure 2.9 below). This revealed that, regardless of prior task experience, temporal underestimation was evident on the four-disk task, whereas overestimation occurred on the three-disk task. Planned comparisons (LSD t-tests, both two-tailed) revealed that the difference between estimated and actual duration was significant on the three-disk ($p < .05$), but not the four-disk task ($p > .10$). This finding suggests that, on the second trial, time predictions were less biased on the four-disk task. The ANOVA produced no other significant interactions (Fs < 1.5, ps > .10).

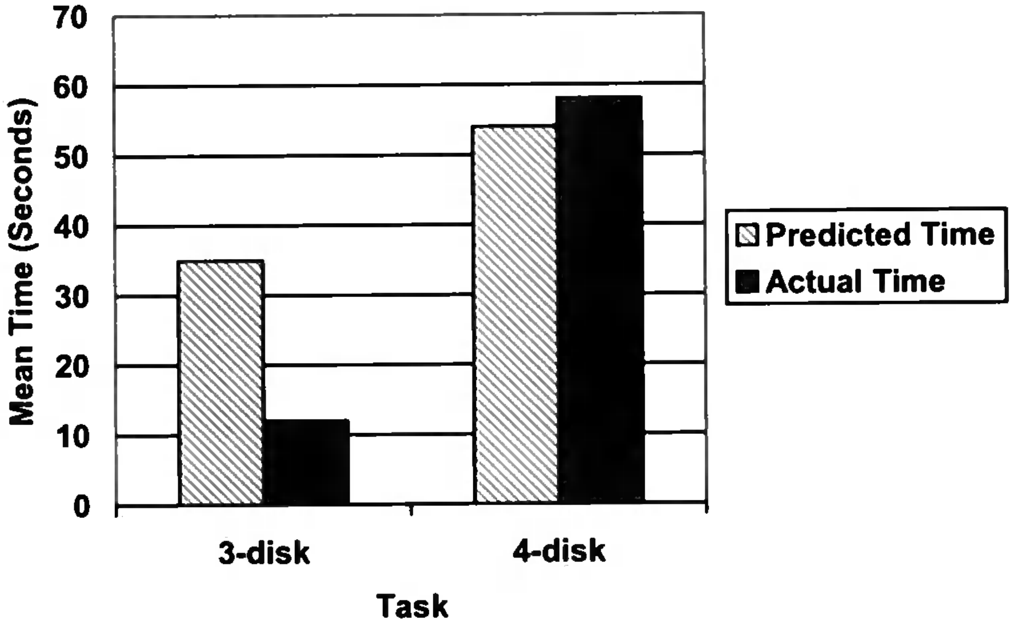

Figure 2.9

Bar graph depicting the interaction between the time and task factors (data from the second task)

2.10 Discussion

As was the case in Experiment 1, previewing the three-disk version of the Tower of Hanoi task and its instructions for two minutes resulted in greater time prediction accuracy. That is, on the first trial, the extent of temporal overestimation on this task was significantly reduced in the mental planning condition. It seems likely that the pre-exposure period enabled these participants to mentally represent the disk moves necessary for task completion, which resulted in them making more accurate time predictions.

This suggestion is broadly consistent with Phillips et al.'s (2001) research, which revealed that participants could mentally preplan up to seven disk moves on the Tower of London task. Since a minimum of only seven disk moves is needed to complete the three-disk version of the Tower of Hanoi task, previewing this task and its instructions for two minutes may well have facilitated mental preplanning. Moreover, given the relative simplicity of this task, the disk moves might have been retained in working memory (Davies, 2000a) during the pre-exposure period, and recalled to make a more realistic time estimate on the first trial.

Task experience was also found to be beneficial on the more complex four-disk version of the Tower of Hanoi task. That is, on the first trial, time estimates were more accurate among participants who previewed this task and its instructions for two minutes. A possible explanation for this finding is that, during the pre-exposure period, participants acquired task-related information, which was used to good effect. For example, participants may have worked out how to complete the task, calculated the amount of time needed to move all the disks in the correct pattern, and based their first temporal estimate on such information.

Although it was suggested that cognitive processing limitations might preclude mental preplanning on the four-disk task, Davies (2000a) found that this was definitely the case on the five-disk task, which involves a minimum of 31 disk moves compared to 15 on the four-disk task. Given this difference in complexity, cognitive processing limitations are likely to preclude mental preplanning on the five-disk task. For example, the 31 disk moves required for optimal task completion will exceed working memory capacity. Whilst the issue of mental preplanning on the four-disk task has not been addressed in previous research, the present study suggests that time prediction accuracy improves when people are encouraged to think about the task completion process for two minutes beforehand.

There was further evidence of temporal overestimation on the three-disk task in the present study. In fact, a pessimistic time prediction bias occurred on this task regardless of prior experience or task performance order. This finding suggests that the temporal underestimation indicative of the planning fallacy does not occur on this task. A plausible explanation for the absence of an optimistic time prediction bias on this task is its simplicity relative to the laboratory tasks used in previous research (e.g., Francis-Smythe & Robertson, 1999). That is, the greater complexity of the laboratory tasks used previously could be responsible for the temporal underestimation that was observed in such research. Thus, over-optimism might only be evident on tasks that are more complex than the three-disk version of the Tower of Hanoi task.

Consistent with this suggestion, there was a hint of an optimistic time prediction bias on the more complex four-disk task. However, the direction in which time predictions were biased on this task seemed to differ according to the order in which it was performed. Specifically, there was a hint of temporal overestimation when the four-disk task was performed first, whereas there was a hint of temporal underestimation when it was performed second. A possible explanation for the latter finding is that time estimates were based on information about the just-completed three-disk task. For example, the number of disk moves needed to complete this task may have formed the basis of estimates on the four-disk task. Given the difference between the tasks in terms of the number of disk moves, using such a judgement strategy would be expected to result in temporal underestimation. However, as temporal overestimation was evident on the three-disk task regardless of performance order, there was no evidence of such a judgment strategy being used on this task.

The relative simplicity of the four-disk task could explain why there was some evidence of temporal overestimation on this task when it was performed first. That is, participants were able to devote sufficient cognitive resources to the process of estimating task duration, a strategy that has been shown to reduce and eliminate overconfidence in non-temporal judgements (e.g., Koriat et al., 1980). Since the four-disk task took less than two minutes to complete, it was of shorter duration than the laboratory tasks used in previous research (e.g., Byram, 1997). As complex tasks are generally of longer duration than simple ones (Buckhalt & Oates, 2002), the four-disk task was presumably also less difficult than the laboratory tasks used previously. Hence, this task may have been too simple to induce overconfidence in participants, a notion that is consistent with previous research (Sniezek et al., 1990).

There was also evidence of greater time prediction accuracy on the four-disk task. That is, on both trials, the extent of temporal misestimation was significantly greater on the three-disk task. Whilst it is unclear why the extent of temporal underestimation on the four-disk task should be less than the magnitude of temporal overestimation on the three-disk task on the second trial, the finding from the first trial is broadly consistent with the difficulty effect (Griffin & Tversky, 1992). Specifically, time predictions were less pessimistic (i.e., less underconfident) on the more complex task, suggesting that participants were more confident as well as more accurate when estimating its duration. In order to confirm the existence of a relationship between task complexity and time prediction accuracy, the three-disk and four-disk tasks were employed again in Experiment 3.

The frequency distributions of time estimates revealed that several participants judged the duration of the three-disk and four-disk tasks using whole minutes. For example, almost 25 percent of the participants judged that the three-disk task would take them one or two minutes to complete. Similarly, half of all participants used temporal units of one to four minutes when judging the duration of the four-disk task. Given that each task took less than two minutes to complete, general temporal overestimation would be expected to occur as a consequence of using units of time such as two or three minutes when judging task duration. As previously mentioned, a more in-depth discussion of this issue will be provided later in this thesis.

The lack of general temporal underestimation on the three-disk and four-disk tasks could be due to the fact that no motivational incentives were on offer in Experiments 1 and 2. Support for this suggestion comes from Buehler et al. (1997), who found that time predictions on an anagram task were optimistically biased only when a monetary incentive was dependent on the speed of task completion. In a similar vein, Byram (1997) found that temporal underestimation was exacerbated when participants were offered a monetary reward for completing an origami task within a certain period of time.

These findings indicate that performance-contingent incentives can affect time estimation bias on laboratory tasks. Hence, the absence of such incentives could be responsible for the findings of Experiments 1 and 2, that is, participants failed to make over-optimistic time predictions because of insufficient motivation (Henry & Sniezek, 1993). In order to determine whether motivational incentives influenced time estimation bias on the Tower of Hanoi task, monetary rewards that were either contingent on the speed of task completion or on the accuracy of time predictions were offered in Experiment 3.

2.11 Experiment 3

A principal aim of this study was to ascertain whether motivational incentives influenced time estimation bias on the Tower of Hanoi task. Research has studied the impact of motivational incentives such as money on judgements of performance on tasks including general knowledge and almanac tests (e.g., Henry, 1994). Such research has found that incentives can improve the accuracy of numerical estimates (Wright & Aboul-Ezz, 1988; Wright & Anderson, 1989), strengthen a person's commitment to a goal, and lead to judgement overconfidence (Henry & Sniezek, 1993). Henry and Sniezek found that participants who were offered money for accurately predicting their performance on a general knowledge test tended to overestimate the number of questions that they would answer correctly. Henry and Sniezek suggest that people are motivated to increase their performance predictions in order to secure an incentive, which results in judgements that are overconfident.

Research by Wright and Anderson (1989) revealed that probabilistic judgements of the outcome of future events were less biased when participants were offered a monetary reward for prediction accuracy. Wright and Anderson suggest that the monetary incentive motivated participants to devote more cognitive resources to predicting their task performance, which resulted in greater judgement accuracy. Support for this suggestion derives from the fact that, relative to the no incentive condition, judgements of the incentive condition were less influenced by previously-presented (and irrelevant) numerical anchor values. This finding indicates that motivational incentives can improve the accuracy of non-temporal judgements of performance on various tasks, as well as reducing people's use of the anchoring and adjustment cognitive heuristics (Tversky & Kahneman, 1982a).

Given the findings of previous research (e.g., Byram, 1997), it was suspected that monetary incentives might mediate time estimation bias on the three-disk and four-disk versions of the Tower of Hanoi task. Consistent with the work of Buehler et al. (1997), monetary incentives that were contingent on the speed of task completion (speed incentive) or the accuracy of time estimates (accuracy incentive) were offered in the present study. It was anticipated that participants would engage in different reasoning strategies according to the type of incentive offered, which in turn would influence the direction in which their time estimates were biased. Following the findings of Buehler et al., it was hypothesised that temporal underestimation would be evident in the speed incentive condition, whereas temporal overestimation would occur in the accuracy incentive condition.

The issue of task complexity was further addressed in this study, where the three-disk and four-disk tasks were employed. Whilst there was a hint of an optimistic time prediction bias on the four-disk task in Experiment 2, temporal underestimation was only observed when the task was performed second. However, on the first trial, the magnitude of temporal overestimation was less on this task compared to the three-disk task. In order to confirm this finding, participants performed either the three-disk or the four-disk task initially. Consistent with the results of Experiment 2, it was hypothesised that temporal overestimation would be evident on both tasks, but that the extent of this pessimistic time prediction bias would be greater on the three-disk task.

The issue of prior task experience was also further addressed by manipulating the order in which the three-disk and four-disk tasks were performed. Experiment 2 revealed that the extent and direction of time prediction bias differed according to the relative complexity of a just-completed version of the same task. That is, on the second trial, the extent of temporal underestimation on the four-disk task was less than the magnitude of overestimation on the three-disk task. Given this finding, the impact of prior performance of a more or less complex similar task on subsequent time estimates was examined in this study.

2.12 Method

2.12.1 Participants

Ninety-nine (80 female and 19 male) students at the University of Plymouth participated voluntarily. Forty-eight participants took part in partial fulfilment of a psychology course requirement whilst the remainder were paid £1.50 each. No biographical information other than gender was recorded.

2.12.2 Materials

The materials were the same as those used in Experiment 2, with the exception of the word association task, which was not employed.

2.12.3 Design and Procedure

A 2 (time: estimated vs. actual) x 2 (task: three-disk vs. four-disk) x 2 (task performance order: three-disk first vs. four-disk first) x 3 (incentive: speed vs. accuracy vs. no incentive) mixed factorial design was used. The time and task factors were repeated-measures, with the order of task performance being fully counterbalanced. Task performance order was a between-groups factor, with participants being randomly assigned to one of two conditions. The incentive factor was also manipulated between groups, with participants being randomly assigned to one of three equal-sized conditions.

Except for the following differences, the procedure was identical to that used in Experiment 2. Since the mental planning condition produced the most accurate time estimates in Experiments 1 and 2, all participants were given two minutes to view each task and its instructions before estimating their completion times. At the end of each pre-exposure period, the performance-contingent incentives were introduced (where applicable) before task duration was estimated.

Regardless of whether they participated in return for money or as part of a course requirement, participants in the accuracy incentive condition were informed that they would receive a financial bonus of 50 pence for accurately estimating the duration of the current task. In order to qualify for this bonus, estimates on the three-disk task had to be within (plus or minus) two seconds of actual task duration, whereas estimates on the four-disk task had to be within seven seconds of actual completion time. The criteria for time prediction accuracy were based on pilot study data, and were within approximately 10% of the mean completion times of the three-disk and four-disk tasks (Ms = 20.25 and 70.56 seconds, respectively). Participants in this condition were informed of the relevant criteria for time prediction accuracy before making a temporal estimate on each task.

Regardless of whether they participated in return for money or as part of a course requirement, participants in the speed incentive condition were informed that they would receive a financial bonus of 50 pence if their completion times on the upcoming task were within the upper quartile of completion times observed in a pilot study. However, these participants were not given any information about the actual task completion times from the pilot study. Participants in the speed and accuracy incentive conditions could thus receive a bonus of £1 for fulfilling the relevant criteria on both tasks.

At the end of the second trial, all participants in the accuracy and speed incentive conditions received the full monetary reward regardless of task performance or time prediction accuracy. All participants in the no incentive condition were asked to be as accurate and realistic as possible when estimating the duration of each task. Each testing session lasted approximately 20 minutes.

2.13 Results

Basic descriptive statistics of estimated and actual duration per task performance order and incentive condition on the three-disk and four-disk tasks are presented in Tables 2.4 and 2.5, respectively. All test statistics used in Experiment 3 are contained in Appendix 3.

Task Performance Order		Incentive Condition		
		Accuracy	Speed	No Incentive
Three-disk First	Time Estimate	M = 32.12 Mdn = 15.00 SD = 55.61 N = 17	M = 19.31 Mdn = 13.50 SD = 16.07 N = 17	M = 21.24 Mdn = 10.00 SD = 19.80 N = 17
	Completion Time	M = 14.65 Mdn = 14.00 SD = 5.42 N = 17	M = 10.75 Mdn = 9.00 SD = 4.09 N = 17	M = 14.88 Mdn = 13.00 SD = 7.46 N = 17
Three-disk Second	Time Estimate	M = 16.06 Mdn = 12.00 SD = 11.74 N = 16	M = 23.76 Mdn = 15.00 SD = 22.55 N = 16	M = 26.31 Mdn = 20.00 SD = 26.56 N = 16
	Completion Time	M = 10.38 Mdn = 10.00 SD = 3.44 N = 16	M = 9.35 Mdn = 8.00 SD = 3.66 N = 16	M = 10.06 Mdn = 9.50 SD = 2.98 N = 16

Table 2.4

Time estimates and completion times on the three-disk task per performance order and incentive condition (in seconds)

Task Performance Order		Incentive Condition		
		Accuracy	Speed	No Incentive
Four-disk First	Time Estimate	M = 85.00 Mdn = 60.00 SD = 62.58 N = 16	M = 64.82 Mdn = 45.00 SD = 43.98 N = 16	M = 109.25 Mdn = 90.00 SD = 80.54 N = 16
	Completion Time	M = 99.69 Mdn = 103.50 SD = 73.02 N = 16	M = 76.35 Mdn = 57.00 SD = 61.00 N = 16	M = 70.25 Mdn = 53.00 SD = 51.78 N = 16
Four-disk Second	Time Estimate	M = 52.47 Mdn = 40.00 SD = 35.02 N = 17	M = 49.94 Mdn = 30.00 SD = 41.92 N = 17	M = 44.29 Mdn = 40.00 SD = 26.27 N = 17
	Completion Time	M = 91.12 Mdn = 63.00 SD = 72.33 N = 17	M = 69.31 Mdn = 43.50 SD = 59.85 N = 17	M = 86.59 Mdn = 98.00 SD = 45.10 N = 17

Table 2.5

Time estimates and completion times on the four-disk task per performance order and incentive condition (in seconds)

As Table 2.4 shows, estimates exceeded completion times in all cells on the three-disk task. This finding is consistent with those of Experiments 1 and 2, and provides further evidence of a general pessimistic time prediction bias on this task. In contrast, there was evidence of temporal underestimation on the four-disk task. That is, in the majority of cells in Table 2.5, temporal estimates were lower than completion times on this more complex task. Bar graphs of predicted and actual completion times from the accuracy, speed and no incentive conditions are presented below in Figures 2.10, 2.11 and 2.12, respectively.

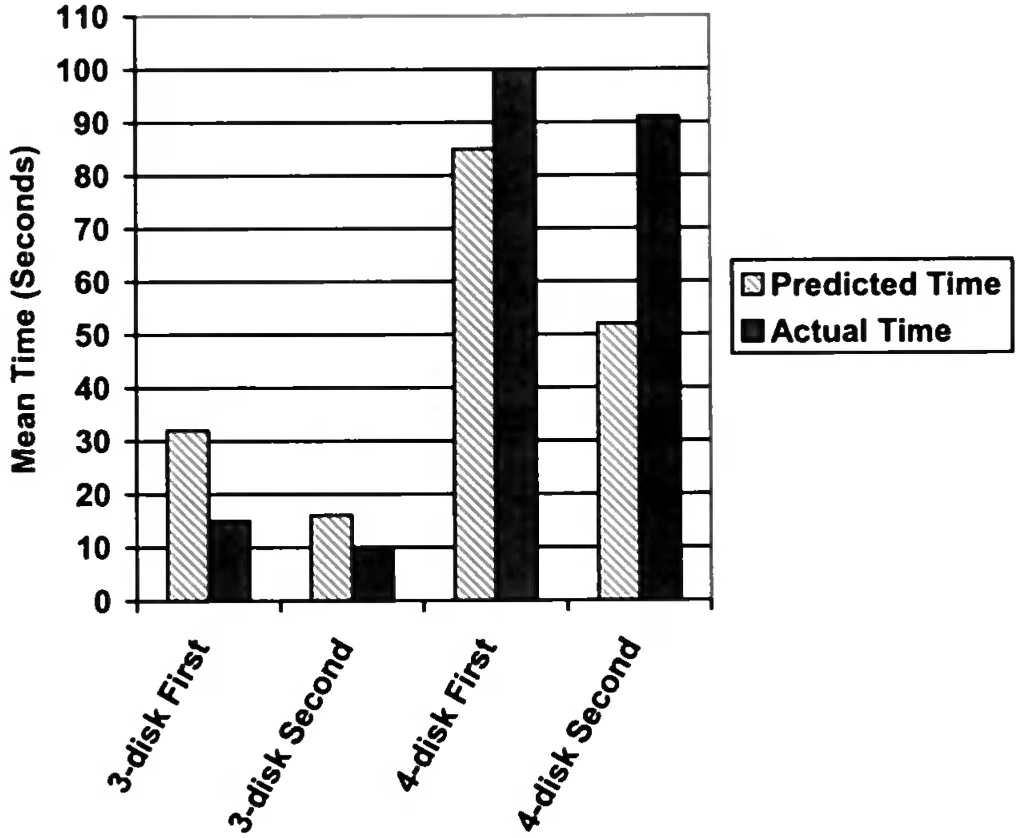

Figure 2.10

Bar graph of predicted and actual completion times per task and task performance order (data from the accuracy incentive condition)

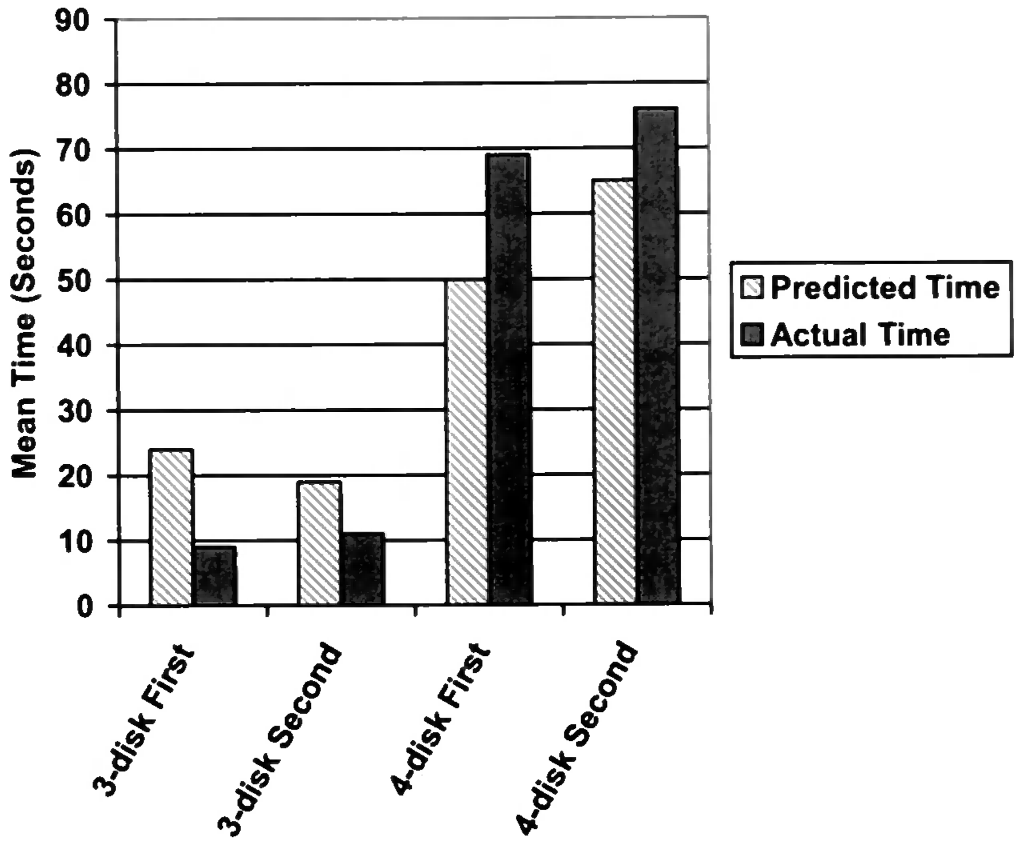

Figure 2.11

Bar graph of predicted and actual completion times per task and task performance order (data from the speed incentive condition)

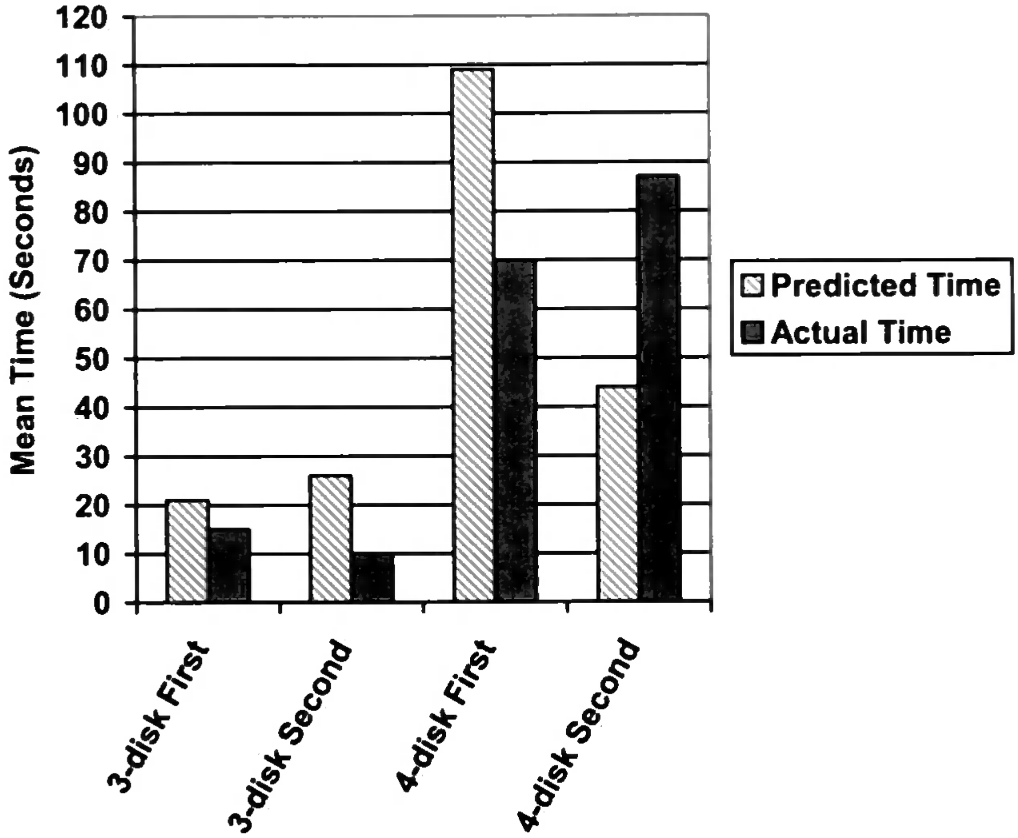

Figure 2.12

Bar graph of predicted and actual completion times per task and task performance order (data from the no incentive condition)

In order to determine how many participants predicted task duration using whole minutes, frequency distributions of time estimates were calculated. Histograms of time estimates on the three-disk and four-disk tasks (regardless of performance order and incentive condition) are presented below in Figures 2.13 and 2.14, respectively.

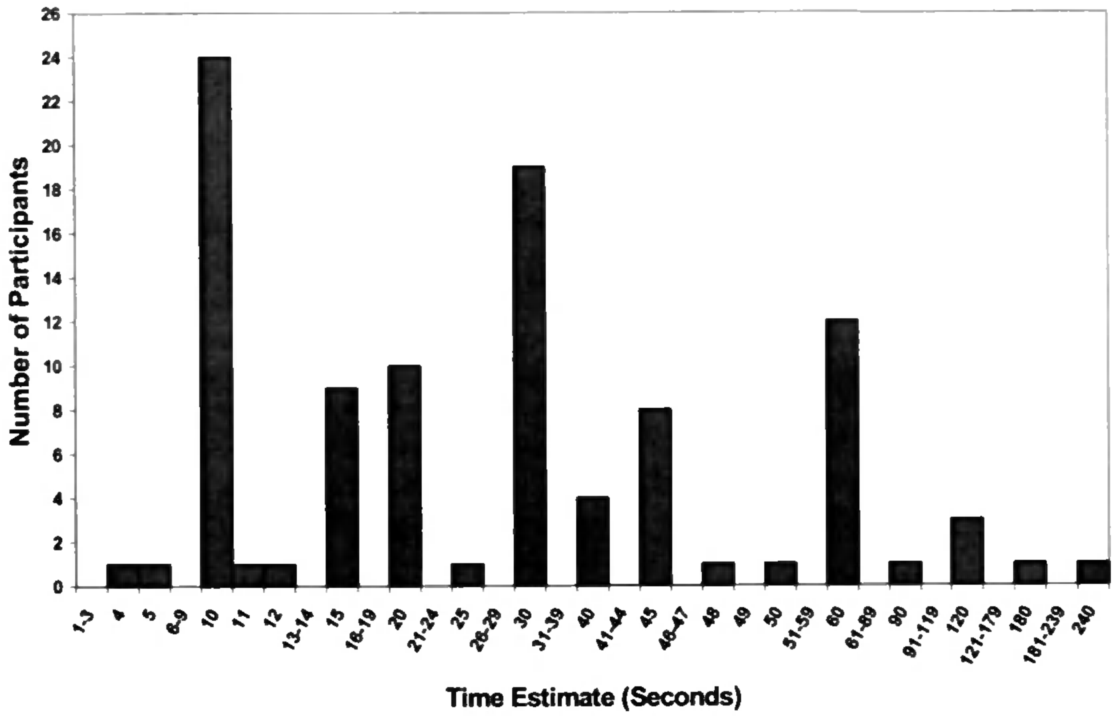

Figure 2.13

Histogram of time estimates on the three-disk task

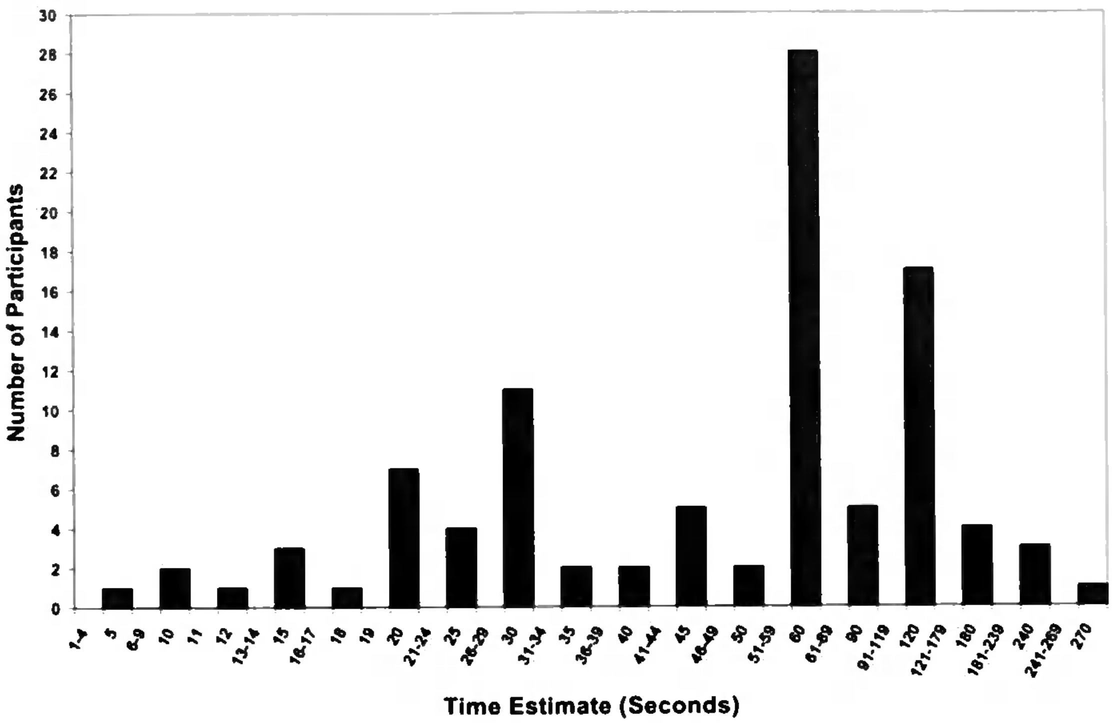

Figure 2.14

Histogram of time estimates on the four-disk task

For the same reasons that were specified in Experiment 1, the estimated and actual completion time data were subjected to a logarithmic transformation before being statistically analysed. In order to address the issue of task complexity, the log-transformed data from the first trial were analysed. A 2 (time) x 2 (task) x 3 (incentive) mixed design ANOVA produced a main effect of task, $F(1,93) = 163.94$, $MSE = .70$, $p < .001$, with overall time being longer on the four-disk task. The main effect of incentive was not significant ($F < 2$, $p > .10$). Likewise, the interaction between the task and time factors was not significant ($F < 2$, $p > .10$). Thus, although estimates exceeded completion times on the three-disk (Ms = 24.22 vs. 13.43 seconds, respectively) and four-disk tasks (Ms = 86.36 vs. 82.10 seconds, respectively), the magnitude of prediction bias did not differ significantly according to the complexity of the first trial. The ANOVA produced no other significant main effects or interactions (Fs < 2.5, ps > .10). Since the incentive factor did not interact with the other factors, there was no evidence that the monetary rewards influenced time prediction bias on the first trial.

The issue of prior task experience was addressed by analysing the log-transformed data from the second trial. A 2 (time) x 2 (task) x 3 (incentive) mixed design ANOVA produced a main effect of task, $F(1,93) = 174.17$, $MSE = .53$, $p < .001$, with overall time being longer on the four-disk task. The main effect of incentive was not significant ($F < 1$, $p > .10$). However, the interaction between the task and time factors was significant, $F(1,93) = 54.41$, $MSE = .24$, $p < .001$ (see Figure 2.15 below).

This interaction revealed that estimates exceeded completion times on the three-disk task (Ms = 20.56 and 10.40 seconds, respectively), whereas completion times exceeded estimates on the four-disk task (Ms = 84.68 and 53.86 seconds, respectively). Planned comparisons (LSD t-tests, both two-tailed) revealed a significant difference between estimated and actual duration on each task (ps < .05). This finding suggests that there was a significant pessimistic time prediction bias on the three-disk task, and a significant optimistic time prediction bias on the four-disk task. The ANOVA produced no other significant main effects or interactions (Fs < 1.5, ps > .10). Since the incentive factor did not interact with the other factors, there was no evidence that the monetary rewards influenced time prediction bias on the second trial.

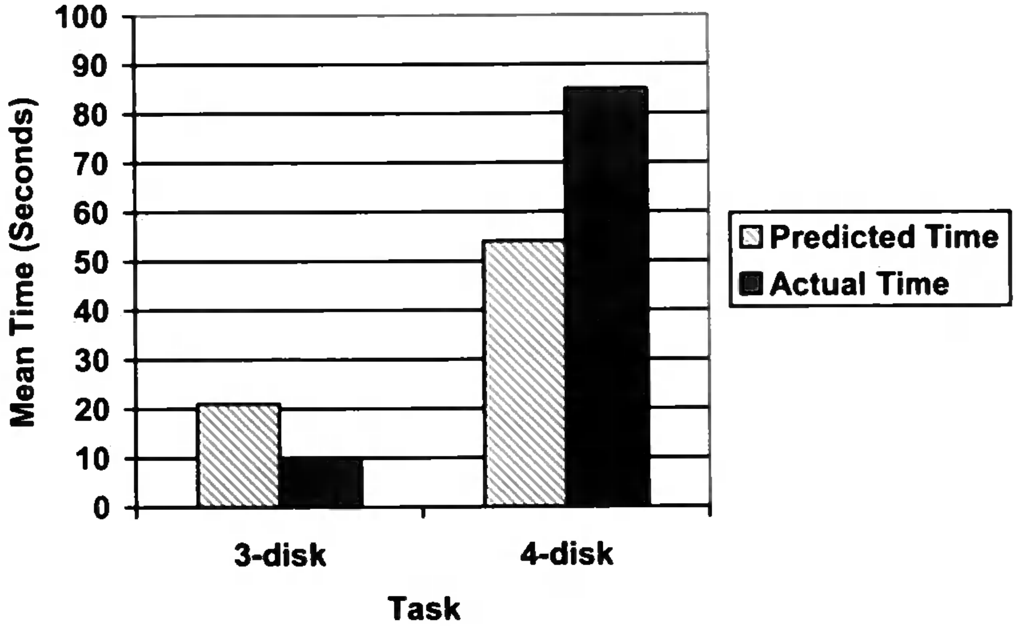

Figure 2.15

Bar graph depicting the interaction between the time and task factors (data from the second task)

Since the monetary rewards did not affect time prediction bias on either trial, the log-transformed task completion time data were analysed as an incentive manipulation check. If the monetary reward had the desired effect, then the speed incentive condition should produce the fastest completion times. A 2 (task) x 3 (incentive) mixed design ANOVA produced a main effect of task, $F(1,96) = 723.95$, $MSE = .22$, $p < .001$, with overall time being longer on the four-disk task. The interaction was not significant ($F < 1$, $p > .10$), suggesting that incentives did not differentially influence the duration of the three- and four-disk tasks. However, the main effect of incentive approached significance, $F(1,96) = 3.04$, $MSE = .42$, $p < .06$. This revealed that overall time was longest in the accuracy condition ($M = 53.92$ seconds), followed by the no incentive condition ($M = 45.61$ seconds), followed by the speed condition ($M = 41.48$ seconds). This finding is broadly consistent with the work of Buehler et al. (1997), which revealed that task completion times were shortest among participants who were offered a financial incentive for the speed of task completion.

2.14 Discussion

There was no evidence that the monetary incentives affected the direction or magnitude of time estimation bias on the three-disk and four-disk tasks. In contrast to the findings of Buehler et al. (1997), it seems that participants did not engage in motivated reasoning (Kunda, 1990) when estimating the duration of these tasks. However, there was some (albeit marginally significant) evidence that the present incentives influenced task duration in the manner predicted. That is, relative to the accuracy incentive condition, overall mean task duration was lower in the speed incentive condition. This finding suggests that participants who were offered the incentive for swift task completion finished the tasks in less time than participants who were offered the incentive for prediction accuracy. However, as it was only overall task duration that differed between the incentive conditions, it is unwise to make too much of this finding.

The magnitude of the incentives on offer could explain the disparity between the findings of Buehler et al. (1997) and those of the present study. That is, participants in Buehler et al.'s study could obtain $4 per task compared to 50 pence per task in this experiment. Thus, the larger incentives offered by Buehler et al. were presumably sufficient to motivate participants to use either a goal-directed or a deliberative reasoning strategy, which resulted in an optimistic or pessimistic time prediction bias, respectively. Conversely, the prospect of receiving only 50 pence was probably insufficient an incentive to induce the present participants to engage in motivated reasoning (Kunda, 1990). Hence, a feasible explanation for the present findings is that the monetary rewards were of insufficient magnitude to influence the direction in which participants' time estimates were biased.

Consistent with the findings of Experiments 1 and 2, there was further evidence of temporal overestimation on the three-disk task. In all incentive and task performance order conditions, time predictions exceeded the actual duration of this simple task. Likewise, there was a hint of temporal overestimation when the four-disk task was performed first, but this pessimistic time prediction bias was not significant. In contrast, there was evidence of a significant optimistic time prediction bias on this more complex task when it was performed second. Thus, as was the case in Experiment 2, the only evidence of temporal underestimation was on the four-disk task when the less complex three-disk task was performed beforehand.

Time estimation bias on the second trial was found to be influenced by the relative complexity of the just-completed task. That is, temporal underestimation was evident on the four-disk task, whereas a significant pessimistic time prediction bias only occurred when the three-disk task was performed second. Since these tasks share the same structure (Kotovsky et al., 1985), participants might have based their second time estimate on information about the nature of the previous task, which was insufficiently adjusted for the demands of the upcoming task. For example, participants may have calculated the amount of time needed to complete the three-disk task, but failed to scale up their estimate according to the greater number of disk moves involved in completing the four-disk task.

Given the difference in complexity between the three-disk and four-disk tasks, the use of such a judgement strategy would be expected to result in temporal underestimation on the four-disk version. Likewise, time predictions would be longer on the three-disk version if they were based on the greater number of disk moves involved in completing the four-disk task. Thus, temporal overestimation would be greater on the three-disk task as a consequence of using such a judgement strategy.

The frequency distributions of time estimates revealed that several participants judged the duration of the three-disk and four-disk tasks using whole minutes. For example, over half of all participants used temporal units of one to four minutes when judging the duration of the four-disk task, whereas 17 individuals used such units on the three-disk task. Given that each task took less than two minutes to complete, general temporal overestimation would be expected to occur as a consequence of using units of time such as two or three minutes when judging task duration. As previously mentioned, a more fine-grained discussion of this issue will be provided later in this thesis.

2.15 General Discussion

There was no evidence of temporal underestimation on the three-disk version of the Tower of Hanoi task. In fact, the present research highlights the directional nature of time prediction bias, as temporal estimates exceeded completion times on this task in every cell in every condition in Experiments 1, 2 and 3. Thus, participants erred on the side of caution when estimating the duration of this well structured laboratory task. This finding is of considerable importance since the temporal underestimation indicative of the planning fallacy has been observed in numerous studies (e.g., Byram, 1997; Josephs & Hahn, 1995; Koole & Van't Spijker, 2000; Newby-Clark et al., 2000) and on a variety of tasks. Moreover, this finding indicates that the planning fallacy is not as prevalent as previous research suggests (e.g., Buehler et al., 1994), and that there is at least one task on which it does not occur, and is in fact reversed.

Task complexity was suggested as a reason for the absence of an optimistic time prediction bias on the three-disk task. That is, the greater complexity of the laboratory tasks used in previous research (e.g., Byram, 1997) was responsible for the temporal underestimation that was observed in those studies. Given that the three-disk task can be completed in just a few seconds and involves making a minimum of only seven disk moves, it is a simple task. In order to ascertain whether an optimistic time prediction bias occurred on a more complex task, the four-disk version of the Tower of Hanoi task was employed in Experiments 2 and 3.

Whilst there was a hint of over-optimism on the four-disk task, the direction in which time predictions were biased was mediated by task performance order in both studies. That is, the only evidence of temporal underestimation was on this task when it was performed after the three-disk task. In contrast, there was some evidence of a pessimistic time prediction bias when the four-disk task was performed first in Experiments 2 and 3. Thus, general temporal underestimation was not evident on this more complex task.

The absence of a general optimistic time prediction bias on the four-disk task could be explained by the fact that it still took less time to complete than the laboratory tasks used in previous research. For example, the four-disk task took less than two minutes to complete, whereas the tasks used previously have taken at least eight minutes to finish (Buehler et al., 1997). Given that longer duration tasks also tend to be more complex than shorter ones (Buckhalt & Oates, 2002), the present tasks were presumably simpler than the ones used previously. Thus, it could be that the four-disk task was still too simple to induce overconfidence in participants, a notion that is broadly consistent with the difficulty effect (Griffin & Tversky, 1992). In order to test this suggestion, a more complex task was employed in the research presented in the next chapter of this thesis.

The present findings indicate that temporal misestimation is reduced when people have the opportunity to think about the task completion process beforehand. That is, the mental planning method resulted in greater time prediction accuracy on the three-disk (Experiments 1 and 2) and four-disk tasks (Experiment 2). It was suggested that the two-minute pre-exposure period enabled participants to acquire task-related information, which was used to good effect. Moreover, since both tasks took less than two minutes to complete, participants may well have mentally simulated the task completion process before estimating task duration. However, as these participants did not perform either task before making a single (Experiment 1) or first time estimate (Experiment 2), they had no experience of performing the task when making an initial prediction. Thus, they were unlikely to possess any kind of distributional information when estimating the duration of the first (Experiment 2) or only task (Experiment 1).

Since these participants were instructed to think about the task completion process during the pre-exposure period, they presumably based their time predictions on information about the current task. For example, time estimates on the three-disk task may have been based on the disk move pattern and the number of disk moves needed for task completion. However, on the basis of previous research (e.g., Buehler et al., 1994), time predictions should be more biased rather than more accurate as a consequence of relying on such singular information. Given the findings of Experiments 1 and 2, it could be that people are able to use singular information to good effect. That is, estimates of task duration are more accurate when people have spent time thinking about the demands of an upcoming task beforehand.

Consistent with previous research (Josephs & Hahn, 1995), prior experience of performing the same task also reduced time prediction bias (Experiment 1). That is, relative to the no experience condition, time prediction accuracy was greater among participants who actively practised performing the three-disk task beforehand. A plausible explanation for this finding is that participants drew upon their prior experience of performing the task when estimating its duration. Indeed, having completed the three-disk task at least once during the pre-exposure period, these participants possessed some kind of distributional information when estimating task duration. Moreover, as the target trial was performed immediately after the pre-exposure period, participants may have recalled information about the just-completed task when making a time estimate. Support for this notion comes from Buehler et al. (1994), who found that time prediction bias was reduced when participants recognised the similarities between previous and current tasks.

The temporal interval between previous and current tasks could explain the well established link between the planning fallacy and the neglect of distributional information. That is, the lengthy temporal interval between the repeated performance of many real world tasks (e.g., painting one's house) might impede the recall of information about personal performance on previous tasks, meaning that time estimates are based on information about the task at hand. As Kahneman and Tversky (1979) have suggested, the use of such singular information should result in temporal underestimation. Conversely, when the temporal interval between previous and current tasks is brief, information about previous task performance is likely to be recalled (Ross & Buehler, 2001), and could be used to make more accurate time estimates. Given the findings of Experiment 1, the effective use of information about previous task performance might be contingent upon task experience being acquired shortly before task duration is estimated.

Prior performance of a more or less complex version of the Tower of Hanoi task was found to influence time prediction bias (Experiment 3). That is, temporal underestimation was only evident on the four-disk task when it was performed second, whereas the extent of overestimation was greater on the three-disk task when it was performed second. It was suggested that this finding occurred because time predictions on the second trial were based on information about the nature of the just-completed task.

If this were the case, then an alternative interpretation of the planning fallacy suggests itself. That is, an optimistic time prediction bias is a consequence of temporal estimates being based on information such as the duration of previous similar tasks, which are less complex than the current one. Thus, rather than neglecting distributional information, people may take account of their previous task performance, but fail to scale up their time estimate in accordance with the greater demands of the task at hand. Given the findings of Experiment 3, the order in which more or less complex versions of the same task were performed was manipulated in the research presented in the next chapter of this thesis.

Contrary to the findings of Buehler et al. (1997), performance-contingent monetary incentives did not affect the direction in which time estimates were biased (Experiment 3). Hence, there was no evidence that participants engaged in motivated reasoning (Kunda, 1990) when estimating the duration of the three-disk and four-disk tasks. The relative magnitude of the incentives seems a plausible explanation for the disparity between the present findings and those of Buehler et al. Specifically, the present monetary rewards may have been too meagre to motivate participants to adopt a goal-directed or a deliberative reasoning strategy.

The success of the incentives used by Buehler et al. could also be due to the fact that participants were given outcome feedback prior to making a time estimate. That is, at the end of each practice trial, Buehler et al.'s participants received a slip of paper detailing the duration of the just-completed task. Thus, when estimating task duration, these participants possessed information that was relevant to the target trial. In order to obtain the monetary reward, such information may have been incorporated into time estimates on the target trial. For example, the duration of the second practice trial might have served as a benchmark for time predictions on the target trial in the speed incentive condition.

Given that Buehler et al.'s speed incentive was contingent on finishing the target trial in less time than the previous trial, predictions on the target trial may have been lower than the duration of the second practice trial. As all trials of the anagram task were of similar complexity and duration (Buehler et al., 1997), temporal underestimation would be expected to occur on the target trial if time predictions were lower than the duration of the second practice trial. Whilst this suggestion is somewhat speculative, the provision of outcome feedback may well have confounded Buehler et al.'s findings. Given this potential confound, and the findings of Experiment 3, the issue of performance-contingent incentives was not addressed further in this research programme.

2.16 Conclusions

A number of potentially important findings have emerged from the research presented in this chapter. In contrast to the findings of previous research (e.g., Buehler et al., 1997; Byram, 1997; Newby-Clark et al., 2000), there was no evidence of temporal underestimation on the three-disk version of the Tower of Hanoi task. In fact, in every cell in every condition of these studies, temporal overestimation was evident on this simple well structured task. There was also some evidence of temporal overestimation on the more complex four-disk task when it was performed first (Experiments 2 and 3). However, there was a hint of temporal underestimation when the four-disk task was preceded by the three-disk task (Experiments 2 and 3).

The present studies also highlight the impact of prior task experience on time estimation bias. In Experiments 1 and 2, prior experience of the same task was found to improve prediction accuracy. This finding suggests that information about a current task can be considered and used to make more accurate time estimates. The effective use of task-related information was also emphasised in Experiment 3, where time prediction bias on the second task differed according to the relative complexity of the previous task. Specifically, the extent of overestimation on the three-disk task was greater when it was performed second rather than first, whereas temporal underestimation was only evident when the four-disk task was performed second.

The research presented in the next chapter of this thesis consists of two experiments. The first study sought to confirm some of the major findings from Experiments 1 to 3, whereas the second study examined the impact of task duration on time estimation bias. In order to further address the issue of task complexity, the five-disk version of the Tower of Hanoi task was employed alongside the three-disk task in Experiment 4. The impact of prior experience of performing the same task was also further examined in this study. The effect of task duration on time estimation bias was investigated in Experiment 5, where task complexity was held constant.

Chapter Three

The Effects of Task Complexity and Duration on Time Estimation Bias

3.1 Overview

The research presented in this chapter comprises two experiments. Given the findings of Experiments 2 and 3, the issue of task complexity was further addressed in Experiment 4. This study sought to ascertain whether temporal underestimation was evident on the more complex five-disk version of the Tower of Hanoi task. In order to confirm the findings of Experiment 3, the issue of prior experience of performing a more or less complex version of the same task was also addressed in this study. Experiment 4 also examined whether prior performance of an identical task improved time estimation accuracy.

The impact of task duration on time estimation bias was examined in Experiment 5, where task complexity was held constant. This study employed short and long duration versions of a simple repetitive disk movement task, which were of similar duration to the three-disk and five-disk versions of the Tower of Hanoi task, respectively. The issue of prior experience of performing a similar task was addressed in Experiment 5, where the order in which the two repetitive tasks were performed was manipulated.

3.2 General Introduction to Experiments 4 and 5

A general optimistic time prediction bias was not evident on the three-disk (Experiments 1 to 3) and four-disk versions of the Tower of Hanoi task (Experiments 2 and 3). This finding contrasts with previous research (e.g., Francis-Smythe & Robertson, 1999), where general temporal underestimation was observed on other laboratory tasks.

The relative simplicity of the three-disk and four-disk tasks was suggested as an explanation for the discrepancy between the findings of Experiments 1 to 3 and those of previous research. That is, general temporal underestimation might only occur on tasks that are more complex than those used in the present research. Indeed, temporal underestimation was evident on laboratory tasks such as constructing self-assembly furniture (Byram, 1997) and proof-reading a manuscript (Josephs & Hahn, 1995), both of which are more complex than the three-disk and four-disk versions of the Tower of Hanoi task.

There is considerable evidence that task complexity mediates bias in non-temporal judgements of task performance (e.g., Juslin et al., 2000; Pulford & Colman, 1997; Winman, 1999). Consistent with the judgement calibration phenomenon known as the difficulty effect (Griffin & Tversky, 1992), such research suggests that over-optimism prevails on more complex tasks whereas pessimism prevails on simpler tasks. Given the extent of support for the difficulty effect (Suantak et al., 1996), it could be that the direction in which time estimates are biased also differs according to task complexity. In fact, there was a hint of an optimistic time prediction bias on the more complex four-disk task, but only when it was performed after the three-disk task (Experiments 2 and 3). In order to determine whether a general optimistic time prediction bias was evident on a more complex task, the five-disk version of the Tower of Hanoi task was employed in Experiment 4.

Whilst tasks such as constructing self-assembly furniture (Byram, 1997) are more complex than the three-disk and four-disk tasks, they are also of longer duration. For example, the proof-reading task employed by Josephs and Hahn (1995) took about 40 minutes to complete, whereas the duration of the three-disk task was about 20 seconds (Experiments 1 to 3). Hence, it could be that the temporal underestimation indicative of the planning fallacy is only evident on laboratory tasks that take more time to complete than the ones used in the present research. In order to determine whether time estimation bias differs according to task duration, cognitive complexity was held constant in Experiment 5. In this study, participants performed a longer and a shorter duration version of a simple repetitive disk movement task.

Different types of prior task experience were found to influence time estimation bias in Experiments 1 to 3. Firstly, predictions were more accurate when participants performed the three-disk task (Experiment 1) or mentally planned how to complete the three-disk and four-disk tasks beforehand (Experiments 1 and 2). It was suggested that time prediction accuracy was greater because participants used the pre-exposure period to think about information such as the steps involved in task completion. That is, when making a time estimate, participants drew on task-related information, which they acquired during the pre-exposure period.

Secondly, prior experience of performing a more or less complex similar task was found to mediate time prediction bias (Experiment 3). Specifically, on the second trial, the magnitude of temporal overestimation was greater on the three-disk task, whereas an optimistic time prediction bias was evident on the four-disk task. It was suggested that this finding occurred because participants used information about the previous task as a basis for their second time prediction, but failed to scale this estimate up or down according to the relative complexity of the current task. Such a judgement strategy would be expected to result in an optimistic time prediction bias when the previous task was less complex than the current one. Likewise, temporal overestimation should be greater when the second task was less complex than the first one. That is, time predictions would be longer if they were based on information about a more complex task that had just been completed.

Given the finding of Experiment 3, the effect of prior task experience on time estimation bias was further explored in the present studies. In Experiments 4 and 5, task performance order was manipulated in order to further address the issue of prior experience of a performing a similar task. The issue of prior performance of the same task was further addressed in Experiment 4, where some participants performed the three-disk or five-disk tasks twice in succession.

3.3 Experiment 4

This study further investigated the impact of task complexity on time prediction bias. Given the findings of Experiments 1 to 3, it was suggested that a general optimistic time prediction bias might only be evident on more complex tasks than the three-disk and four-disk versions of the Tower of Hanoi task. In order to test this suggestion, task complexity was increased in this study by using the five-disk version of the Tower of Hanoi task, which was employed alongside the three-disk version. Given that over-optimism prevails in non-temporal judgements of performance on complex tasks (e.g., Griffin & Tversky, 1992), it could be that temporal underestimation will be evident on the five-disk task. However, as this task has not been used in previous research into time estimation, no prediction was made about the direction of judgement bias on it. Consistent with the findings of the previous experiments, general temporal overestimation was expected to occur on the three-disk task.

The issue of prior task experience was also further addressed in this study. That is, whether prior performance of a similar or an identical task influenced time prediction bias on the next task. To this end, some participants performed the same task twice in succession, whereas others completed a more or less complex version of the Tower of Hanoi task initially. Given the finding of Experiment 3, it was hypothesised that time estimation bias would differ according to the relative complexity of the previous task. Specifically, an optimistic time prediction bias was expected to occur when the five-disk task was preceded by the three-disk task, whereas temporal estimates should be more pessimistically biased when the three-disk task was performed after the five-disk task. Since prior performance of the same task has been found to reduce time estimation bias (Josephs & Hahn, 1995), it was anticipated that predictions would be more accurate when the same rather than a different version of the Tower of Hanoi task was performed initially.

3.4 Method

3.4.1 Participants

Ninety-four (88 female and 6 male) students at the University of Plymouth participated voluntarily in partial fulfilment of a psychology course requirement. No biographical information other than gender was recorded.

3.4.2 Materials

The materials were the same as those used in Experiment 3 except that the Tower of Hanoi apparatus contained five different-sized disks. The two largest disks were removed to form the three-disk task.

3.4.3 Design and Procedure

A 2 (time: estimated vs. actual) x 2 (task: first vs. second) x 4 (task experience: three-disk twice vs. five-disk twice vs. three- then five-disk vs. five- then three-disk) mixed factorial design was used. The time and task factors were repeated-measures. The order in which the tasks were performed was determined by random assignment to a level of the task experience factor. Task experience was a between-groups factor, with participants being randomly assigned to one of four conditions.

The procedure was identical to that used in Experiment 3 except that participants were given 20 seconds (after reading the task instructions) to make a time prediction on each task. Each testing session lasted approximately 20 minutes.

3.5 Results

Table 3.1 below contains basic descriptive statistics of estimated and actual duration per task and task experience condition. All test statistics used in Experiment 4 are contained in Appendix 4.

Task		Task Experience Condition			
		Three-disk Twice	Five- Then Three-disk	Five-disk Twice	Three- Then Five-disk
First	Time Estimate	M = 93.54 Mdn = 90.00 SD = 49.88 N = 24	M = 243.96 Mdn = 270.00 SD = 169.01 N = 23	M = 268.70 Mdn = 240.00 SD = 160.41 N = 24	M = 71.96 Mdn = 60.00 SD = 46.65 N = 23
	Completion Time	M = 32.71 Mdn = 24.50 SD = 23.11 N = 24	M = 248.75 Mdn = 219.50 SD = 35.14 N = 23	M = 247.22 Mdn = 222.00 SD = 109.18 N = 24	M = 31.09 Mdn = 29.00 SD = 19.66 N = 23
Second	Time Estimate	M = 30.63 Mdn = 25.00 SD = 22.33 N = 24	M = 82.50 Mdn = 60.00 SD = 52.29 N = 23	M = 204.57 Mdn = 180.00 SD = 121.85 N = 24	M = 120.22 Mdn = 90.00 SD = 77.00 N = 23
	Completion Time	M = 19.17 Mdn = 16.00 SD = 13.93 N = 24	M = 18.00 Mdn = 14.00 SD = 10.03 N = 23	M = 203.87 Mdn = 207.00 SD = 117.57 N = 24	M = 169.39 Mdn = 161.00 SD = 85.17 N = 23

Table 3.1

Time estimates and completion times per task and task experience condition (in seconds)

As Table 3.1 shows, estimates exceeded completion times on the three-disk task regardless of prior experience. That is, temporal overestimation was evident in the three-disk twice, the five- then three-disk and the three- then five-disk task conditions. This finding is consistent with Experiments 1 to 3, and indicates that general temporal overestimation occurred on this simple task. However, Table 3.1 also shows that temporal underestimation was evident when the five-disk task was performed before or after the three-disk version, whereas overestimation occurred on the first and second trials in the five-disk task twice condition. Bar graphs of predicted and actual completion times on the first and second tasks are presented below in Figures 3.1 and 3.2, respectively.

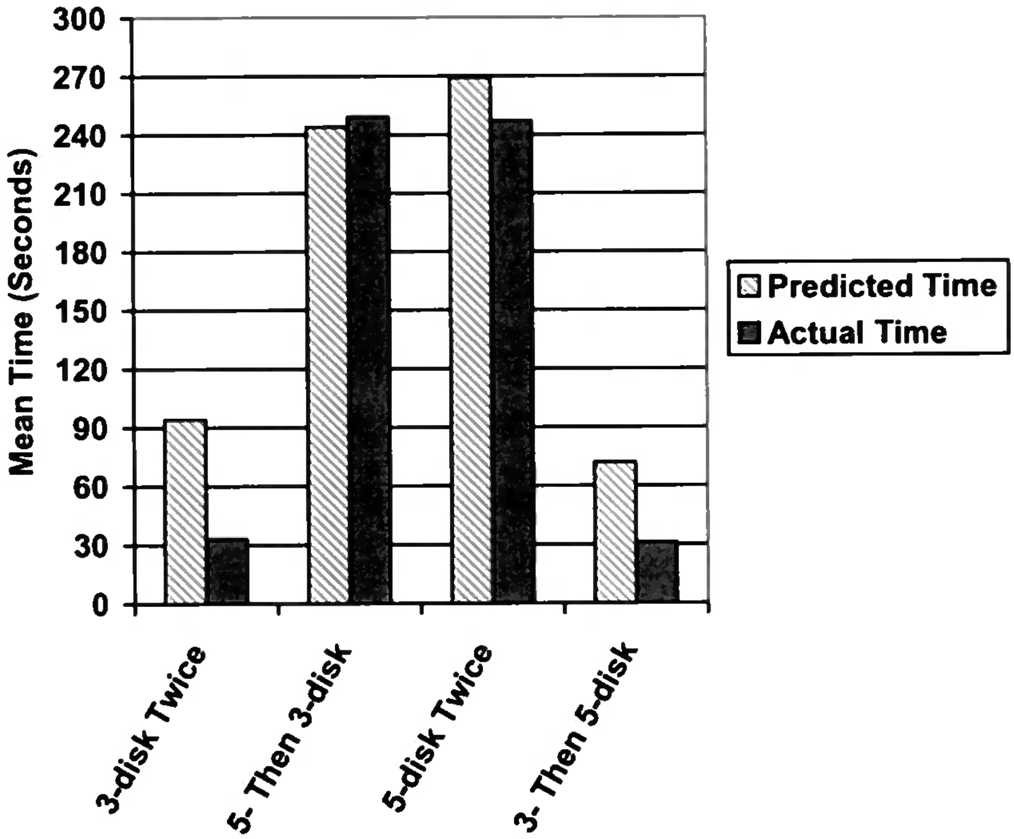

Figure 3.1

Bar graph of predicted and actual completion times per task experience condition (data from the first task)

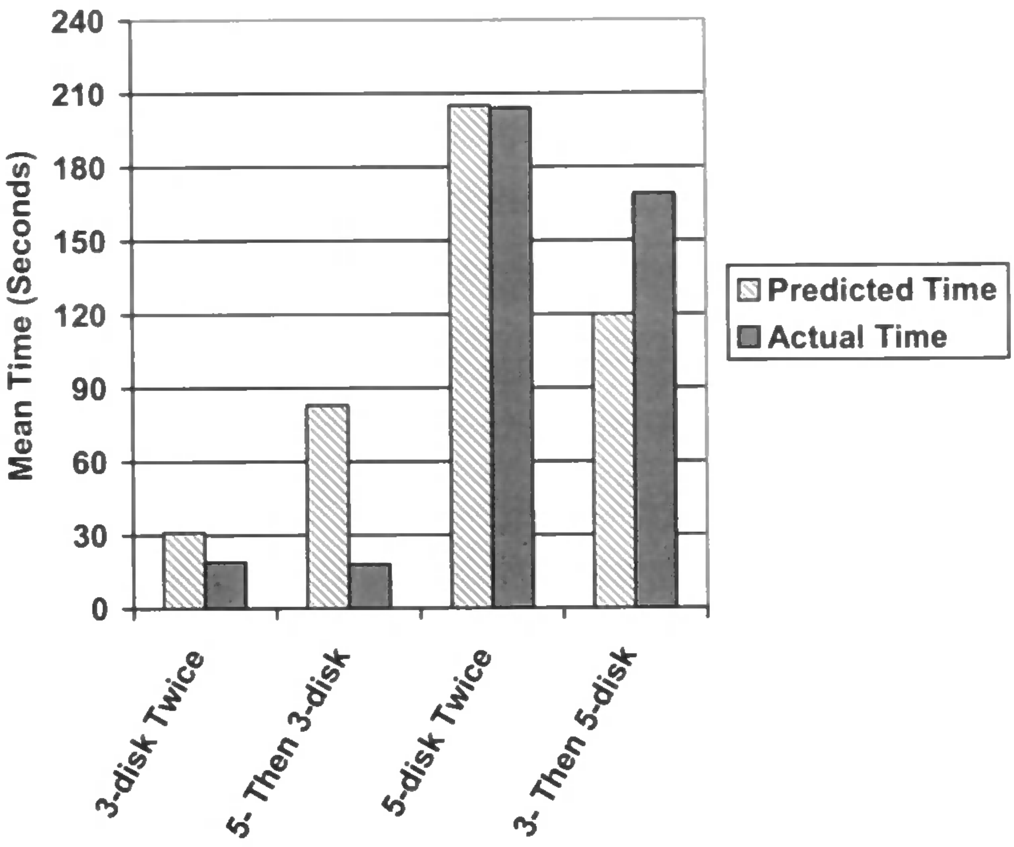

Figure 3.2

Bar graph of predicted and actual completion times per task experience condition (data from the second task)

In order to determine how many participants predicted task duration using whole minutes, frequency distributions of time estimates were calculated. Histograms of time estimates on the three-disk and five-disk tasks (regardless of task experience condition) are presented below in Figures 3.3 and 3.4, respectively.

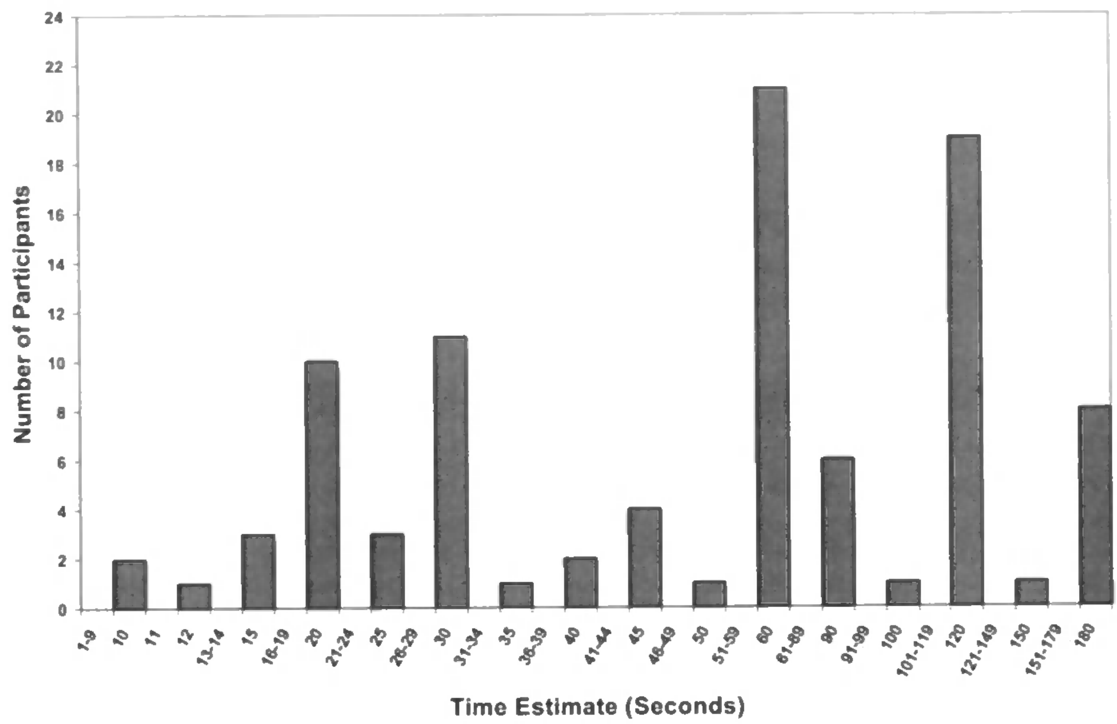

Figure 3.3

Histogram of time estimates on the three-disk task

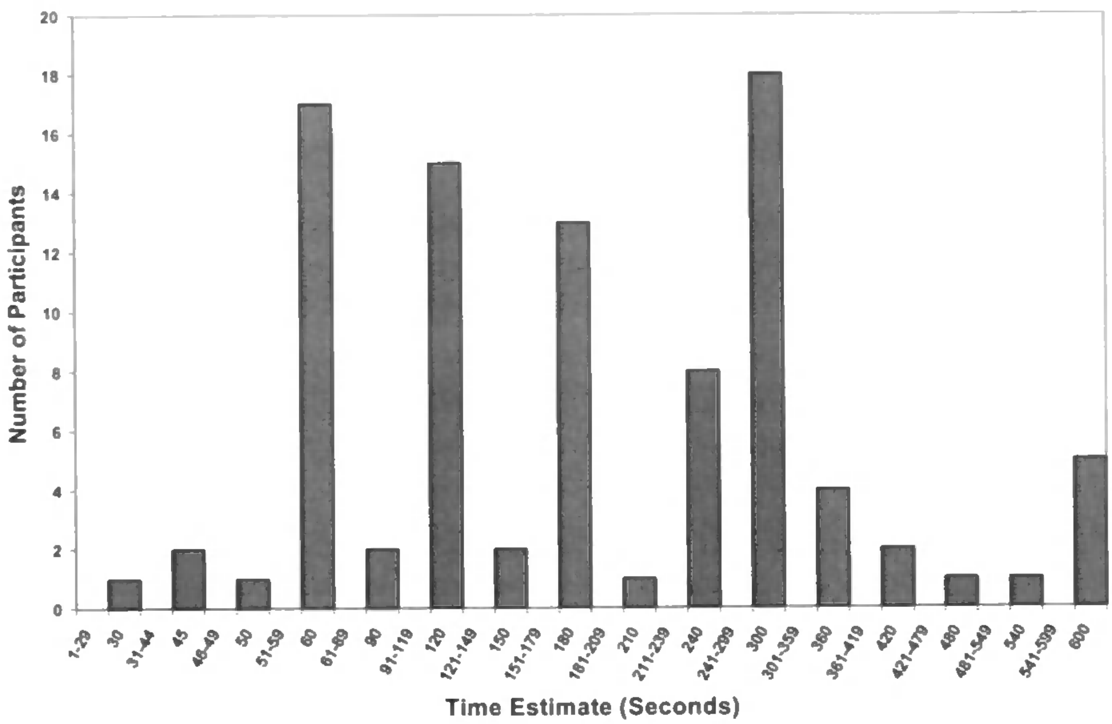

Figure 3.4

Histogram of time estimates on the five-disk task

For the same reasons that were specified in Experiment 1, the estimated and actual completion time data were subjected to a logarithmic transformation before being statistically analysed. In order to address the issue of task complexity, the log-transformed data from the first trial were analysed. Since the task experience factor was only relevant to the second trial, the data from the first trial were collapsed across the two conditions that performed either the three-disk or the five-disk task. A 2 (task: three-disk vs. five-disk) x 2 (time: estimated vs. actual) mixed design ANOVA produced a main effect of time, $F(3,90) = 27.89$, $MSE = .32$, $p < .001$, with estimates exceeding completion times (Ms = 169.54 and 139.94 seconds, respectively). This finding indicates that general temporal overestimation was evident on the first trial. There was also a main effect of task, $F(3,90) = 226.70$, $MSE = .53$, $p < .001$, with overall time being shorter on the three-disk task.

The interaction was also significant, $F(3,90) = 38.00$, $MSE = .32$, $p < .001$ (see Figure 3.5 below). This revealed that the difference between estimates and completion times was greater on the three-disk task (Ms = 82.75 and 31.90 seconds, respectively) than on the five-disk task (Ms = 256.33 and 247.99 seconds, respectively). Planned comparisons (LSD t-tests, both two-tailed) revealed a significant difference between estimated and actual duration on the three-disk ($p < .05$), but not the five-disk task ($p > .10$). Thus, the extent of temporal overestimation was less on the five-disk task, suggesting that task complexity might be positively related to time prediction accuracy.

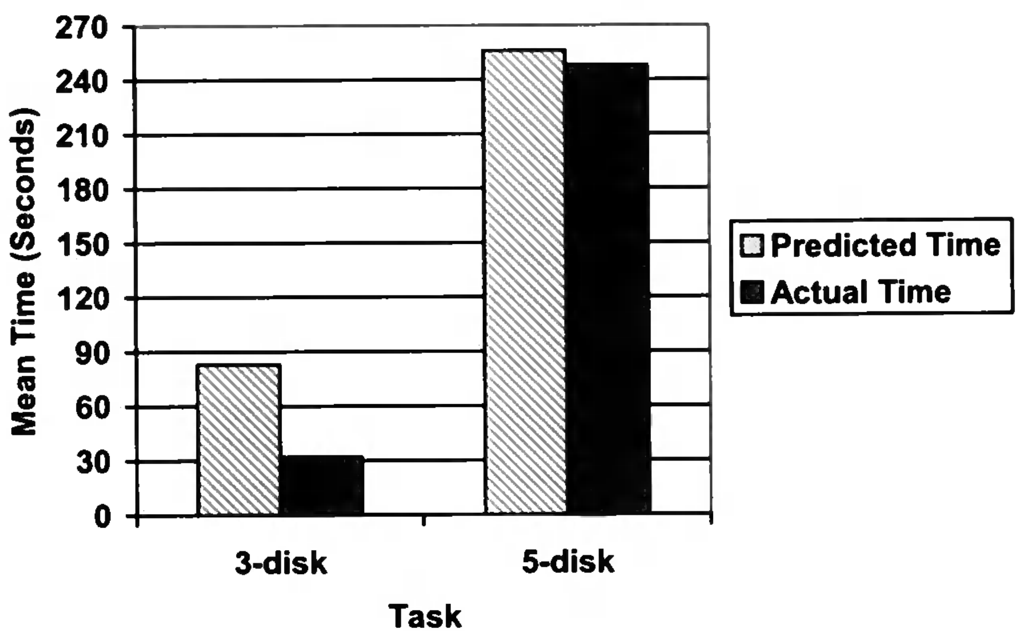

Figure 3.5

Bar graph depicting the interaction between the task and time factors (data from the first task partially collapsed across the task experience factor)

In order to address the issue of prior task experience, the log-transformed data from the second trial were analysed. A 2 (time) x 4 (task experience) mixed design ANOVA produced a main effect of time, $F(1,90) = 20.92$, $MSE = .29$, $p < .001$, with estimates exceeding completion times (Ms = 109.48 and 102.61 seconds, respectively). Thus, there was evidence of general temporal overestimation on the second trial. There was also a main effect of task experience, $F(3,90) = 120.64$, $MSE = .41$, $p < .001$, with overall time being longest in the five-disk task twice condition. Scheffé pairwise comparisons revealed significant differences between the means of all conditions ($ps < .05$) except those of the five-disk task twice and the three- then five-disk task conditions ($p > .05$).

The interaction was also significant, $F(3,90) = 25.20$, $MSE = .29$, $p < .001$ (see Figure 3.2 above). This revealed that temporal overestimation was evident in both conditions on the three-disk task, whereas the direction of time prediction bias on the five-disk task differed according to prior experience. That is, temporal underestimation was evident when the three-disk task was performed beforehand, whereas temporal overestimation occurred when the five-disk task was performed for a second time. Planned comparisons (LSD t-tests, all two-tailed) revealed that estimates and completion times differed significantly when the five-disk task was preceded by the three-disk task ($p < .05$), but not when it was performed for a second time ($p > .10$). On the three-disk task, the difference between estimated and actual duration was greater when the five-disk task ($p < .001$) rather than the three-disk task was performed beforehand ($p < .05$). These findings indicate that, on the three-disk and five-disk tasks, time predictions were more accurate when the same rather than a different version of the Tower of Hanoi task had been performed beforehand.

3.6 Discussion

Consistent with the task experience hypothesis, time predictions on the second trial were found to be more accurate when the same rather than a different version of the Tower of Hanoi was performed beforehand. A plausible explanation for this finding is that participants mentally represented the task components during the first trial, and drew upon pertinent information such as the disk moves needed for task completion when making a second time estimate. Such a judgement strategy would be expected to result in greater time prediction accuracy when an identical rather than a different task had just been completed. This finding suggests that the effect of prior task performance that was observed in Experiment 1 generalises beyond the three-disk task to the more complex five-disk version of the Tower of Hanoi task.

The present study suggests that prior experience of performing a more or less complex similar task influenced time prediction bias. That is, temporal overestimation was greater when the three-disk task was preceded by the five-disk task, whereas temporal underestimation was only evident when the five-disk task was performed after the three-task. This finding is consistent with the notion that participants used a judgement strategy whereby information about the previous task formed the basis of a second time prediction, which was insufficiently adjusted according to the greater or lesser demands of the current task.

Given the difference in complexity between the three-disk and five-disk tasks, time prediction bias would be a consequence of using such a judgement strategy. For example, over-optimism would occur if time predictions were based on the disk moves needed to complete the three-disk task, but were insufficiently scaled up to take account of the greater number of disk moves involved in finishing the five-disk task. Since a similar finding was observed in Experiment 3, the issue of prior experience of performing a different version of the same task was further addressed in Experiment 5.

There was evidence that, on the first trial, time estimates were less biased on the more complex five-disk task. This finding is broadly consistent with previous research (e.g., Suantak et al., 1996), which suggests that bias in non-temporal judgements of performance differs according to task complexity. Since time predictions on the five-disk task were less pessimistically biased, participants presumably displayed greater confidence (as well as greater accuracy) when judging the duration of the more complex task. This finding not only suggests that task complexity might mediate time prediction accuracy, but also highlights the potential applicability of the difficulty effect (Griffin & Tversky, 1992) to the process of estimating task duration.

As was the case in Experiments 1 to 3, temporal overestimation was evident on the three-disk task regardless of the order in which it was performed. However, a general optimistic time prediction bias did not occur on the five-disk task. In fact, temporal underestimation was only observed when this more complex task was preceded by the three-disk task, whereas temporal overestimation was evident when it was performed first. This finding suggests that increasing the cognitive complexity of the Tower of Hanoi task did not result in general temporal underestimation.

Task duration could explain the discrepancy between the present findings and those of previous research where the planning fallacy was evident (e.g., Buehler et al., 1994). That is, although the average duration of the five-disk task was almost three times that of the four-disk task in Experiment 3 (Ms = 217.31 and 78.42 seconds, respectively), it still took less time to complete than the laboratory tasks used previously. Thus, general temporal underestimation might only occur on laboratory tasks that are of longer duration than the tasks used in Experiments 1 to 4. Indeed, it has been found that people tend to use temporal units such as 10 or 15 minutes when estimating the duration of everyday activities (Fraisse, 1984).

Extrapolating from the work of Fraisse, it could be that the propensity to use such temporal units results in an overestimation of time on tasks that are of shorter duration than these units. Given the duration of the three-disk, four-disk and five-disk tasks, time predictions would be expected to be pessimistically biased as a consequence of using temporal units such as five or 10 minutes. Consistent with this suggestion, the frequency distribution of time estimates on the five-disk task revealed that several participants judged the duration of this task as being either five or 10 minutes. This finding provides support for the notion that the tendency to estimate the duration of everyday activities using such temporal units might generalise to this well structured laboratory task.

Since more complex tasks tend to be of longer duration than simpler ones (Buckhalt & Oates, 2002), the absence of a general optimistic time prediction bias could also be due to the relative simplicity of the three-disk, four-disk and five-disk tasks. Specifically, because these tasks were relatively undemanding, participants may well have been able to devote sufficient cognitive resources to the process of predicting task duration. For example, participants might have worked out the disk pattern involved in completing the three-disk task beforehand, and used this information when estimating its duration. On the basis of previous research (e.g., Sniezek et al., 1990), predictions would not be expected to optimistically biased as a consequence of cognitively processing task-related information. A key aim of Experiment 5 was to ascertain whether task duration or cognitive complexity was responsible for the absence of general temporal underestimation in Experiments 1 to 4.

3.7 Experiment 5

This study sought to determine whether task duration or cognitive complexity was responsible for the lack of a general optimistic prediction bias in Experiments 1 to 4. In order to achieve this goal, task duration was manipulated whilst cognitive complexity was held constant. A repetitive manual disk movement task was employed, which used the Tower of Hanoi task apparatus. Duration was manipulated by employing two versions of this task, which involved moving either three or five disks for a certain number of times. Unlike the Tower of Hanoi task, there were few restrictions on disk movement in the repetitive task, which was thus cognitively undemanding and simple to perform. Moreover, because there were few restrictions on disk movement (e.g., larger disks could be placed on top of smaller ones), there was no need for participants to create subgoals in order to complete this task. Hence, the three-disk and five-disk versions of the repetitive task did not vary in cognitive complexity, but differed principally in the amount of time needed for completion.

If task complexity is an important determinant of time estimation bias, then participants might be over-optimistic about their ability to complete such a simple repetitive task. That is, few cognitive resources are likely to be devoted to predicting the duration of this rather mundane task, a judgement strategy that has been shown to result in an optimistic bias (e.g., Arkes et al., 1988). Alternatively, if task duration is an important factor, then the same general temporal overestimation should be evident on the repetitive task as was observed on the tasks used in Experiments 1 to 4. Specifically, general temporal overestimation occurred because the present tasks were of shorter duration than the laboratory tasks used in previous research where the planning fallacy was evident (e.g., Byram, 1997). Task performance order was also manipulated in order to ascertain whether prediction bias on the second task differed according to the relative duration of the first task.

The repetitive task involved moving the disks one at a time from the left-hand peg to the centre peg to the right-hand peg and back again in reverse order. This process was repeated until all disks had been placed on the pegs for a certain number of times. The two repetitive tasks were designed to be of similar duration to the three-disk and five-disk versions of the Tower of Hanoi task (when they were performed first) in Experiment 4 (Ms = 31.90 and 247.99 seconds, respectively). These mean completion times served as a benchmark for the number of times that the disks had to be stacked on the pegs. Pilot testing revealed that, on the three-disk task, all disks could be placed on eight pegs in 32 seconds. On the five-disk task, pilot testing revealed that all disks could be placed on 36 pegs in 248 seconds. Completion of the three-disk and five-disk tasks thus involved stacking all of the disks on either eight or 36 pegs, respectively.

3.8 Method

3.8.1 Participants

Fifty (44 female and 6 male) students at the University of Plymouth participated voluntarily in partial fulfilment of a psychology course requirement. No biographical information other than gender was recorded.

3.8.2 Materials

These were identical to the materials used in Experiment 4.

3.8.3 Design and Procedure

A 2 (time: estimated vs. actual) x 2 (task: three-disk vs. five-disk) x 2 (task performance order: three-disk then five-disk vs. five-disk then three-disk) mixed factorial design was used. The time and task factors were repeated-measures. The order in which the tasks were performed was fully counterbalanced. The task performance order factor was manipulated between groups, with participants being randomly assigned to one of two equal-sized conditions.

Except for the following differences, the procedure was identical to that of Experiment 4. The researcher recorded the number of pegs on which the disks were stacked and informed participants when each task had been completed. All participants were informed about the number of pegs upon which the disks had to be placed before estimating task duration. Each testing session lasted approximately 20 minutes.

3.9 Results

Table 3.2 below contains basic descriptive statistics of estimated and actual duration per task and task performance order condition. All test statistics used in Experiment 5 are contained in Appendix 5.

Task		Task Performance Order Condition	
		Three- Then Five-disk	Five- Then Three-disk
Three-disk	Time Estimate	M = 59.60 Mdn = 60.00 SD = 30.92 N = 25	M = 63.20 Mdn = 60.00 SD = 41.05 N = 25
	Completion Time	M = 30.48 Mdn = 30.00 SD = 4.62 N = 25	M = 26.08 Mdn = 24.00 SD = 6.72 N = 25
Five-disk	Time Estimate	M = 184.00 Mdn = 180.00 SD = 74.89 N = 25	M = 255.60 Mdn = 240.00 SD = 106.46 N = 25
	Completion Time	M = 210.68 Mdn = 207.00 SD = 38.89 N = 25	M = 228.36 Mdn = 217.00 SD = 47.44 N = 25

Table 3.2

Time estimates and completion times per task and task performance order condition (in seconds)

As Table 3.2 shows, temporal estimates exceeded completion times on the three-disk task regardless of the order in which it was performed. This finding suggests that participants made pessimistically biased predictions on the shorter duration repetitive task.

However, Table 3.2 shows that the direction of time prediction bias differed according to the order in which the five-disk task was performed. That is, temporal underestimation was evident when this task was performed second, whereas temporal overestimation occurred when it was performed first. A bar graph of predicted and actual completion times per task and task performance order condition is presented below in Figure 3.6.

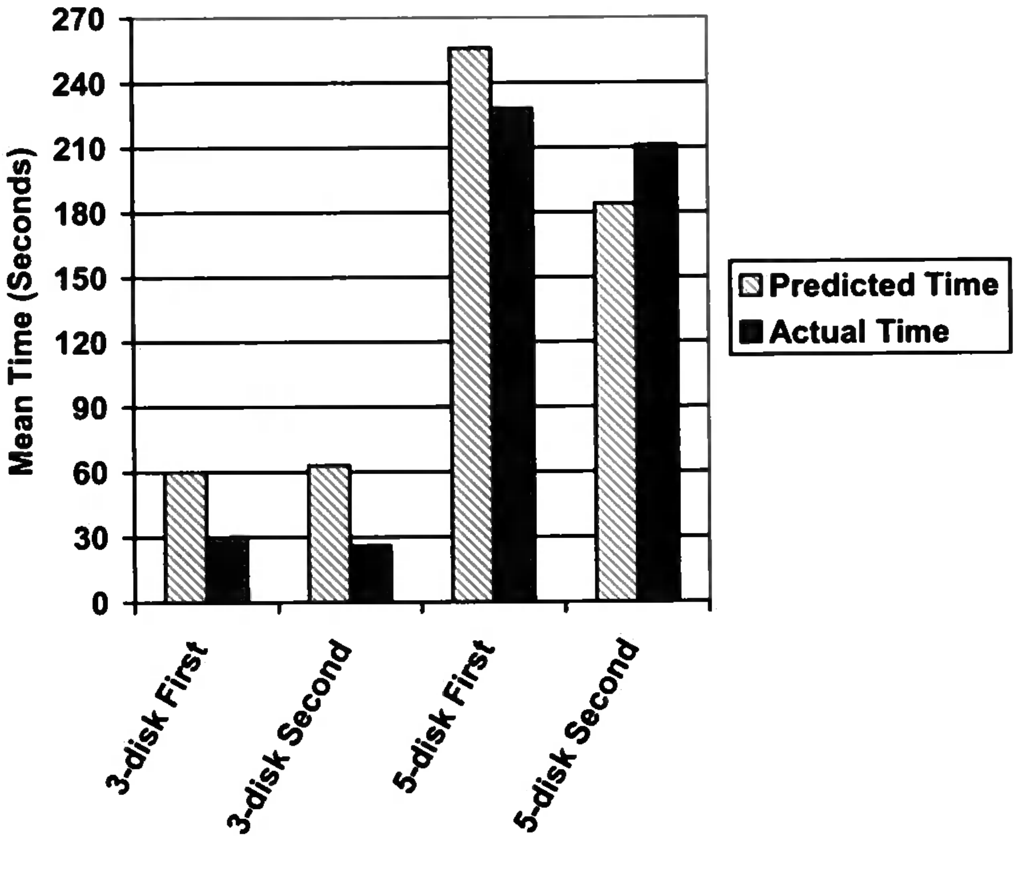

Figure 3.6

Bar graph of estimated and actual completion times per task and performance order condition

In order to determine how many participants predicted task duration using whole minutes, frequency distributions of time estimates were calculated. Histograms of time estimates on the three-disk and five-disk tasks (regardless of performance order) are presented below in Figures 3.7 and 3.8, respectively.

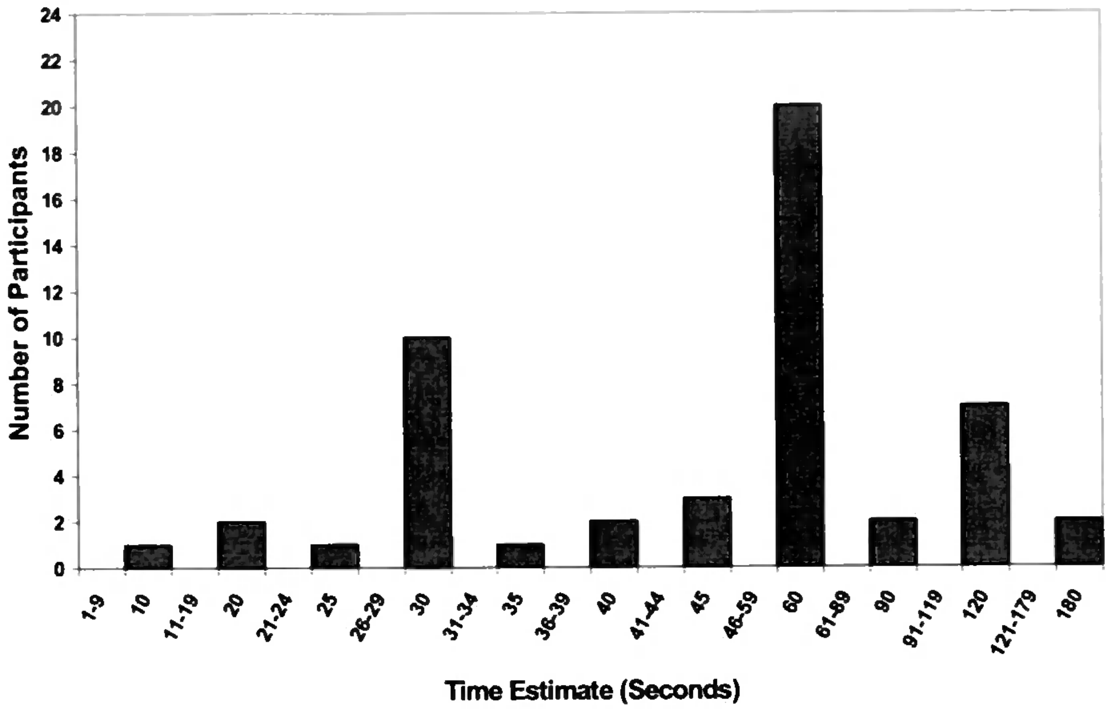

Figure 3.7

Histogram of time estimates on the three-disk task

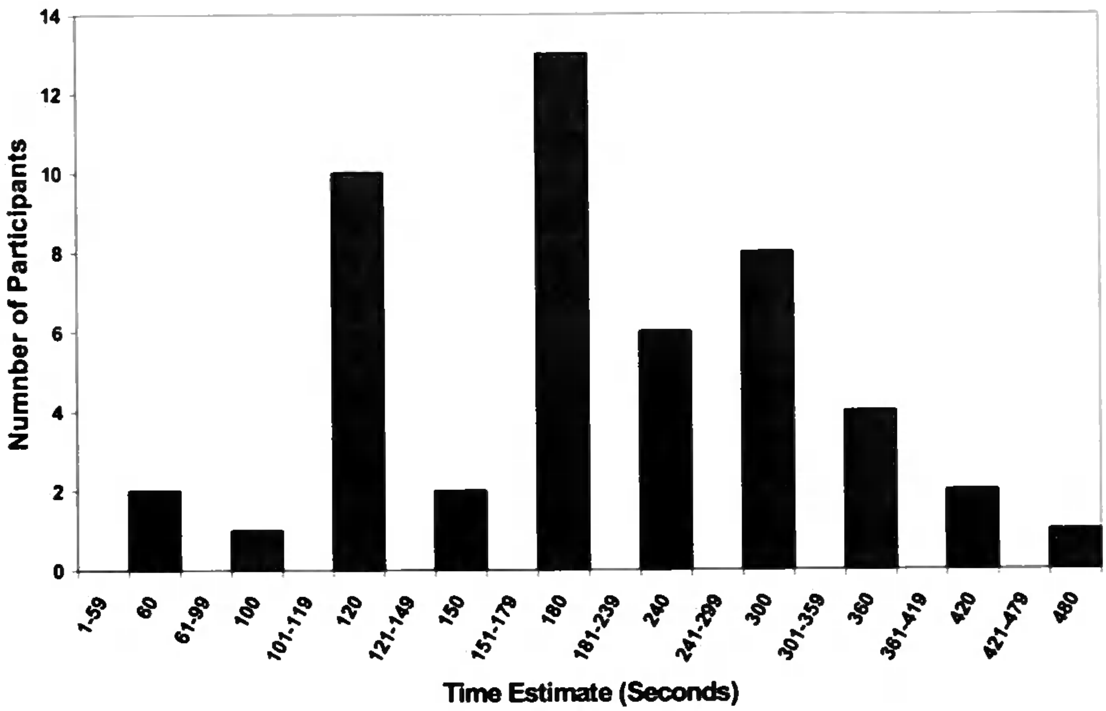

Figure 3.8

Histogram of time estimates on the five-disk task

For the same reasons that were specified in Experiment 1, the estimated and actual completion time data were subjected to a logarithmic transformation prior to being statistically analysed. In order to address the issue of task duration versus cognitive complexity, the log-transformed data from the first trial were analysed. A 2 (time) x 2 (task) mixed design ANOVA produced a main effect of time, $F(1,48) = 21.39$, $MSE = .11$, $p < .001$, with estimates exceeding completion times (Ms = 157.60 and 129.42 seconds, respectively). This finding suggests that general temporal overestimation was evident on the first trial. There was also a main effect of task, $F(1,48) = 501.72$, $MSE = .15$, $p < .001$, with overall time being longer on the five-disk task.

The interaction was also significant, $F(1,48) = 14.03$, $MSE = .11$, $p < .001$, and revealed that the extent of temporal overestimation was greater on the three-disk task (see

Figure 3.9 below). Planned comparisons (LSD t-tests, both two-tailed) revealed that estimates and completion times differed significantly on the three-disk ($p < .05$), but not the five-disk task ($p > .10$). This finding indicates that time predictions were less pessimistically biased (and thus more accurate) on the longer duration version of the repetitive disk movement task.

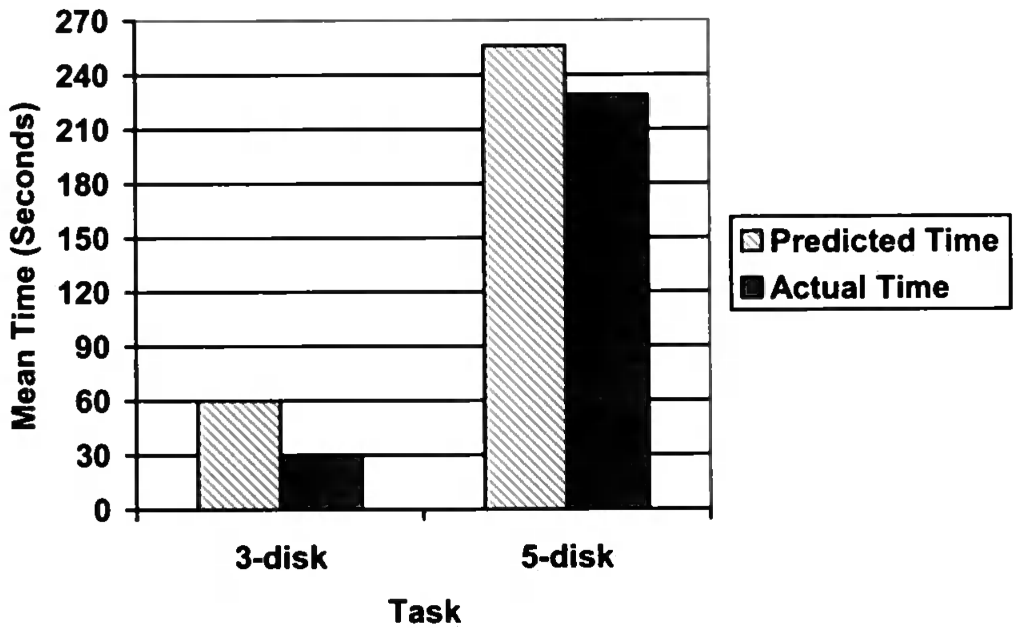

Figure 3.9

Bar graph depicting the interaction between the task and time factors (data from the first task)

In order to examine the impact of task performance order on time estimation bias, the log-transformed data from the second trial were analysed. A 2 (time) x 2 (task) mixed design ANOVA produced a main effect of time, $F(1,48) = 16.60$, $MSE = .12$, $p < .001$, with estimates exceeding completion times (Ms = 123.60 and 118.38 seconds, respectively). This finding suggests that general temporal overestimation was evident on the second trial. There was also a main effect of task, $F(1,48) = 424.13$, $MSE = .16$, $p < .001$, with overall time being longer on the five-disk task.

The interaction was also significant, $F(1,48) = 48.84$, $MSE = .12$, $p < .001$. This revealed that estimates exceeded completion times on the three-disk task, whereas completion times exceeded estimates on the five-disk version (see Figure 3.10 below). Planned comparisons (LSD t-tests, both two-tailed) revealed a significant difference between estimated and actual duration on each task ($ps < .05$).

This finding indicates that the direction in which time predictions on the second task were biased differed according to the relative duration of the first task. That is, an optimistic time prediction bias was evident on the longer task, whereas a pessimistic time prediction bias occurred on the shorter task. However, whilst the difference between predicted and actual duration was greater when the three-disk task was performed second (see Table 3.2), time estimates (log-transformed) did not differ significantly according to performance order on this task, $t(48) = .24$, $p > .10$, two-tailed. Thus, time predictions were not significantly longer when the shorter task was performed after the longer task rather than before it.

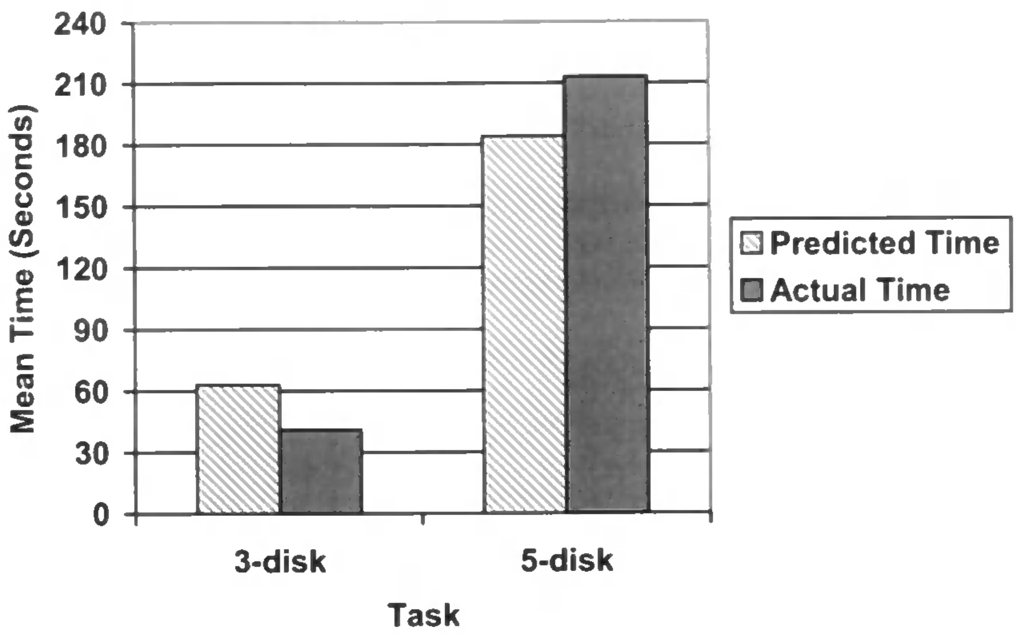

Figure 3.10

Bar graph depicting the interaction between the task and time factors (data from the second task)

3.10 Discussion

There was no evidence of a general optimistic time prediction bias on either version of the repetitive disk movement task. Specifically, time predictions were pessimistically biased on the three-disk task regardless of performance order, and there was a hint of temporal overestimation on the five-disk task when it was performed first. In fact, temporal underestimation was only evident when the five-disk task was preceded by the three-disk task. Given that the three-disk and five-disk versions of the repetitive task did not differ in cognitive complexity, the absence of general temporal underestimation suggests that task duration is responsible for the findings of the present research. That is, the present lack of general temporal underestimation could be due to the shorter duration of the tasks used in

Experiments 1 to 5 relative to those tasks employed in previous research where the planning fallacy has been observed (e.g., Byram, 1997).

Given this suggestion, the direction in which time predictions are biased might differ according to task duration. That is, temporal overestimation occurs on laboratory tasks that take less than one minute to complete (e.g., the three-disk repetitive task), whereas temporal underestimation may prevail on such tasks when they are of longer duration than those used in Experiments 1 to 5. Consistent with this notion, an optimistic time prediction bias has been observed on laboratory tasks that took between eight minutes (Buehler et al., 1997) and just over one hour to complete (Byram, 1997). Thus, the planning fallacy might only be evident on laboratory tasks that are of longer duration than the present ones. In order to test the validity of this claim, laboratory tasks that were of longer duration than those used in Experiments 1 to 5 were employed in the remainder of this research programme.

The impact of task duration on the accuracy of time predictions was emphasised in the present study. That is, on the first trial, the extent of temporal overestimation was greater on the three-disk task relative to the five-disk task. This finding indicates that, after only 20 seconds of pre-exposure, participants were more accurate (i.e., less pessimistic) when estimating the duration of the longer task.

A possible explanation for this finding is that greater cognitive resources were devoted to the process of estimating the duration of the five-disk task. Since participants were aware in advance that this task involved placing all disks on each peg 12 times (i.e., 36 pegs in total), they presumably realised that it could not be completed in a matter of seconds. With such information in mind, they may have used the pre-exposure period to consider the amount of work involved in the task completion process, and based their second time estimate on such information. Conversely, as the three-disk task involved placing the disks on each peg only two or three times (i.e., eight pegs in total), few cognitive resources may have been devoted to estimating the duration of this quick and simple task. On the basis of previous research (Koriat et al., 1980), judgements would be expected to be more accurate when greater cognitive resources are devoted to the process of predicting task performance.

There was some evidence that time estimation bias on the second task differed according to the relative duration of the first task. Specifically, temporal underestimation was only evident on the five-disk task when it was performed second, whereas there was a hint of greater temporal overestimation on the three-disk task when it was performed second. This finding is broadly consistent with those of Experiments 3 and 4, and suggests that information about the first task might have formed the basis of time estimates on the second task. For example, when making a second time prediction, participants may have estimated the duration of the just-completed task, but failed to take sufficient account of the greater or lesser number of disks involved in completing the current task. Given the findings of Experiments 3 to 5, the impact of previous task duration on subsequent time estimates was further examined in the remainder of this research programme.

3.11 General Discussion

There was further evidence of temporal overestimation on the three-disk Tower of Hanoi task regardless of task performance order or prior task experience in Experiment 4. Likewise, Experiment 5 revealed that temporal overestimation was evident on the three-disk version of the repetitive task regardless of the order in which it was performed. Moreover, there was a hint of general temporal overestimation on the five-disk Tower of Hanoi task (Experiment 4) and the five-disk repetitive task (Experiment 5). These findings suggest that the temporal underestimation indicative of the planning fallacy does not occur on well structured laboratory tasks, which range in duration from about 30 seconds to four minutes. However, there was some evidence of an optimistic time prediction bias on both of the five-disk tasks, but only when they were performed second. That is, participants tended to underestimate their completion times only when they had completed the three-disk version of the same task beforehand.

The present studies revealed that, on the first trial, participants were more accurate at estimating the duration of the longer tasks. For example, time predictions were less pessimistically biased on the five-disk version of the Tower of Hanoi relative to the three-disk task (Experiment 4). As temporal estimates were also more accurate on the four-disk version of the Tower of Hanoi (Experiments 2 and 3), the present studies provide further evidence of a positive relationship between task complexity and time prediction accuracy. There was also evidence that temporal estimates were less pessimistically biased on the five-disk version of the repetitive task (Experiment 5), suggesting that task duration might be positively related to time prediction accuracy. Whilst these findings could be due to participants devoting greater cognitive resources to predicting the duration of the longer tasks, an alternative explanation is that people tend to estimate time in whole minutes rather than seconds, or longer temporal units such as 10 or 15 minutes (Fraisse, 1984).

Given the duration of both types of three-disk and five-disk task used here (i.e., approximately 30 seconds and four minutes, respectively), the tendency to estimate time in whole minutes may have resulted in greater time prediction accuracy on the longer tasks. For example, participants might have given estimates of four or five minutes on these tasks. Conversely, greater time prediction bias may have been a consequence of estimating the duration of the shorter tasks using temporal units such as one or two minutes. Although participants were asked to estimate task duration in seconds, several of them gave time predictions in whole or part minutes, which they then converted into seconds at the request of the researcher.

Consistent with the notion that people tend not to judge task duration in seconds, the frequency distributions of time estimates revealed that the majority of the present participants predicted the duration of all tasks in units of whole minutes. For example, over half of the participants estimated the duration of the three-disk Tower of Hanoi task (Experiment 4) and the shorter duration repetitive disk movement task (Experiments 5) using temporal units of whole minutes (see Figures 3.3 and 3.7, respectively). Similarly, on the five-disk Tower of Hanoi task (Experiment 4) and the longer duration repetitive disk movement task (Experiment 5), over ninety percent of participants estimated their completion times using whole minutes (see Figures 3.4 and 3.8, respectively). Hence, there was evidence that individuals estimated the duration of both types of three-disk and five-disk task using temporal units of whole minutes rather than seconds. Indeed, on the five-disk Tower of Hanoi task (Experiment 4), a sizeable minority of participants estimated task duration in temporal units of five and 10 minutes. Likewise, some participants predicted that the longer duration repetitive disk movement task would take them five minutes to complete (Experiment 5). Thus, it seems that people use such temporal units when judging the duration of certain laboratory tasks.

Experiment 5 revealed that task duration could be responsible for the discrepancy between the findings of the present studies and those of previous research where the planning fallacy was evident (e.g., Josephs & Hahn, 1995). That is, a general optimistic time prediction bias might only prevail on longer duration tasks than those used in Experiments 1 to 5. Given that the shortest task used in previous research took about eight minutes to complete (Buehler et al., 1997), tasks may need to be of at least this duration before general temporal underestimation occurs. In order to test this suggestion, a laboratory task that was of similar duration to some of those used in previous studies (e.g., Francis-Smythe & Robertson, 1999) was employed in Experiments 6 and 7.

The present lack of general temporal underestimation could also be due to the nature of the tasks used in Experiments 1 to 5. For example, whilst the Tower of Hanoi task has been the subject of research into task planning (e.g., Goel & Grafman, 1995; Karat, 1982; Spitz et al., 1984), it is a rather novel task. Thus, it could be that participants erred on the side of caution when making a time estimate because they were unsure of the task completion process. Moreover, as the Tower of Hanoi task and the repetitive task are both novel, participants presumably had little or no experience of performing them before the experiment.

In contrast, in previous research, participants may be been over-optimistic when predicting the duration of tasks such as writing an essay (e.g., Koole & Van't Spijker, 2000) because they were familiar with the completion process. Hence, people might be over-optimistic when estimating the duration of a familiar task, but pessimistic when judging the duration of a novel activity. Given this suggestion, it was decided to increase task familiarity in the research presented in the next chapter of this thesis. To this end, laboratory tasks that are more realistic than the Tower of Hanoi task and the repetitive disk movement task were employed in Experiments 6 and 7.

The importance of prior task experience in mediating time prediction bias was further emphasised in the present studies. For example, time estimates were more accurate when the three-disk and five-disk versions of the Tower of Hanoi task were performed for a second time (Experiment 4). This finding is consistent with previous research (Josephs & Hahn, 1995), and suggests that information about personal performance on a previous task can be used to good effect when it is acquired shortly before estimating task duration.

Given the brief temporal interval between the first and second trials in the present studies, participants were likely to recall information about the just-completed task (e.g., the correct disk move pattern) when making a second time estimate. Likewise, participants may have thought about any errors that they committed on the first trial (e.g., moving a disk to the wrong peg), and incorporated such information into their second time prediction. In the light of the research reported by Taylor et al. (1998), thinking about factors that might delay task completion would be expected to reduce temporal misestimation.

Prior experience of performing a similar task was also found to influence time prediction bias in the present studies. That is, an optimistic time prediction bias was only evident when either of the five-disk tasks were performed second (Experiments 4 and 5), whereas temporal overestimation was greater on the three-disk tasks when a more complex (Experiment 4) or a longer duration version of the same task was completed initially (Experiment 5). This finding is consistent with that of Experiment 3, and suggests that participants took account of their performance on the previous task when making a second time estimate.

A possible explanation for these findings is that time predictions on the second task were based on information about the first task. Given that the anchoring and adjustment heuristics (Tversky & Kahneman, 1982b) have been shown to influence temporal judgements (Buehler et al., 1995), participants might have used these heuristics when making a second time estimate. For example, in Experiment 4, over-optimism may have been a consequence of time predictions being anchored on the shorter duration of the just-completed three-disk task, with insufficient adjustment for the greater demands of the upcoming five-disk task.

In relation to the time estimation process, evidence of anchoring effects implies that people base their predictions about their performance on a current task on what happened on a previous task. However, in contrast to the standard anchoring paradigm (Chapman & Johnson, 2002), no numerical anchor values were presented to participants in the present studies. Thus, if anchoring and adjustment occurred in these experiments, numerical anchor values would have to be self-generated. Since participants generated a numerical value before each task (i.e., a time estimate), they might also have done so at the end of each trial. That is, they may have estimated how long the just-completed task took to complete, and used this information as a basis for their second estimate.

Alternatively, participants may have recalled their time prediction on the first task, and used this numerical value as an anchor for estimating the duration of the second task. Whilst these suggestions are rather speculative, there is evidence of anchoring and adjustment in the absence of externally-presented numerical values (Epley & Gilovich, 2002). Hence, it could be that participants misestimated the duration of the second task because they used these cognitive heuristics.

3.12 Conclusions

A number of potentially important findings have emerged from the research presented in this chapter. There was no evidence of a general optimistic time prediction bias on the five-disk Tower of Hanoi task (Experiment 4). Hence, increasing the complexity of this well structured task did not result in general temporal underestimation. In addition to providing further evidence that prior task experience mediates time estimation accuracy (Experiment 4), task duration was also identified as a potentially important determinant of prediction bias (Experiment 5). Specifically, general temporal underestimation was not evident on three-disk and five-disk versions of a simple repetitive task, suggesting that the shorter duration of the present tasks could be responsible for the absence of a general optimistic time prediction bias in Experiments 1 to 5. This finding indicates that general temporal underestimation does not occur, and seems to be reversed, on well structured laboratory tasks that take up to four or five minutes to complete.

The next chapter of this thesis comprises two studies. A key aim of Experiments 6 and 7 was to ascertain whether general temporal underestimation occurred on a laboratory task that was of longer duration than those used in Experiments 1 to 5. In order to achieve this goal, a task that took approximately 11 minutes to complete was employed alongside tasks that were of similar duration to the ones used in the present studies. The issue of prior task experience was further addressed in Experiments 6 and 7 by manipulating task performance order. These studies also employed tasks that are less artificial than the Tower of Hanoi task and the repetitive disk movement task. That is, tasks that bear greater resemblance to various well structured activities that are performed in everyday life, such as constructing self-assembly furniture (Byram, 1997) or cooking a meal by following a recipe (Byrne, 1977).

Chapter Four

Time Estimation Bias on Less Artificial Well Structured Laboratory Tasks

4.1 Overview

The research presented in this chapter consists of two experiments. A principal aim of Experiments 6 and 7 was to ascertain whether general temporal overestimation would occur on well structured laboratory tasks that are less artificial than the ones used in Experiments 1 to 5. That is, whether time predictions were pessimistically biased on more realistic tasks that were of similar duration to the three-disk and five-disk versions of the Tower of Hanoi task (and the repetitive disk movement task). The present studies also sought to determine whether general temporal underestimation would occur on a well structured task that was of longer duration than the ones used in Experiments 1 to 5. This task was also of similar duration to a number of the laboratory tasks used in previous research where an optimistic time prediction bias prevailed (e.g., Byram, 1997).

The impact of task duration on time prediction bias was further examined in Experiment 6, which employed two different tasks. One of these tasks took longer to complete than the ones used in Experiments 1 to 5, whereas the other was of similar duration to the five-disk Tower of Hanoi task. In order to reflect the design of Experiments 1 to 5, where different versions of the Tower of Hanoi task and the repetitive disk movement task were used, task structure was held constant in Experiment 7. In this study, three different versions of the same task were employed. Two of these tasks were shorter duration sub-component versions of the long duration task from Experiment 6, and were employed alongside the long task itself. The issue of prior task experience was further addressed by manipulating task performance order in both studies.

4.2 General Introduction to Experiments 6 and 7

As has been shown, a general optimistic time prediction bias did not occur on three versions of the Tower of Hanoi task (Experiments 1 to 4) and two versions of a repetitive disk movement task (Experiment 5). In general, there was evidence of temporal overestimation on the tasks used in Experiments 1 to 5. This finding contrasts with previous research (e.g., Josephs & Hahn, 1995), which has produced evidence of optimistically biased time estimates on other laboratory and real world tasks. Given the findings of Experiment 5, task duration was suggested as an explanation for the lack of a general optimistic time prediction bias in the present research. That is, general temporal underestimation might only be evident on longer duration tasks than those used in Experiments 1 to 5.

Consistent with this suggestion, the laboratory tasks used previously were of longer duration than those employed in Experiments 1 to 5, which took up to four or five minutes to complete. For example, Byram (1997) reports that the furniture assembly task took a little over one hour to complete, whereas it took about 10 minutes to finish the origami task. Likewise, the crossword puzzle used by Francis-Smythe and Robertson (1999) took about 10 minutes to complete, whereas the anagram task employed by Buehler et al. (1997) took about eight minutes to finish.

Whilst task duration might explain the findings of Experiments 1 to 5, the nature of the tasks used in those studies could also account for the absence of a general optimistic time prediction bias. That is, participants erred on the side of caution because they were unfamiliar with the completion process of the Tower of Hanoi task and the repetitive disk movement task. Given this suggestion, it could be that people are only over-optimistic when estimating the duration of tasks that are less novel than those used in Experiments 1 to 5. Moreover, although the Tower of Hanoi task has been the subject of research into task planning (e.g., Davies, 2000b; Goel & Grafman, 1995; Welsh et al., 1999), it bears little resemblance to various well structured activities that are performed in everyday life.

In contrast, some of the laboratory tasks used in previous research (e.g., Francis-Smythe & Robertson, 1999) were more akin to certain well structured everyday activities, suggesting that participants may have had some prior experience of completing the same or similar tasks. For example, participants in Byram's (1997) research might have been familiar with the process of assembling an item of flat-packed furniture. If this were the case, then temporal underestimation would be expected to occur if these individuals failed to take account of the fact that self-assembly furniture components tend not to fit together as easily as the task instruction booklet suggests.

Support for this notion comes from Byram's research, which revealed that participants failed to consider factors that might delay task completion when estimating the duration of a furniture assembly task. Hence, when predicting the duration of laboratory tasks that are identical or similar to tasks encountered previously, it could be that individuals fail to consider distributional information, a strategy that has been shown to lead to temporal underestimation (Buehler et al., 1994). Conversely, participants in Experiments 1 to 5 were unlikely to possess distributional information about novel tasks such as the Tower of Hanoi because of a lack of prior task experience. Thus, they may have erred on the side of caution when estimating task duration.

Given these suggestions, a key aim of Experiments 6 and 7 was to explore the time estimation process using well structured laboratory tasks that are more realistic than the Tower of Hanoi task and the repetitive disk movement task. The issue of task duration was also further addressed in the present studies, where a well structured laboratory task that took about 11 minutes to complete was employed alongside ones that were of similar duration to some of those used in Experiments 1 to 5.

The tasks used in Experiments 6 and 7 were toy construction kits manufactured by Playmobil®. Completing these tasks involved assembling miniature castles by slotting together moulded plastic components in accordance with pictorial information presented in a step-by-step instruction booklet. Since the Playmobil® tasks had to be completed in a sequential manner, they were well structured like the tasks used in Experiments 1 to 5. However, relative to the Tower of Hanoi task and the repetitive disk movement task, the Playmobil® tasks were more akin to some of the laboratory tasks used in previous research. For example, the furniture assembly task used by Byram (1997) also had to be completed sequentially, and involved fitting components together in accordance with information that was specified in an instruction booklet. Thus, many of the features of such real world tasks were also evident in the Playmobil® construction kits, suggesting that they were less artificial than the tasks used in Experiments 1 to 5.

4.3 Experiment 6

This study examined whether time estimation bias differed according to the duration of well structured laboratory tasks, which were more realistic than the ones used in Experiments 1 to 5. Task duration was manipulated by employing two different Playmobil® construction kit toys. One of these tasks was of similar duration to the five-disk Tower of Hanoi task, whereas the other was longer than the tasks used in the previous experiments. Pilot testing revealed that the long task took just over 11 minutes to complete, whereas the duration of the short task was a little over four and a quarter minutes (Ms = 667.87 and 258.43 seconds, respectively). Each Playmobil® task involved constructing a different miniature castle by slotting together moulded plastic components in accordance with instructions, which were presented in a step-by-step booklet. As would be expected, the number of plastic components and the number of steps necessary for task completion was greater on the long duration task.

As the short Playmobil® task was of similar duration to the five-disk Tower of Hanoi task (Experiment 4), it was hypothesised that general temporal overestimation would occur on this task. However, as the long Playmobil® task was of similar duration to some of the laboratory tasks used in previous research where an optimistic time prediction bias prevailed (e.g., Francis-Smythe & Robertson, 1999), general temporal underestimation was expected to occur on this task. Task performance order was manipulated in order to ascertain whether time prediction bias on the second task was influenced by the relative duration of the first task.

4.4 Method

4.4.1 Participants

Sixty-one (51 female and 10 male) students at the University of Plymouth participated voluntarily. Forty-nine participants took part in partial fulfilment of a psychology course requirement whilst the remainder were paid £2.50 each. No biographical information other than gender was recorded.

4.4.2 Materials

Two Playmobil® construction kits each accompanied by a step-by-step pictorial instruction booklet were used. The short duration task involved constructing a single-towered castle, whereas the long duration task involved building a multi-turreted castle with a surrounding jetty, platform and battlements. A digital stopwatch was used to measure task duration.

4.4.3 Design and Procedure

A 2 (time: estimated vs. actual) x 2 (task: short duration vs. long duration) x 2 (task performance order: short then long task vs. long then short task) mixed factorial design was used. The time and task factors were repeated-measures, with the order of task performance being fully counterbalanced. Task performance order was a between-groups factor, with participants being randomly assigned to one of two conditions.

The procedure was similar to that of Experiment 1. After participants were briefed about the experimental rationale, the researcher presented the instruction booklet for the first task, displayed the plastic components on the table, and then explained what the task involved. There then followed a planning period that was different for each task.

Pilot testing revealed that it took, on average, 40 seconds to read the instruction booklet for the short task and 80 seconds to read the booklet for the long task. Dependent on task performance order, participants were given these amounts of time to read the instruction booklet before estimating the duration of the current task. In order to ensure that all pages of the instruction booklet were previewed, participants were informed when 10 seconds of each planning period remained.

At the end of each planning period, participants were asked to close the instruction booklet temporarily and estimate task duration. Participants then re-opened the instruction booklet and began performing the task. The stopwatch was activated at this point and was stopped once each task had been completed. This procedure was repeated for the second task. Upon completion of the second task, participants were fully de-briefed about the experimental rationale. Each testing session lasted approximately 30 minutes.

4.5 Results

Table 4.1 below contains basic descriptive statistics of estimated and actual duration per task and task performance order condition. All test statistics used in Experiment 6 are contained in Appendix 6.

Task		Task Performance Order Condition	
		Short Then Long	Long Then Short
Short Duration	Time Estimate	M = 285.48 Mdn = 240.00 SD = 191.96 N = 31	M = 348.00 Mdn = 300.00 SD = 124.72 N = 30
	Completion Time	M = 233.94 Mdn = 212.00 SD = 87.12 N = 31	M = 244.90 Mdn = 224.00 SD = 73.34 N = 30
Long Duration	Time Estimate	M = 492.58 Mdn = 450.00 SD = 207.52 N = 31	M = 557.00 Mdn = 600.00 SD = 245.33 N = 30
	Completion Time	M = 642.48 Mdn = 617.00 SD = 167.26 N = 31	M = 696.07 Mdn = 656.00 SD = 222.71 N = 30

Table 4.1

Time estimates and completion times per task and task performance order condition (in seconds)

As Table 4.1 shows, temporal estimates exceeded completion times on the short duration Playmobil® task regardless of the order in which it was performed. In contrast, completion times exceeded temporal estimates regardless of the order in which the long duration Playmobil® task was performed. A bar graph of predicted and actual completion times is presented in Figure 4.1 below.

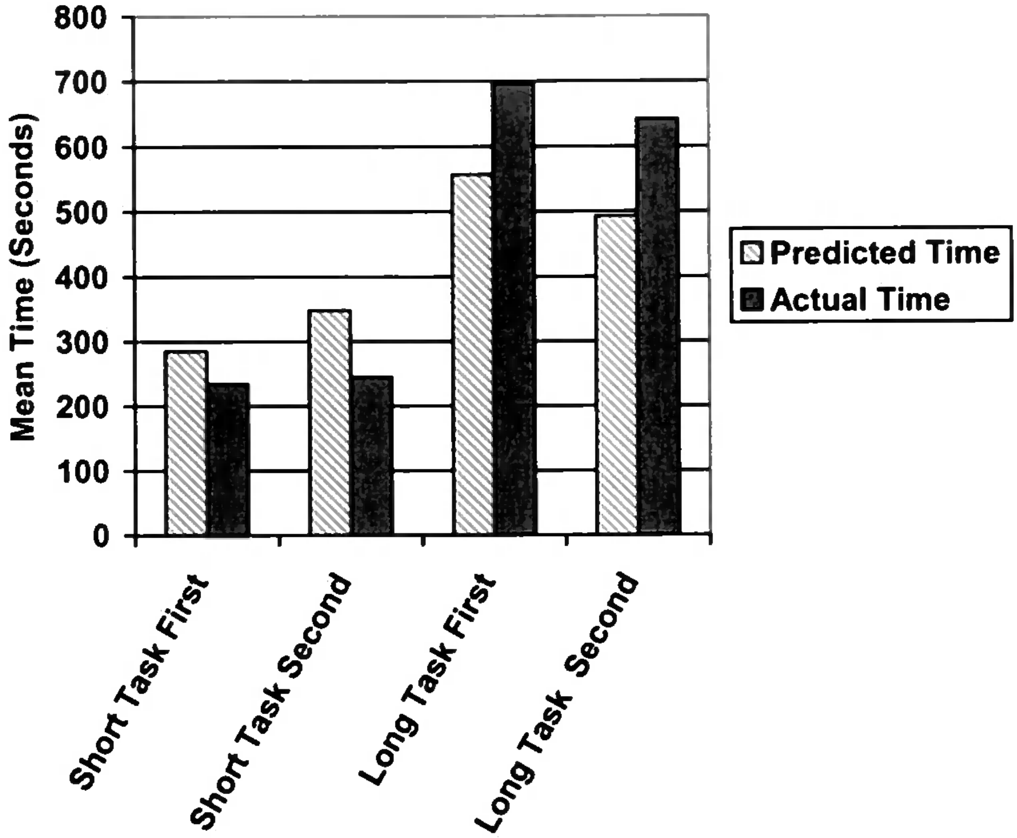

Figure 4.1

Predicted and actual completion times per task and task performance order condition

In order to determine how many participants predicted task duration using whole minutes, frequency distributions of time estimates were calculated. Histograms of time estimates on the short and long duration Playmobil® tasks (regardless of performance order) are presented below in Figures 4.2 and 4.3, respectively.

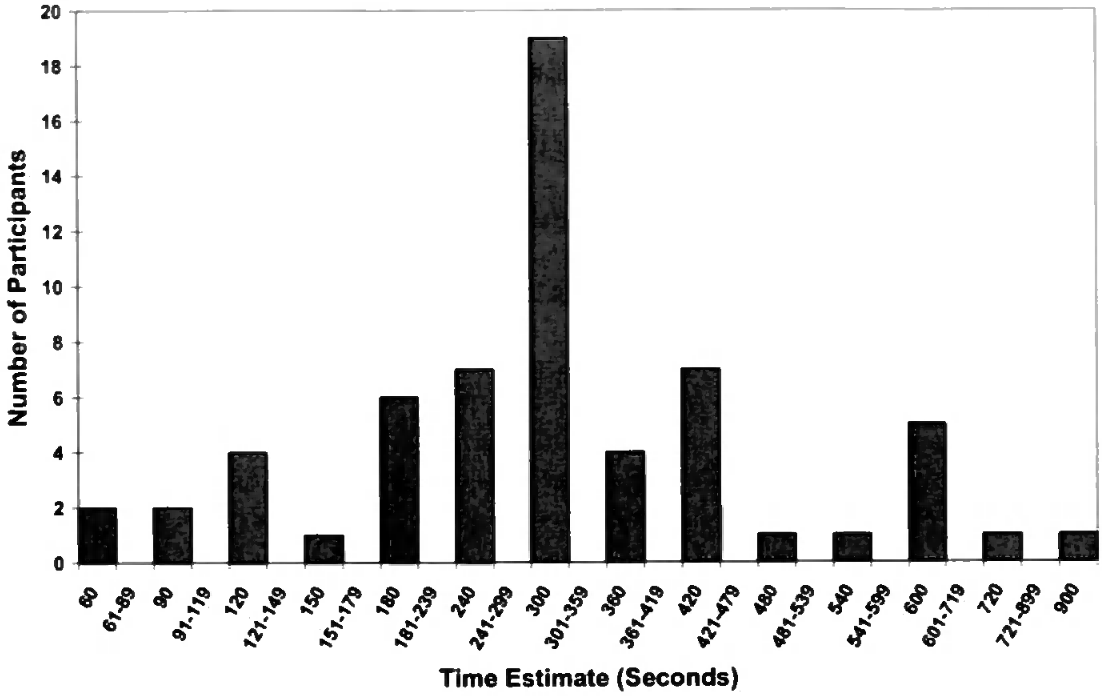

Figure 4.2

Histogram of time estimates on the short duration Playmobil® task

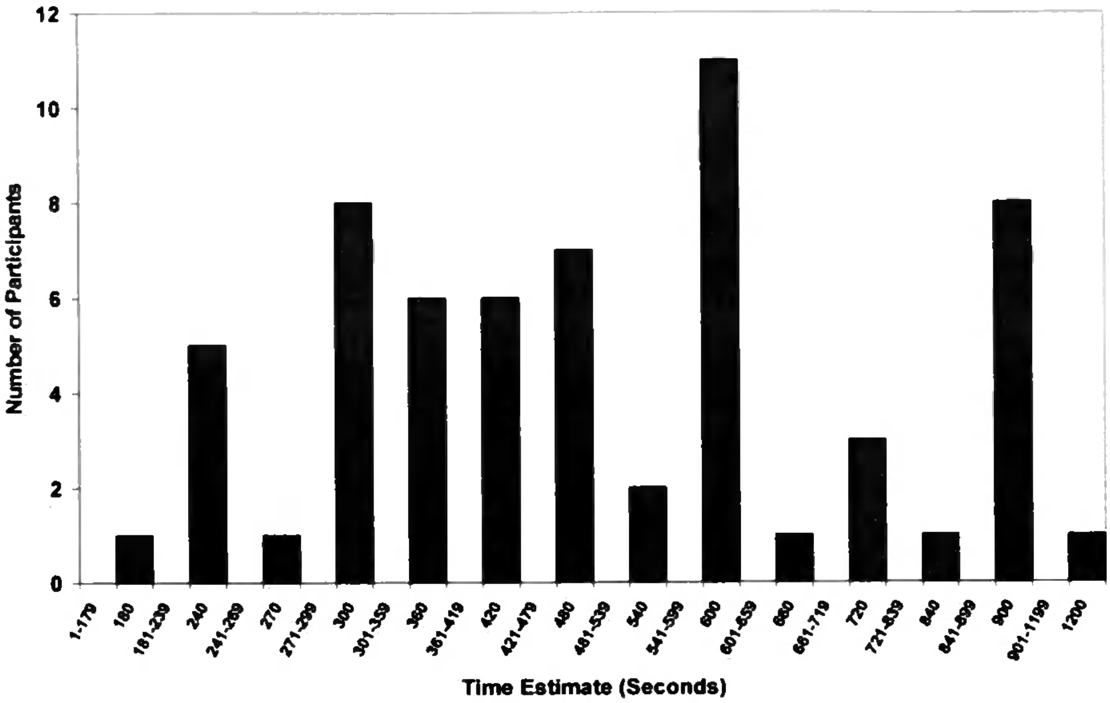

Figure 4.3

Histogram of time estimates on the long duration Playmobil® task

For the same reasons that were specified in Experiment 1, the estimated and actual completion time data were subjected to a logarithmic transformation before being statistically analysed. The log-transformed data were subjected to a 2 (task) x 2 (time) x 2 (task performance order) mixed design ANOVA. This analysis produced a main effect of task, $F(1,59) = 625.01$, $MSE = .06$, $p < .001$, with overall time being longer on the long duration task. The main effect of task performance order approached significance, $F(1,59) = 3.23$, $MSE = .38$, $p < .08$, with overall time being longer in the long then short task condition.

The interaction between task and time was significant, $F(1,59) = 86.65$, $MSE = .04$, $p < .001$ (see Figure 4.4 below). This revealed that estimates exceeded completion times on the short task (Ms = 316.74 and 239.42 seconds, respectively), whereas completion times exceeded estimates on the long task (Ms = 669.28 and 524.79 seconds, respectively). This finding suggests that, regardless of performance order, the direction in which time predictions were biased differed between the long and short duration tasks. That is, temporal overestimation occurred on the short task, whereas temporal underestimation prevailed on the long task. Planned comparisons (LSD t-tests, both two-tailed) revealed a significant difference between estimated and actual duration on each task ($ps < .05$).

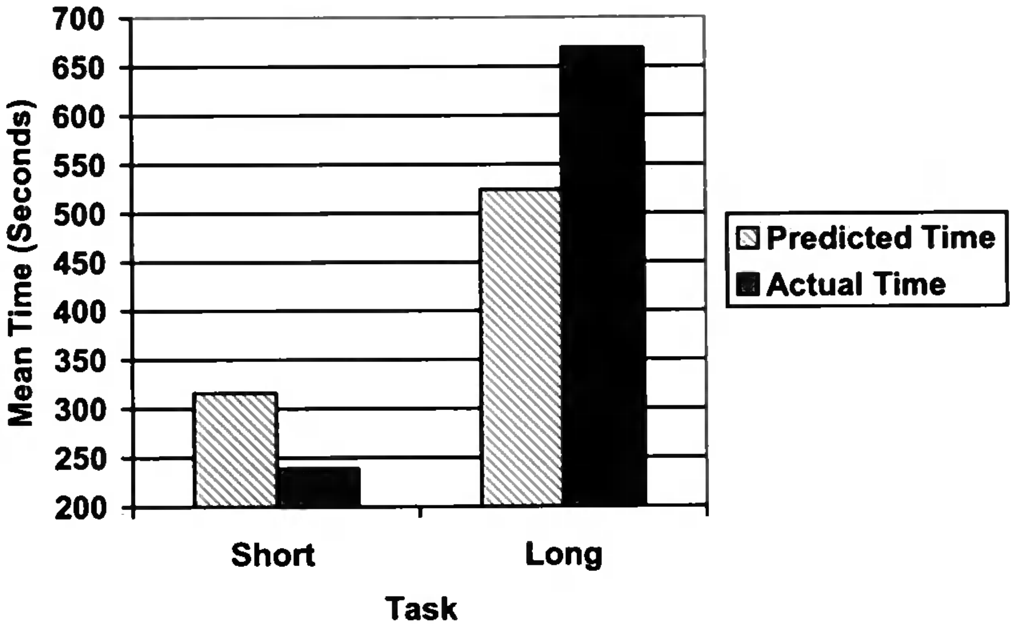

Figure 4.4

Bar graph depicting the interaction between the task and time factors

The interaction between time and task performance order approached significance, $F(1,59) = 2.95$, $MSE = .12$, $p < .10$. The main effect of time and the task by task performance order interaction were not significant ($Fs < 2.5$, $ps > .10$). However, there was a significant three-way interaction, $F(1,59) = 4.58$, $MSE = .04$, $p < .05$ (see Figure 4.1 above). This revealed that the extent of time estimation bias was greater on each task when it was performed second rather than first. Planned comparisons (LSD t-tests, all two-tailed) revealed that estimated and actual duration differed significantly on the long task regardless of performance order (ps < .05). In contrast, the difference between estimated and actual duration was significant when the short task was performed second ($p < .05$), but not first ($p > .10$). This finding suggests that the extent of temporal overestimation was greater when the short task was performed second rather than first. However, the magnitude of temporal underestimation on the long task did not differ significantly according to task performance order.

4.6 Discussion

The present study indicates that the direction in which time estimates were biased differed between the two Playmobil® tasks. As predicted, general temporal overestimation was evident on the short duration task, whereas general temporal underestimation occurred on the long duration task. The prevalence of temporal underestimation on the long task suggests that the lack of a general optimistic time prediction bias in Experiments 1 to 5 may have been due to the durations of the different versions of the Tower of Hanoi task and the repetitive disk movement task. That is, a general optimistic time prediction bias might only occur on tasks that are of longer duration than those used in the previous experiments. Moreover, as the long Playmobil® task was of similar duration to some of the laboratory tasks used in previous research (e.g., Buehler et al., 1997), such tasks may need to be of around 11 minutes' duration before a general optimistic time prediction bias is evident. Whilst it is not known whether the present findings are applicable to other tasks, this study emphasises the fact that task duration is a factor that mediates time estimation bias.

The existence of general temporal overestimation on the short duration Playmobil® task is consistent with the findings of Experiments 4 and 5, where there was a hint of a general pessimistic time prediction bias on tasks that were of similar duration (i.e., the five-disk Tower of Hanoi task and the five-disk repetitive task). This finding suggests that people do not underestimate the duration of well structured laboratory tasks that take between four and five minutes to complete. Moreover, the present study indicates that the absence of general temporal underestimation is not confined to artificial problem solving tasks, but also applies to a different type of well structured laboratory task, which is more akin to various real world activities.

A possible explanation for the general temporal overestimation that prevailed on the short Playmobil® task is that people tend to use temporal units such as 10 or 15 minutes when judging task duration in everyday life (Fraisse, 1984). Thus, on tasks that take only a few minutes to complete, temporal overestimation may be a consequence of the tendency to use longer units of time when judging task duration. Consistent with the notion that people use longer temporal units when estimating task duration, there was evidence that several participants predicting their completion time on the short and long duration Playmobil® tasks in units such as 10 or 15 minutes (see Figures 4.2 and 4.3, respectively). In fact, the frequency distributions of time estimates revealed that the vast majority of participants judged the duration of the short Playmobil® task in units of whole minutes, whereas this was the case for all participants on the long task (see Figures 4.2 and 4.3, respectively). This finding suggests that people prefer to use whole minutes rather than seconds when judging task duration.

There was some evidence that temporal misestimation was influenced by the relative duration of the previous task, as the magnitude of time prediction bias was greater when the short task was performed second rather than first. That is, time predictions were more pessimistically biased on this task among participants who had just completed the long duration task. Conversely, there was a hint of temporal underestimation being greater when the long task was performed second rather than first. However, the extent to which time predictions were optimistically biased did not differ significantly between the first and second trials on the long task. Taken together, these findings are broadly consistent with those of Experiments 3 to 5, and indicate that information about the first task may have formed the basis of time estimates on the second task.

It could be that participants used the anchoring and adjustment heuristics (Tversky & Kahneman, 1982a) when estimating the duration of the second task. The number of large plastic components involved in completing the previous task may have served as an anchor for predictions, which were insufficiently adjusted according to the relative demands of the current task (e.g., the greater or lesser number of large plastic components). An anchoring and adjustment strategy should result in time predictions on the short task being more pessimistically biased if they were anchored on such information about the just-completed long duration task. By the same token, temporal underestimation should be greater on the long duration task as a consequence of using such a judgement strategy.

Whilst similar findings were observed when different versions of the same task were performed in succession (Experiments 3 to 5), two different Playmobil® tasks were employed here. Hence, in contrast to those previous experiments, no specific information about how to complete the second task was acquired whilst performing the first task in this study. Given that both of the present tasks involved fitting plastic components together in a pre-specified manner, participants may have acquired some kind of general information about the nature of the second trial during the first trial (e.g., the actions needed for task completion). However, participants did not perform any part of the second task before estimating its duration.

In contrast, as different versions of the Tower of Hanoi task share the same structure (Karat, 1982), participants in Experiments 3 and 4 would have acquired information about the nature of the second task whilst performing the first task. For example, after the first trial, participants might have realised that completing the second task would also involve removing some disks from their final destination peg (Kotovsky et al., 1985). Likewise, participants in Experiment 5 would have obtained relevant information about how to complete the five-disk repetitive task whilst performing the three-disk task beforehand. The use of different versions of the same task meant that participants in Experiments 3 to 5 possessed some prior experience of the second task when estimating its duration.

In order to ensure that participants acquired prior experience of performing part or all of the second task, tasks that shared the same structure were employed Experiment 7. That is, two versions of the same Playmobil® task were performed consecutively. Although these tasks differed in duration, the shorter task was a sub-component version of the longer task. Hence, some kind of pertinent information such as the steps and components involved in completing the second task would be acquired whilst performing the first task.

4.7 Experiment 7

This study sought to ascertain whether prior experience of performing a longer or shorter duration version of the same Playmobil® task would influence time prediction bias on the next task. In order to achieve this goal, task duration was manipulated whilst task structure was held constant. The long duration Playmobil® task from Experiment 6 was used again in this study. This task was divided into medium and short duration sub-component versions, which were employed alongside the long task itself. Given the findings of Experiment 6, the present study also sought to determine whether temporal underestimation was still evident on the long Playmobil® task when it preceded, or was performed after, one of the two shorter duration sub-component tasks.

Since the short and medium tasks were sub-components of the long Playmobil® task, differences in duration and complexity should be more akin to those associated with different versions of the Tower of Hanoi task. For example, the three-disk Tower of Hanoi task is not only a sub-component version of the five-disk task (Kotovsky et al., 1985), but is also less complex and takes less time to complete. Indeed, the two sub-component versions of the long Playmobil® task were selected because they approximated the duration of the three-disk and five-disk tasks from Experiment 4. Pilot testing revealed that the medium version was of similar duration to the five-disk task (Ms = 242.56 and 247.99 seconds, respectively), whereas the short version approximated the duration of the three-disk task (Ms = 27.19 and 31.90 seconds, respectively).

As the short and medium sub-component tasks were of similar duration to the three-disk and five-disk versions of the Tower of Hanoi task, it was hypothesised that general temporal overestimation would occur on these tasks. However, as was the case in Experiment 6, it was hypothesised that general temporal underestimation would be evident on the long duration Playmobil® task. Given that prior experience of performing the same task has been found to reduce temporal misestimation (Josephs & Hahn, 1995), it was anticipated that time predictions on the long task would be more accurate when the medium rather than the short duration sub-component task was performed initially. That is, completing the medium duration version beforehand would provide participants with greater information about the long task (e.g., how the plastic components slotted together), which they would use to good effect.

4.8 Method

4.8.1 Participants

Eighty (64 female and 16 male) students at the University of Plymouth participated voluntarily. Forty-seven participants took part in partial fulfilment of a psychology course requirement whilst the remainder were paid £3 each. No biographical information other than gender was recorded.

4.8.2 Materials

The long duration Playmobil® task from Experiment 6 was used. The short duration sub-component version of this task entailed constructing one wall of the Playmobil® castle by following the instructions on page one of the booklet. The medium duration sub-component version involved building the walls and floor of the castle, and part of the surrounding jetty. This task entailed following the instructions on pages one to four of the booklet. The long duration task involved building a multi-turreted castle with a surrounding jetty, platform and battlements. This task involved following the instructions on pages one to eight of the booklet. A word association checklist comprising 40 words (Berkowitz & Troccoli, 1990) was also used. Task duration was measured using a digital stopwatch.

4.8.3 Design and Procedure

A 2 (time: estimated vs. actual) x 2 (task: first vs. second) x 4 (task experience: short then long duration vs. long then short duration vs. medium then long duration vs. long then medium duration) mixed factorial design was used. Task experience was a between-groups factor, with participants being randomly assigned to one of four equal-sized conditions. The time and task factors were repeated-measures, with task performance order being fully counterbalanced.

The procedure was similar to that of Experiment 6. Once participants were informed of the nature of each task, there followed a planning period that was different for each task. Pilot testing revealed that, on average, 20 seconds were needed to read the instruction booklet for the short task, 40 seconds for the medium task, and 80 seconds for the long task. Dependent on task performance order, participants were given these amounts of time to read the instruction booklet before estimating task duration. In order to ensure that all relevant pages of the instruction booklet were previewed, all participants were informed when 10 seconds of each planning period remained. At the end of each planning period, all participants gave a time estimate. This procedure was applicable to the first and second trials.

In order to give the researcher time to dismantle the apparatus of each task, participants performed a word association checklist (Berkowitz & Troccoli, 1990) at the end of the first trial. Once the apparatus of each task had been dismantled or after two minutes had elapsed (whichever was longer), participants were asked to stop performing the word association task. The word association checklist was then removed from the table and participants were presented with the plastic components and the instruction booklet for the second task. Each testing session lasted approximately 30 minutes.

4.9 Results

Basic descriptive statistics of estimated and actual duration per task and task experience condition are presented in Table 4.2 below. All test statistics used in Experiment 7 are contained in Appendix 7.

Task		Task Experience Condition			
		Short Then Long Task	Long Then Short Task	Medium Then Long Task	Long Then Medium Task
First	Time Estimate	M = 38.75 Mdn = 30.00 SD = 25.85 N = 20	M = 696.00 Mdn = 720.00 SD = 210.92 N = 20	M = 354.00 Mdn = 300.00 SD = 142.59 N = 20	M = 657.00 Mdn = 600.00 SD = 244.28 N = 20
	Completion Time	M = 22.00 Mdn = 20.00 SD = 8.58 N = 20	M = 629.00 Mdn = 618.00 SD = 113.89 N = 20	M = 290.40 Mdn = 267.50 SD = 88.88 N = 20	M = 598.30 Mdn = 610.00 SD = 123.40 N = 20
Second	Time Estimate	M = 435.00 Mdn = 405.00 SD = 181.70 N = 20	M = 28.15 Mdn = 22.50 SD = 20.12 N = 20	M = 550.50 Mdn = 540.00 SD = 129.59 N = 20	M = 254.25 Mdn = 240.00 SD = 109.20 N = 20
	Completion Time	M = 556.25 Mdn = 545.00 SD = 147.18 N = 20	M = 18.55 Mdn = 18.00 SD = 3.76 N = 20	M = 497.85 Mdn = 498.00 SD = 124.14 N = 20	M = 178.95 Mdn = 176.00 SD = 35.68 N = 20

Table 4.2

Time estimates and completion times per task and task experience condition (in seconds)

As Table 4.2 shows, temporal estimates exceeded completion times in all task experience conditions on the first task, and in three out of the four conditions on the second task. This finding suggests that general temporal overestimation was evident in this study. That is, a pessimistic time prediction bias occurred on the short and medium duration tasks when they were performed first or second, whereas an optimistic time prediction bias was only evident on the long task when it was preceded by the short duration sub-component version. Thus, in contrast to the findings of Experiment 6, general temporal underestimation was not evident on the long duration Playmobil® task. Bar graphs of predicted and actual completion times on the short, medium and long duration tasks are presented below in Figures 4.5, 4.6 and 4.7, respectively.

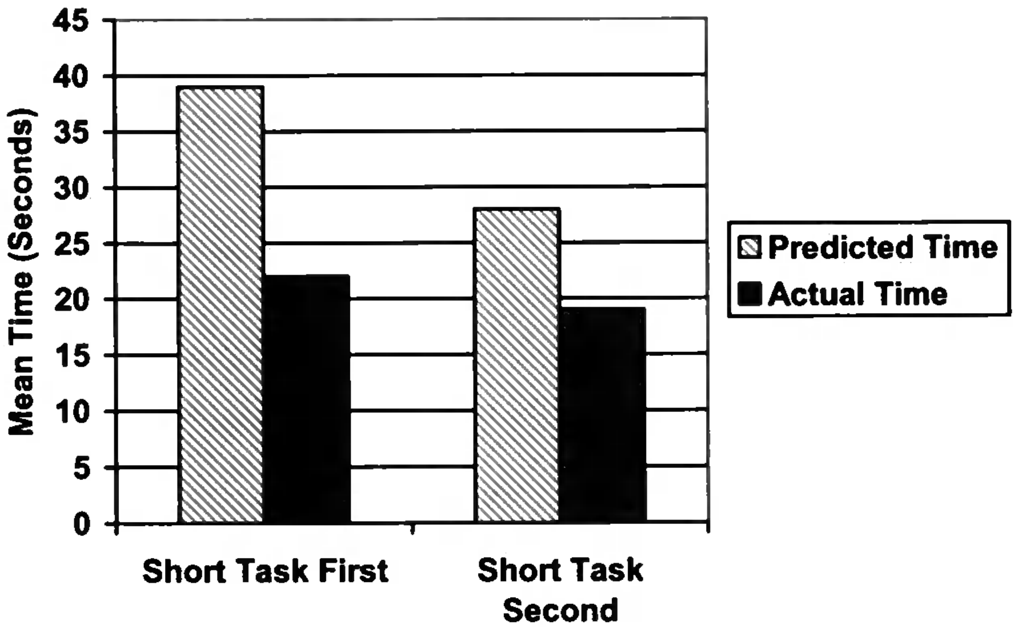

Figure 4.5

Bar graph of predicted and actual completion times on the short duration task per task performance order

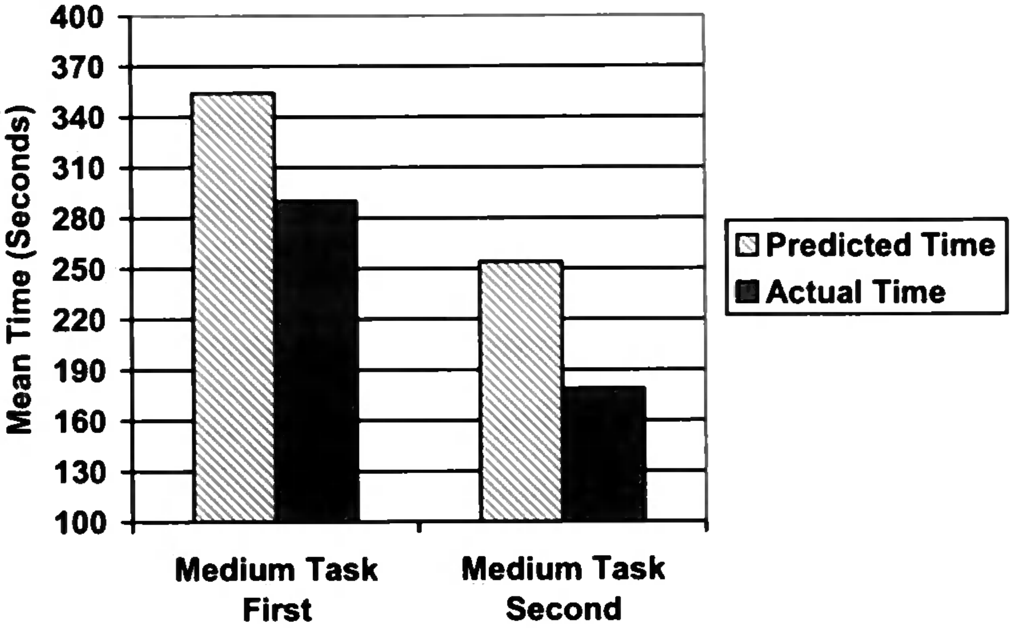

Figure 4.6

Bar graph of predicted and actual completion times on the medium duration task per task performance order

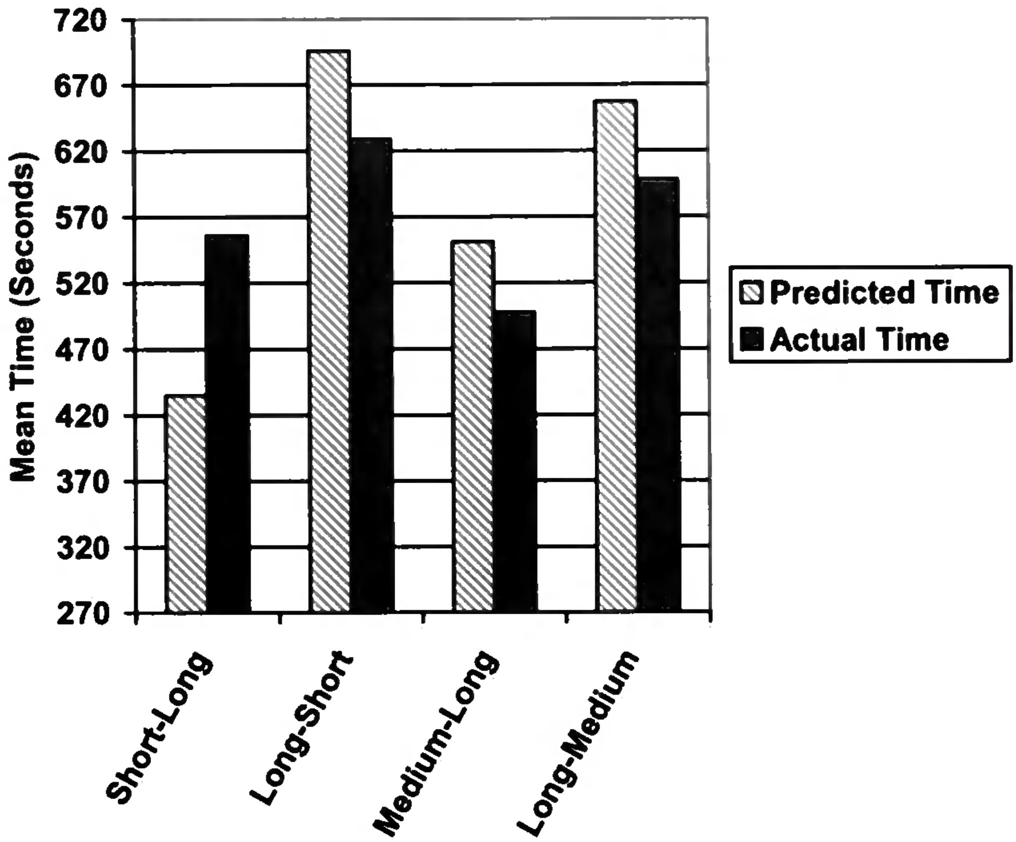

Figure 4.7

Bar graph of predicted and actual completion times on the long duration task per task experience condition

In order to determine how many participants predicted task duration using whole minutes, frequency distributions of time estimates were calculated. Histograms of time estimates on the short, medium and long duration Playmobil® tasks (regardless of performance order and task experience condition) are presented below in Figures 4.8, 4.9 and 4.10, respectively.

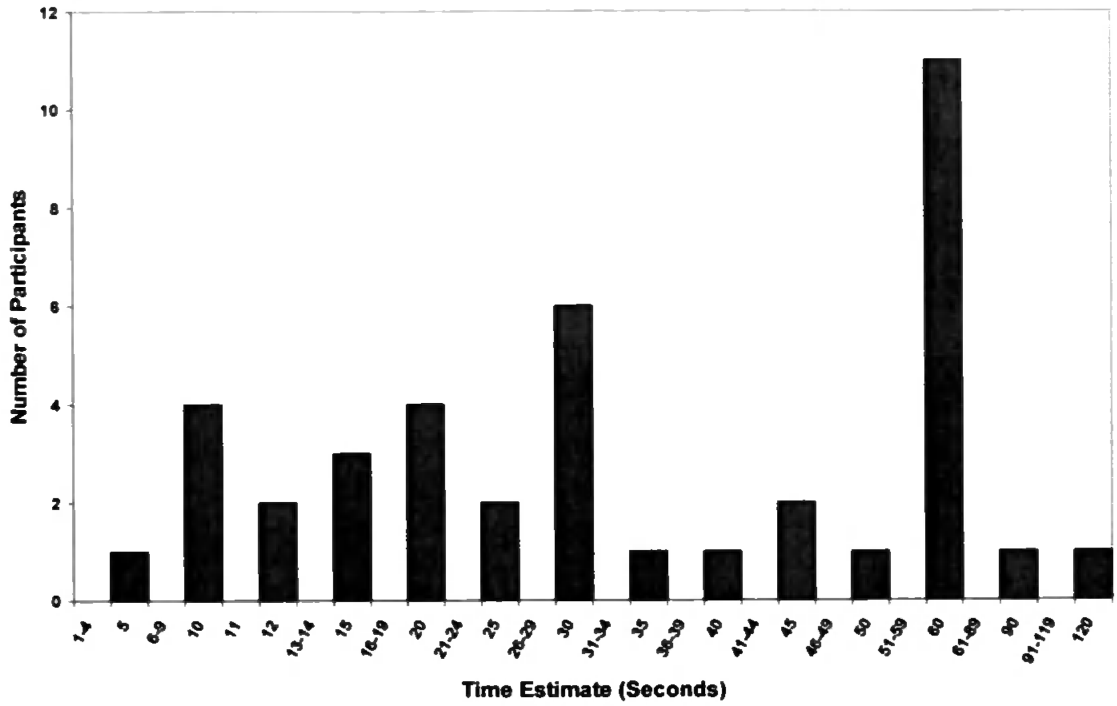

Figure 4.8

Histogram of time estimates on the short duration Playmobil® task

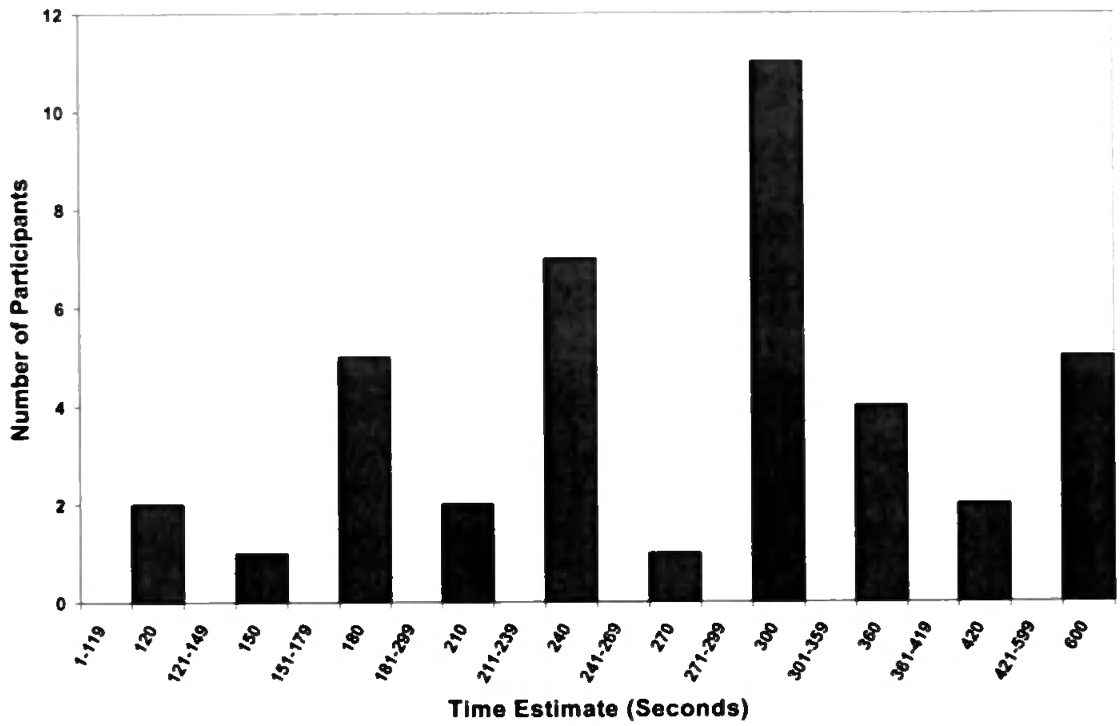

Figure 4.9

Histogram of time estimates on the medium duration Playmobil® task

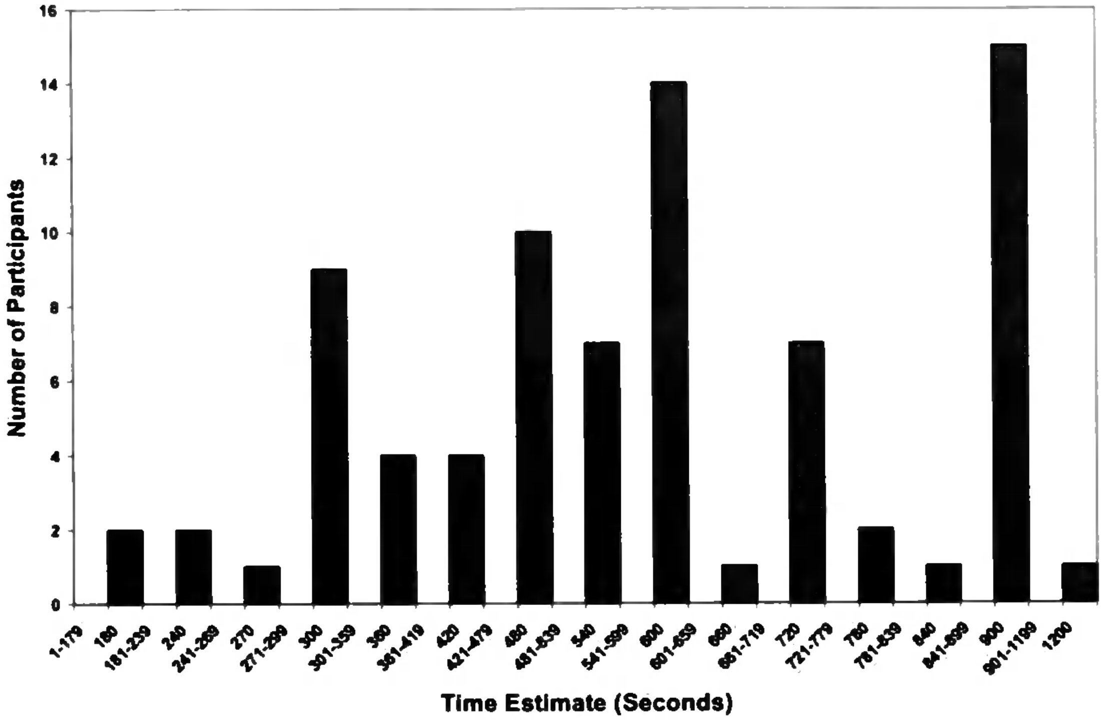

Figure 4.10

Histogram of time estimates on the long duration Playmobil® task

For the same reasons that were specified in Experiment 1, the estimated and actual completion time data were subjected to a logarithmic transformation before being statistically analysed. In order to address the issue of prior task experience, the log-transformed data from the second trial were analysed. A 2 (time) x 4 (task experience) mixed design ANOVA produced a main effect of task experience, $F(3,76) = 637.66$, $MSE = .14$, $p < .001$, with overall time being longest in the medium then long task condition. Scheffé pairwise comparisons revealed significant mean differences between all conditions ($ps < .05$) except the medium then long and the short then long task conditions ($p > .10$). The main effect of time was not significant ($F < 3$, $p > .10$).

There was a significant interaction, $F(3,76) = 6.12$, $MSE = .11$, $p < .01$ (see Figure 4.11 below). This revealed evidence of temporal overestimation on the medium and short tasks, whereas the direction in which time predictions were biased on the long task differed according to the relative duration of the previous task. That is, temporal underestimation was evident when the short task was performed beforehand, whereas overestimation occurred when the medium task was completed initially. Planned comparisons (LSD t-tests, all two-tailed) revealed that estimated and actual duration differed significantly on the medium task ($p < .05$), but not on the short task ($p > .10$). On the long duration task, the difference between estimated and actual completion time was significant when it was preceded by the short task ($p < .05$), but not by the medium task ($p > .10$). This finding suggests that performing the medium rather than the short duration sub-component task beforehand resulted in greater time prediction accuracy on the long duration task.

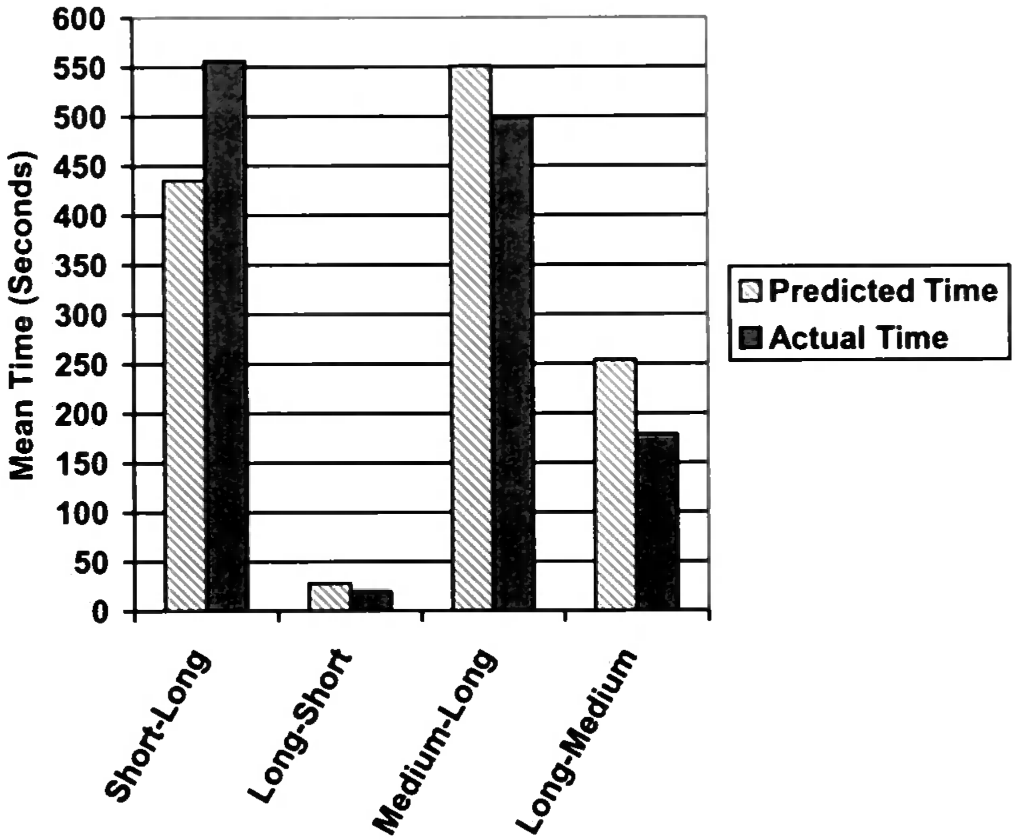

Figure 4.11

Bar graph depicting the interaction between the time and task experience factors (data from the second task)

The issue of prior task experience was further addressed by analysing the log-transformed time prediction data from the short and medium duration tasks. If predictions were influenced by the relative duration of the previous task, then they should be longer when the tasks were performed second rather than first. That is, when the short and medium sub-component tasks were preceded by the long duration task. Whilst time predictions were longer on each task when they were performed second (see Table 4.2), independent-groups t-tests (both two-tailed) revealed that the difference between the two trials was significant on the medium task, $t(38) = 2.61$, $p < .05$, but not on the short task,

$t(38) = 1.67$, $p > .10$. Thus, the longer duration of the just-completed task may have influenced time predictions on the medium sub-component task only.

In order to determine whether time prediction accuracy differed between the short, medium and long duration tasks, the log-transformed data from the first trial were analysed. The data from the long duration task were not collapsed across the task experience factor, as it was decided to keep sample sizes equal, that is, not to have 40 out of 80 participants in one of three treatment conditions. A 2 (time: estimated vs. actual) x 4 (task condition: short vs. medium vs. long then short vs. long then medium) mixed design ANOVA produced a main effect of time, $F(1,76) = 13.79$, $MSE = .09$, $p < .001$, with temporal estimates exceeding completion times (Ms = 436.44 and 385.06 seconds, respectively). This finding suggests that general temporal overestimation was evident on the first trial.

The ANOVA also produced a main effect of task condition, $F(3,76) = 464.12$, $MSE = .20$, $p < .001$, with overall time being longest in the long then short task condition. Scheffé pairwise comparisons revealed significant differences between the means of all conditions ($ps < .05$) except those of the long then medium and the long then short task conditions ($p > .10$). The interaction was not significant ($F < 2$, $p > .10$). The lack of a significant interaction suggests that the extent to which time predictions were pessimistically biased did not differ between the short, medium and long duration tasks when they were performed first. Moreover, the magnitude of temporal overestimation did not differ between the long then medium task and the long then short task conditions on the first trial.

4.10 Discussion

Contrary to the findings of Experiment 6, there was no evidence of general temporal underestimation on the long duration Playmobil® task in this study. In fact, general temporal overestimation was observed when this task was performed first. Likewise, there was a hint of temporal overestimation when it was preceded by the medium duration sub-component version. The discrepancy between the findings of Experiments 6 and 7 is not easily explicable, as the procedure associated with this task was identical in both studies. That is, all of the plastic components were displayed on the table in front of participants, who were given the same amount of time to preview the task and its instructions, and received the same warning when the pre-exposure period was coming to an end. Given these similarities, participants in both studies would presumably have possessed the same amount of information the nature of the long task when estimating its duration. Thus, it is not clear why general temporal underestimation was evident on the long task in Experiment 6, whereas general temporal overestimation occurred in the present study.

The present study also produced some evidence of temporal overestimation on the short and medium duration sub-component versions of the long Playmobil® task regardless of the order in which they were performed. This finding suggests that general temporal overestimation is not confined to artificial problem solving tasks such as the Tower of Hanoi, but might also generalise to more realistic well structured laboratory tasks with a duration of up to four or five minutes. However, there was some evidence of temporal underestimation, but only on the long duration Playmobil® task when it was preceded by the short sub-component task. Thus, participants tended to underestimate their completion times only when they had completed one part (i.e., one page in the instruction booklet) of the long duration task beforehand.

The presence of temporal underestimation on the long task when it was performed after the short task suggests that time predictions may have been influenced by the duration of the short task relative to that of the long task. For example, the duration of the short task may have served as a basis for time predictions on the long task. However, there was no evidence of such a judgement strategy being used on the long task when it was preceded by the shorter duration medium sub-component task. Specifically, there was a hint of temporal overestimation on the long task when the medium sub-component version had just been performed. Taken together, these findings suggest that the direction in which time predictions on the long task were biased (i.e., over or underestimation) differed according to the relative duration of the previous task

A possible explanation for these findings is that, due to differences in the amount of prior task experience, participants used different kinds of task-related information when estimating the duration of the long task. As participants who performed the short task initially constructed only one wall of the Playmobil® castle, they possessed little information about how to complete the long task when estimating its duration. Given this lack of prior task experience, these participants may have used a heuristic information processing strategy when making a second time estimate. For example, time predictions might have been anchored on the perceived duration of the previous task, but insufficiently adjusted according to the longer duration of the upcoming task. The use of such an anchoring and adjustment judgement strategy should lead to temporal underestimation when the previous task was of shorter duration than the current task. Thus, participants in the short then long task condition may have used these cognitive heuristics when making a second time estimate.

In contrast, assembling half of the long task beforehand (i.e., the medium duration sub-component task) meant that participants in this condition possessed considerable information about the nature of the long task when estimating its duration. Given the extent of their prior task experience, these individuals might have engaged in a more thorough information processing strategy when making a second time estimate. For example, they may have calculated the number of large plastic components required to complete the first task, and appropriately scaled up this figure as a function of the greater number of major components involved in finishing the second task. Using such a judgement strategy could result in temporal overestimation if it involved thinking about factors that delayed the completion of the previous task (e.g., fitting some plastic components together incorrectly). Thus, these participants may have erred on the side of caution because they took account of their previous task performance. In fact, time predictions were more accurate when the long task was preceded by the medium rather than the short task, suggesting that these individuals used such distributional information to good effect.

There was evidence that time predictions on the medium task may have been influenced by the relative duration of the just-completed task. That is, relative to the first trial, time predictions were longer when this task was performed second (i.e., after the long duration task). A possible explanation for this finding is that participants calculated the number of major plastic components involved in completing the previous task (e.g., the components needed to build the walls of the castle), but overestimated the number of large plastic components needed to finish the current task. Time predictions should be longer as a consequence of using this kind of judgement strategy when the current task comprises fewer such components than the previous one. In contrast, there was no evidence of such a judgement strategy being used on the short duration sub-component task, as time predictions were slightly longer when it was performed first rather than second. However, as the length of time predictions on the short duration task did not differ significantly according to task performance order, it is unwise to make too much of this finding.

There was evidence that participants used whole minutes rather than seconds when estimating their completion times on the short, medium and long duration Playmobil® tasks. For example, the frequency distributions of time estimates revealed that nearly all of the participants used whole minutes when judging the duration of the medium and long Playmobil® tasks (see Figures 4.9 and 4.10, respectively). Moreover, almost half of these individuals used temporal units such as 10 and 15 minutes when estimating the duration of these two tasks. Whilst no participants used these temporal units when judging the duration of the short sub-component task, a sizeable minority estimated their completion times using whole minutes rather than seconds (see Figure 4.8). These findings provide support for the notion that people tend not to judge task duration using seconds, but prefer instead to use whole minutes or longer temporal units when estimating their task completion times.

4.11 General Discussion

The present studies have produced evidence of general temporal overestimation on three Playmobil® tasks that were of similar duration to the tasks used in Experiments 1 to 5. Specifically, there was some evidence that time predictions were pessimistically biased on the short Playmobil® task (Experiment 6) and the two shorter duration sub-component versions of the long task (Experiment 7). This finding indicates that people tend to err on the side of caution when estimating the duration of well structured laboratory tasks that are less artificial than the Tower of Hanoi and the repetitive disk movement task. Moreover, this finding suggests that the temporal underestimation indicative of the planning fallacy does not occur on two very different types of laboratory task that take up to five minutes to complete.

In Experiment 6, a general optimistic time prediction bias was evident on the Playmobil® task that took about 11 minutes to complete, providing support for the notion that temporal underestimation prevails on longer tasks than those used in Experiments 1 to 5. An optimistic time prediction bias was also evident on this task in Experiment 7, but only when it was preceded by the short duration sub-component version, which took about 30 seconds to complete. Conversely, a general pessimistic time prediction bias was evident on the long duration task when it was performed first, and there was a hint of temporal overestimation when this task was preceded by the medium duration sub-component task (Experiment 7). Since there was no evidence of a general optimistic time prediction bias on the long duration task in Experiment 7, the present studies provide conflicting evidence concerning the direction in which temporal estimates were biased on this task.

Given that the long duration Playmobil® task had not been employed in research before Experiment 6, it was not known whether participants would under or overestimate its duration. In fact, it was anticipated that an optimistic time prediction bias would occur on this task because it was of similar duration to some of the tasks used in previous research where temporal underestimation prevailed (e.g., Francis-Smythe & Robertson, 1999). However, as temporal underestimation has also been observed on laboratory tasks with a duration of over one hour (Byram, 1997), it could be that the long Playmobil® task was too short for a general optimistic time prediction bias to occur.

Consistent with this suggestion, people's propensity for estimating time using longer units such as 15 or 20 minutes (Fraisse, 1984) could result in temporal predictions being pessimistically biased on tasks that are of shorter duration than such intervals. Thus, the general temporal overestimation that occurred in Experiment 7 might reflect the direction in which time predictions are biased on the long duration Playmobil® task. Given the rather conflicting findings of the present studies, the long duration Playmobil® task was employed in Experiments 8 and 9 in order to determine the direction of time prediction bias on it.

The benefits of prior task experience were emphasised in Experiment 7, where time predictions were more accurate among participants who had performed half of the long duration Playmobil® task beforehand. That is, the extent of temporal misestimation on the long task was significantly less when the medium rather than the short duration sub-component task was performed beforehand. A plausible explanation for this finding is that performing the medium task provided participants with considerable task-related information, which they used to good effect. Moreover, it could be that these participants considered factors that delayed the completion of the first task (e.g., incorrectly assembling some plastic components), and incorporated such information into their second time estimate. Given the findings of Buehler et al. (1994), taking account of information about personal performance on a previous similar task would be expected to result in more accurate time estimates.

The impact of prior experience of performing a longer or shorter duration task on time prediction bias was also demonstrated in the present studies. That is, time prediction bias on the second task was found to differ according to the relative duration of the first task. For example, on the second trial, there was a hint of greater temporal underestimation on the long duration Playmobil® task, whereas the magnitude of temporal overestimation on the short task was greater relative to the first trial (Experiment 6). Similarly, temporal underestimation occurred when the long task was preceded by the short task, whereas there was evidence of temporal overestimation on the medium and short tasks when they were performed after the long task (Experiment 7). These findings are broadly consistent with those of Experiments 3 to 5, and indicate that participants may have used information about the nature of the first task when estimating the duration of the second task.

Whilst information concerning the previous task might have formed the basis of time estimates on the current task, the nature of this task-related information is not known. Given that the tasks used in Experiments 3 to 7 differed in duration, it could be that the perceived duration of the previous task served as a basis for time predictions on the current task. That is, participants estimated how long they had taken to complete the first task, and used this information as a basis for predicting the duration of the second task. This kind of judgement strategy would be expected to result in temporal underestimation when the current task was of longer duration than the previous one. Conversely, temporal overestimation should be greater when the longer duration of a just-completed task formed the basis of time predictions on a subsequent shorter task.

The anchoring and adjustment heuristics (Tversky & Kahneman, 1982a) might also be used in this kind of judgement strategy, with temporal misestimation occurring when successive tasks differ in duration. That is, time predictions are anchored on the perceived duration of the previous task, but are insufficiently scaled up or down according to the longer or shorter duration of the upcoming task. For example, in Experiment 7, temporal underestimation may have been a consequence of time predictions being anchored on the perceived duration of the just-completed short sub-component task, with insufficient adjustment for the longer duration of the long Playmobil® task.

Although the tasks used in Experiments 3 to 7 differed in duration, they were also similar. Thus, participants would have acquired some kind of information about the general nature of the second task whilst performing the first task. Moreover, as different versions of the same task were employed in all but one of these studies (i.e., Experiment 6), most participants had some experience of performing all or part of the second task when estimating its duration. For example, in Experiment 7, performing the short sub-component task initially meant that participants had already built part of the long duration Playmobil®

task before making a second time prediction. Likewise, as the Playmobil® tasks used in Experiment 6 both involved fitting together plastic components in a pre-specified manner, participants would have acquired some kind of general information about the nature of the second task whilst performing the first one (e.g., the actions needed for task completion).

Given the similarity between the first and second trials in Experiments 3 to 7, information other than the perceived duration of the previous task may have formed the basis of time predictions on the second task. That is, time predictions might have been based on information about the nature of the previous task such as the number of plastic Playmobil® components or the number of Tower of Hanoi disks. A possible judgement strategy for using such information might entail a calculation of the number of large plastic components involved in completing the previous Playmobil® task, and an estimation of the number of such components needed to finish the to-be-completed task.

Whilst this kind of judgement strategy would presumably require greater cognitive resources than one involving the perceived duration of the previous task, it could also lead to temporal misestimation. For example, the optimistic time prediction bias in Experiment 7 may have been due to participants underestimating the number of large plastic components involved finishing the upcoming long duration Playmobil® task, and basing their second temporal estimate on this information. Similarly, the anchoring and adjustment heuristics might be used in this kind of judgement strategy, with temporal misestimation occurring because of a failure to take sufficient account of the greater or lesser amount of work involved in completing the current task (e.g., the number of large Playmobil® task components).

The use of different judgement strategies and different kinds of task-related information could thus explain the effects of prior task performance that were observed in

Experiments 3 to 7. In order to determine the type of task-related information that formed the basis of subsequent time estimates, the relevance of prior experience was manipulated in the studies presented in the next chapter of this thesis. To this end, Experiments 8 and 9 employed tasks that were either related or unrelated to each other, which participants performed consecutively.

4.12 Conclusions

Experiments 6 and 7 have produced a number of potentially important findings. There was some evidence that participants made pessimistically biased time predictions on well structured laboratory tasks that are similar to some of the tasks used in previous research (e.g., Byram, 1997), but are less artificial than the ones employed in Experiments 1 to 5. Given the duration of most of the tasks used in Experiments 6 and 7, it is clear that general temporal underestimation does not occur on a different type of well structured laboratory task that takes up to four or five minutes to complete. This finding suggests that the absence of a general optimistic time prediction bias in Experiments 1 to 5 may have been due to the shorter duration of the Tower of Hanoi task and the repetitive disk movement task rather than the nature of these tasks (e.g., their artificiality).

The present studies have produced evidence of an optimistic time prediction bias on a well structured laboratory task that took about 11 minutes to complete. That is, temporal underestimation occurred on the long duration Playmobil® task regardless of performance order in Experiment 6, and was also evident on this task when it was preceded by the short duration sub-component task in Experiment 7. The latter finding is broadly consistent with previous studies (e.g., Experiment 5), and suggests that temporal underestimation could be due to time predictions on longer tasks being based on information about previous tasks, which are of shorter duration. Indeed, in Experiments 6 and 7, there was some evidence that time prediction bias on the current task differed according to the relative duration of the previous task. This finding suggests people do take account of information about a just-completed task when subsequently estimating the duration of a similar task (Experiment 6) or a different version of the same task (Experiment 7).

The next chapter of this thesis comprises the final two experiments of this research programme. These studies sought to determine the type of information about a previous task that formed the basis of time predictions on a current task. In order to achieve this goal, related or unrelated tasks were performed in succession in Experiments 8 and 9. The issue of the relevance of prior experience was addressed in Experiment 8, where a similar or a different shorter duration task was performed before the long Playmobil® task. The relevance of prior task experience was also the focus of Experiment 9, where the long Playmobil® task was either performed for a second time or was preceded by an unrelated task, which was of similar, shorter or longer duration.

Chapter Five

The Impact of Prior Task Performance on Time Estimation Bias

5.1 Overview

This chapter comprises the final two experiments of the present research programme. Given the findings of Experiments 3 to 7, the issue of prior experience of performing a similar or different task was further addressed in Experiments 8 and 9. Specifically, these studies sought to determine the type of information about a just-completed task that formed the basis of time estimates on the long duration Playmobil® task from Experiments 6 and 7. In order to achieve this goal, tasks that were either related or unrelated to the long duration Playmobil® task were performed beforehand in both of the present studies. Completing a related task initially should provide participants with information that was relevant to the long duration Playmobil® task, whereas no such information was acquired whilst performing an unrelated task beforehand.

Experiment 8 sought to ascertain whether time prediction bias on the long Playmobil® task differed according to the type of shorter duration task that had just been completed. The relevance of task experience was manipulated by having participants perform either the short duration sub-component Playmobil® task (Experiment 7) or the three-disk version of the Tower of Hanoi task beforehand. The issue of the relevance of prior task experience was also addressed in Experiment 9. In this study, participants either completed one of three versions of the Tower of Hanoi task before the long duration Playmobil® task or performed this task twice in succession.

5.2 General Introduction to Experiments 8 and 9

Prior experience of performing a task of different duration was shown to influence time prediction bias in Experiments 3 to 7. That is, in all but one of these studies (i.e., Experiment 6), the only evidence of the temporal underestimation indicative of the planning fallacy (Kahneman & Tversky, 1979) was on longer duration tasks when they were preceded by shorter tasks. For example, in Experiment 3, an optimistic time prediction bias only occurred when the four-disk Tower of Hanoi task was preceded by the three-disk task. Likewise, temporal underestimation was only evident when the long duration Playmobil® task was performed after the short sub-component task in Experiment 7.

There was also some evidence that prior experience of performing longer duration tasks led to an increase in temporal overestimation on shorter tasks. For example, in Experiment 4, relative to when the three-disk Tower of Hanoi task was performed for a second time, the extent of overestimation on this task was greater when the five-disk task was completed beforehand. This finding indicates that participants might have used information about the previous task when making a second time prediction.

Given that similar findings were observed when different versions of the same task (e.g., Experiment 4) or similar tasks (Experiment 6) were performed consecutively, it could be that time estimates were based on general rather than specific information about the nature of the previous task. That is, instead of thinking about precise information such as how certain plastic Playmobil® components fitted together, participants may have considered general information such as the number of steps involved in completing the previous task.

Support for this notion comes from Experiment 6, where the magnitude of time estimation bias was greater when the long and short duration Playmobil® tasks were performed second rather than first. As these tasks were different, no information about how to complete the second task would have been acquired whilst performing the first one. Hence, the extent of temporal misestimation should have been similar on both trials if participants did not use some kind of information about the nature of the previous task. Since each task involved constructing a miniature castle, participants may well have based their second time prediction on some kind of general information about the nature of the previous task (e.g., the number of large plastic Playmobil® components).

The perceived duration of the previous task was suggested as another possible source of task-related information. That is, participants may have estimated how long it had taken them to complete the first task, and used this information as a basis for making a second time prediction. Such a judgement strategy should lead to temporal misestimation if individuals failed to increase or decrease their second time prediction sufficiently according to the longer or shorter duration of the upcoming task. Similarly, failing to take sufficient account of the greater or lesser number of large plastic components needed to assemble the upcoming Playmobil® task should also result in temporal misestimation if time predictions were based on the number of large plastic components involved in completing the previous task. Thus, relying on either type of task-related information when making a second time prediction should lead to temporal misestimation if individuals used an anchoring and adjustment judgement strategy.

It has been shown that externally-presented numerical values are not a prerequisite of anchoring and adjustment (Epley & Gilovich, 2002), suggesting that these cognitive heuristics are used when numerical values are self-generated. In the present research, self-generated numerical anchor values may have taken the form of estimates of the number of components (e.g., the number of Tower of Hanoi disks) or the duration of the previous task. Thus, potential sources of the anchoring component of an anchoring and adjustment temporal judgement strategy were evident in the present research.

A possible source of the adjustment component of this kind of judgement strategy could be the failure to appreciate the increase in task complexity with each additional Tower of Hanoi disk. For example, participants in Experiment 4 may have based their second time prediction on the number of disk moves involved in completing the three-disk task, but increased this estimate as a function of the extra disks associated with the upcoming five-disk task instead of the greater number of disk moves required to finish this task. Such insufficient adjustment from self-generated numerical anchor values would be expected to result in predictions being optimistically biased when the current task is more complex than the previous one (Suantak et al., 1996).

Whilst some kind of information about the first task may well have served as a basis for estimating the duration of the second task in Experiments 3 to 7, it is not clear what form this information takes. In order to determine the type of task-related information upon which subsequent time estimates are based, the issue of the relevance of prior task experience was addressed in the present studies. To this end, tasks that differed in duration but were either related or unrelated to each other were performed consecutively in Experiments 8 and 9.

Completing a related task initially would provide participants with information about the nature of the upcoming task, whereas no such relevant prior experience would be acquired whilst performing an unrelated task. Thus, if information about the nature of the previous task formed the basis of time predictions, then judgement accuracy should be greater when a related task had just been completed. In contrast, if time predictions were based on the perceived duration of the previous task, then the extent of temporal misestimation should not differ according to the relevance of prior task experience. A key aim of Experiment 8 was to ascertain whether prior performance of a related or an unrelated shorter duration task influenced time prediction bias on the long duration Playmobil® task.

5.3 Experiment 8

This study sought to determine the type of information about a just-completed shorter duration task that formed the basis of time estimates on the long Playmobil® task. That is, whether participants used information about the nature (e.g., the number of task components) or the perceived duration of the previous task when making a second time prediction. In order to achieve this goal, the relevance of prior experience was manipulated by employing two shorter duration tasks, one of which was related to the long Playmobil® task. The related task was the short duration sub-component task from Experiment 7, whereas the unrelated task was the three-disk Tower of Hanoi task. Pilot testing revealed that the three-disk task and the short Playmobil® task were of similar duration (Ms = 28.59 and 25.37 seconds, respectively).

Performing the short sub-component task initially would provide participants with information about the nature of the long duration Playmobil® task, whereas no relevant prior experience would be acquired whilst completing the three-disk task. Hence, if temporal estimates were based on information about the nature of the previous task, then time predictions on the long duration task should be more accurate when the related task was performed beforehand. Conversely, if temporal estimates were based on the perceived duration of the previous task, then the extent of time prediction bias on the long duration task should not differ according to the relevance of prior task experience.

Given the findings of Experiment 7, it was anticipated that temporal underestimation would occur on the long Playmobil® task when it was preceded by the short duration sub-component task. However, no prediction was made about the direction (i.e., under or overestimation) or extent of time prediction bias on the long duration task when the three-disk Tower of Hanoi task was performed initially. As temporal overestimation was evident previously on the three-disk (e.g., Experiment 1) and short duration Playmobil® tasks (Experiment 7), it was expected that time predictions would be pessimistically biased on these tasks here.

5.4 Method

5.4.1 Participants

Fifty-six (42 female and 14 male) students at the University of Plymouth participated voluntarily. Forty-three participants took part in partial fulfilment of a psychology course requirement whilst the remainder were paid £2.50 each. No biographical information other than gender was recorded.

5.4.2 Materials

The long duration Playmobil® task from Experiments 6 and 7, and a wooden Tower of Hanoi apparatus containing three different-sized disks were used. A digital stopwatch was used to record task duration.

5.4.3 Design and Procedure

A 2 (task: short duration vs. long Playmobil®) x 2 (time: estimated vs. actual) x 2 (prior task experience: three-disk Tower of Hanoi task first vs. short Playmobil® task first) mixed factorial design was used. Prior task experience was a between-groups factor, with participants being randomly assigned to one of two equal-sized conditions. The time and task factors were repeated-measures, with task performance order being fully counterbalanced.

The procedure was similar to that of Experiment 7. Pilot testing revealed that 20 seconds were needed to preview the three-disk Tower of Hanoi apparatus and its instruction sheet. Participants in the three-disk task condition were thus given 20 seconds to view this task before estimating its duration. Participants in the short duration Playmobil® task condition were also given 20 seconds to view the components and instruction booklet of this task before estimating its duration. As was the case in Experiments 6 and 7, participants viewed the plastic components and the instruction booklet of the long Playmobil® task for 80 seconds before estimating its duration. All participants were informed when 10 seconds of each pre-exposure period remained. Each testing session lasted approximately 20 minutes.

5.5 Results

Table 5.1 below contains basic descriptive statistics of estimated and actual duration per task and prior task experience condition. All test statistics used in Experiment 8 are contained in Appendix 8.

Task		Prior Task Experience Condition	
		Three-disk Tower of Hanoi Task	Short Duration Playmobil® Task
Long Duration Playmobil® (Second Task)	Time Estimate	M = 412.50 Mdn = 360.00 SD = 137.99 N = 28	M = 432.86 Mdn = 420.00 SD = 177.66 N = 28
	Completion Time	M = 614.04 Mdn = 599.50 SD = 125.32 N = 28	M = 599.07 Mdn = 597.00 SD = 140.91 N = 28
Short Duration (First Task)	Time Estimate	M = 124.29 Mdn = 120.00 SD = 120.00 N = 28	M = 40.14 Mdn = 30.00 SD = 32.94 N = 28
	Completion Time	M = 27.25 Mdn = 20.00 SD = 16.29 N = 28	M = 22.61 Mdn = 22.00 SD = 6.61 N = 28

Table 5.1

Time estimates and completion times per task and prior task experience condition (in seconds)

As Table 5.1 shows, temporal estimates exceeded completion times on both of the short duration tasks, suggesting that a general pessimistic time prediction bias was evident on these tasks. Conversely, a general optimistic time prediction bias was evident on the long duration Playmobil® task, with completion times exceeding estimates regardless of prior task experience. Bar graphs of predicted and actual duration on the long and short duration tasks are presented below in Figures 5.1 and 5.2, respectively.

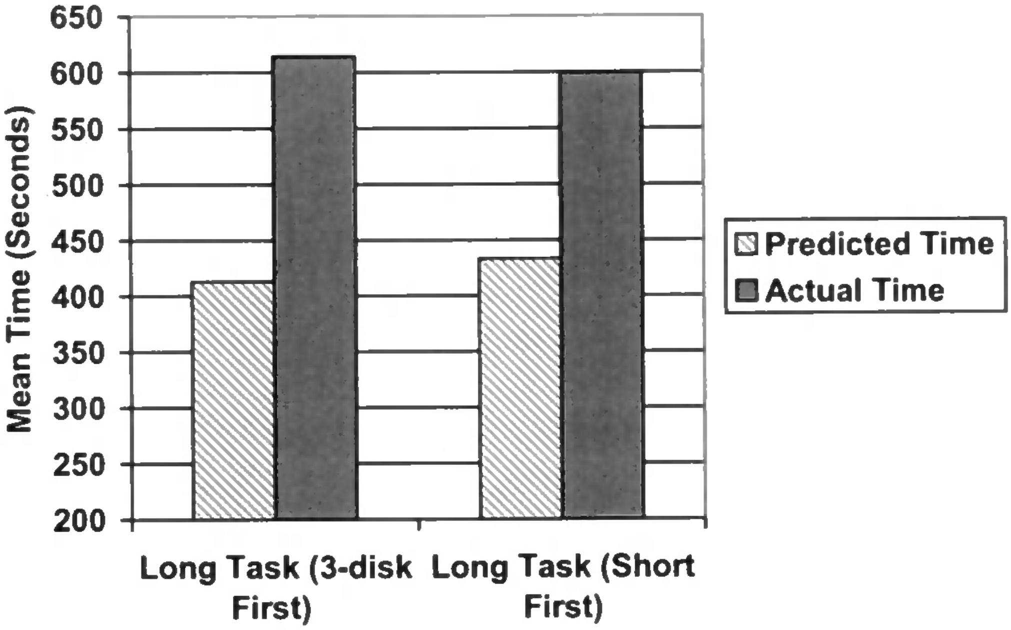

Figure 5.1

Predicted and actual completion times per prior task experience condition (data from the long duration Playmobil® task)

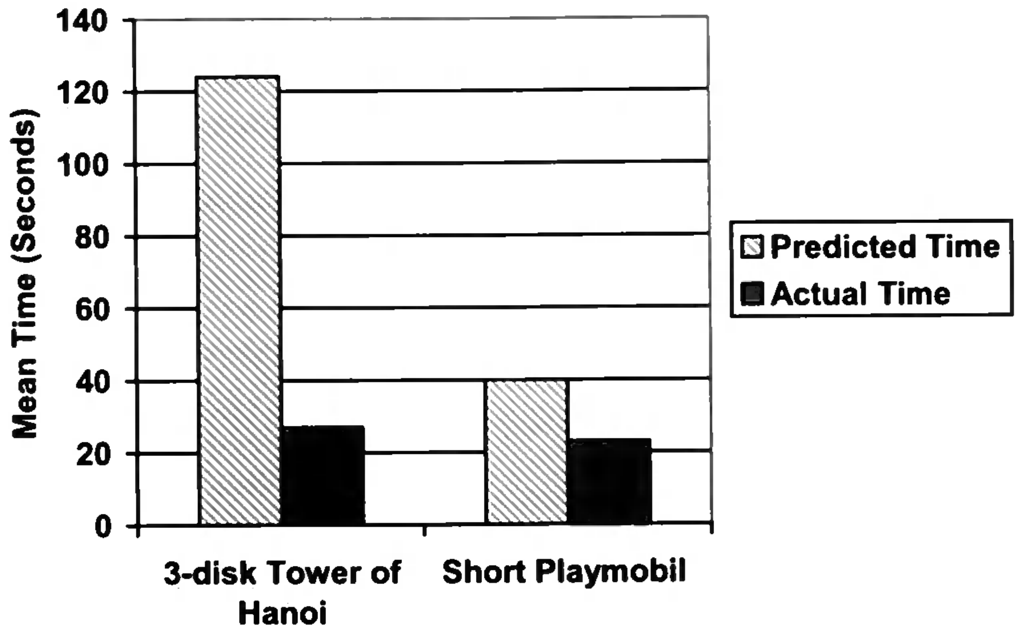

Figure 5.2

Predicted and actual completion times on each of the short duration tasks

In order to determine how many participants predicted task duration using whole minutes, frequency distributions of time estimates were calculated. Histograms of time estimates on the three-disk Tower of Hanoi, the short Playmobil® task and the long duration Playmobil® task (regardless of prior experience condition) are presented below in Figures 5.3, 5.4 and 5.5, respectively.

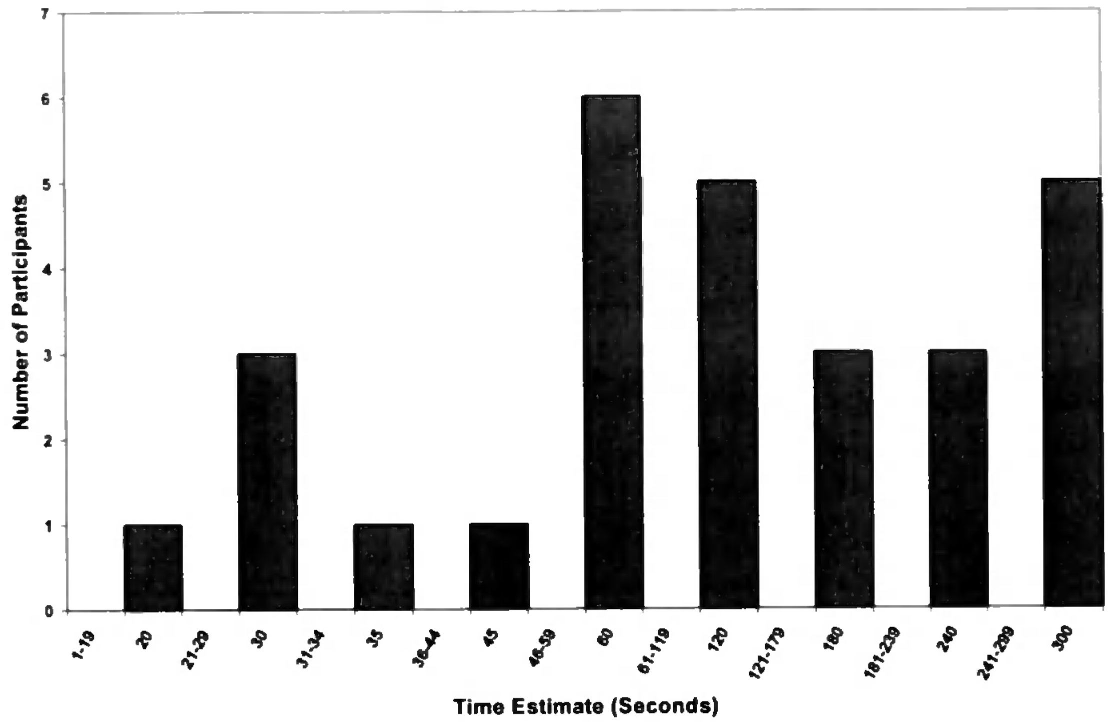

Figure 5.3

Histogram of time estimates on the three-disk Tower of Hanoi task

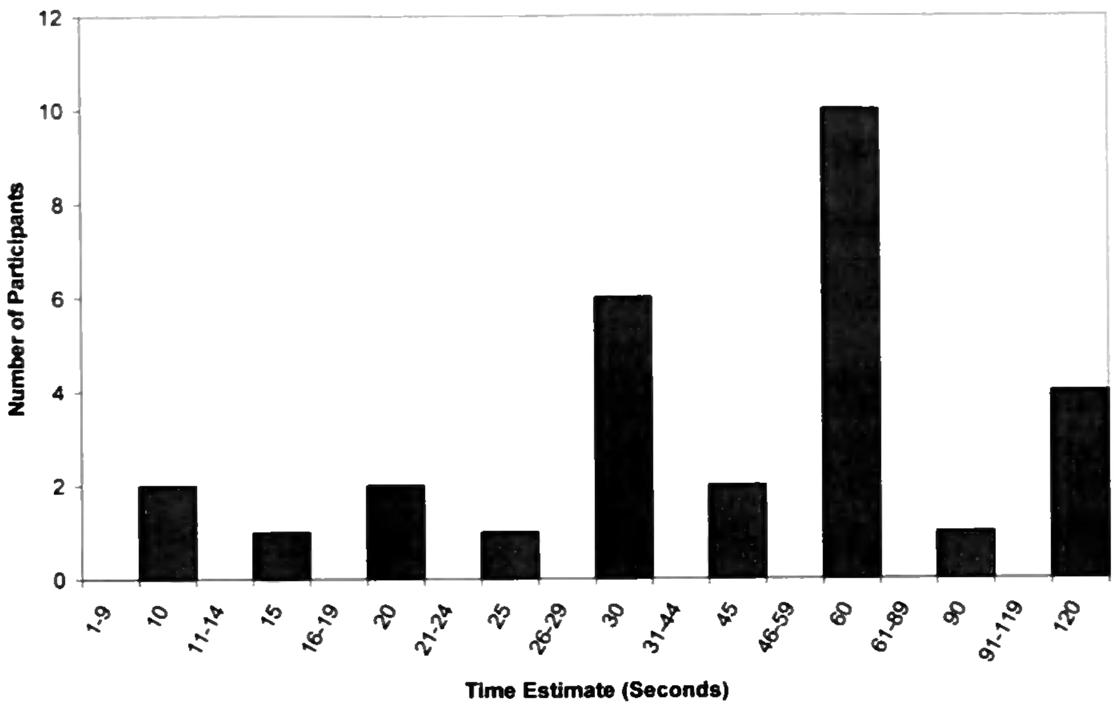

Figure 5.4

Histogram of time estimates on the short duration Playmobil® task

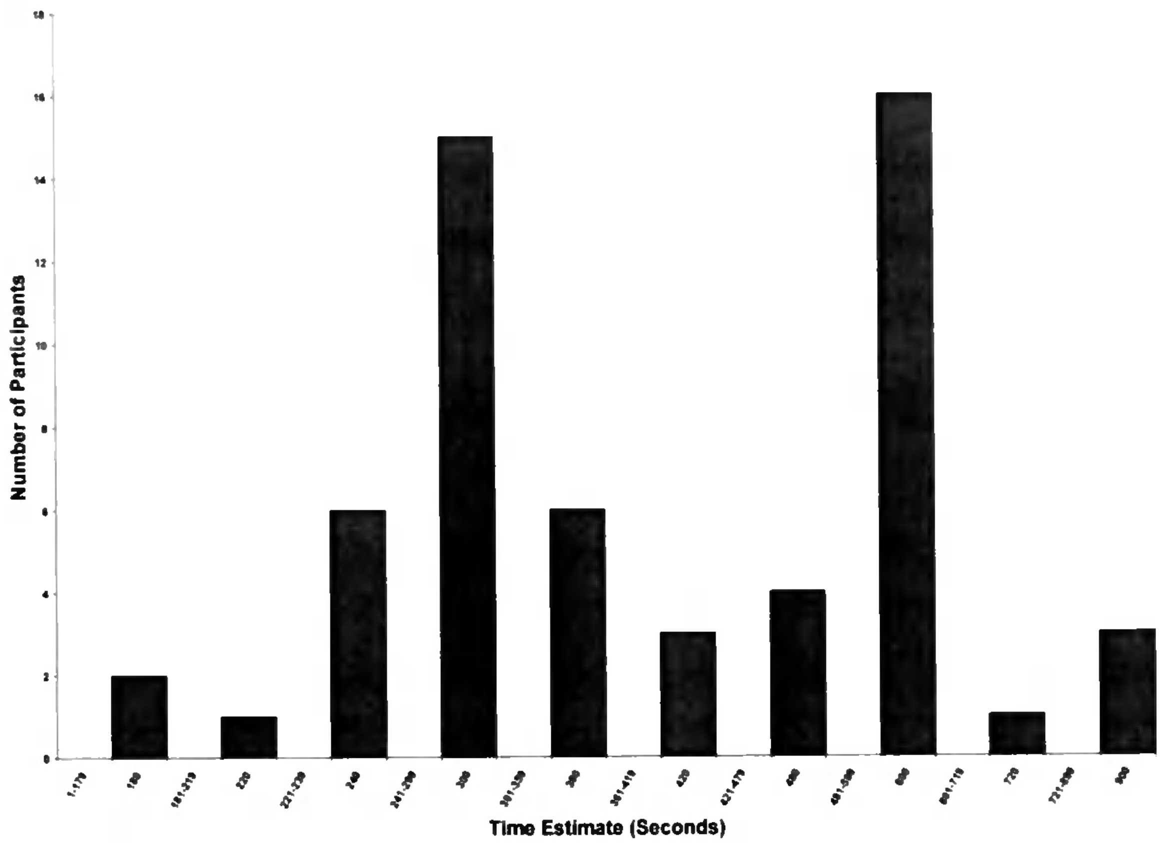

Figure 5.5

Histogram of time estimates on the long duration Playmobil® task

For the same reasons that were specified in Experiment 1, the estimated and actual completion time data were subjected to a logarithmic transformation before being statistically analysed. The issue of prior task experience was addressed by analysing the log-transformed data from the long duration Playmobil® task. A 2 (time) x 2 (prior experience) mixed design ANOVA produced a main effect of time, $F(1,54) = 61.81$, $MSE = .07$, $p < .001$, with completion times exceeding estimates (Ms = 606.56 and 422.68 seconds, respectively). This finding suggests that a general optimistic time prediction bias was evident on this task. However, the main effect of prior experience and the interaction were not significant ($Fs < 1$, $ps > .10$). The lack of a significant interaction indicates that the extent of temporal underestimation on the long task did not differ according to the type of short duration task that was performed beforehand.

In order to ascertain whether time prediction bias differed between the two short duration tasks, the log-transformed data from the first trial were analysed. A 2 (time: estimated vs. actual duration) x 2 (task: three-disk Tower of Hanoi vs. short duration Playmobil®) mixed design ANOVA produced a main effect of task, $F(1,54) = 15.56$, $MSE = .65$, $p < .001$, with overall time being longer on the three-disk task. There was also a significant main effect of time, $F(1,54) = 45.73$, $MSE = .40$, $p < .001$, with temporal estimates exceeding completion times (Ms = 82.22 and 24.93 seconds, respectively). This finding suggests that general temporal overestimation was evident on the three-disk task and the short duration Playmobil® task.

The interaction was also significant, $F(1,54) = 19.61$, $MSE = .40$, $p < .001$, and revealed that the difference between estimated and actual duration was greater on the three-disk task (see Figure 5.2 above). Follow-up analyses (LSD t-tests, both two-tailed) revealed that estimates and completion times differed significantly on the three-disk task ($p < .05$), but not on the short duration Playmobil® task ($p > .10$). These findings indicate that time predictions were less pessimistically biased on the short duration Playmobil® task.

5.6 Discussion

There was evidence of temporal underestimation on the long duration Playmobil® task when the three-disk Tower of Hanoi task was performed beforehand. Similarly, as predicted, temporal underestimation was evident on the long Playmobil® task when it was preceded by the short duration sub-component task. However, the extent of this optimistic time prediction bias did not differ according to the relevance of prior task experience. That is, relative to those individuals who performed the unrelated problem solving task beforehand, time predictions were no less optimistically biased among participants who had just constructed part of the Playmobil® castle. Thus, participants with prior experience of performing part of the long duration task were no more accurate at estimating their completion time on the second trial.

This finding suggests that information about the nature of the previous task was not used as a basis for estimating the duration of the long Playmobil® task. Instead, the perceived duration of the previous task seems to have been the type of information upon which time estimates on the long duration Playmobil® task were based. Specifically, as the extent of judgement bias was similar when the related or the unrelated task was performed beforehand, temporal underestimation on the long Playmobil® task may have been a consequence of time predictions being based on the shorter duration of the previous task.

The presence of an optimistic time prediction bias on the long duration Playmobil® task is consistent with the notion that participants used the anchoring and adjustment heuristics (Tversky & Kahneman, 1982a) when making a second temporal estimate. Given that time prediction bias did not differ according to the relevance of prior task experience, participants may have used an anchoring and adjustment judgement strategy involving the perceived duration of the previous task when making a second temporal estimate. That is, temporal underestimation was a consequence of time predictions being anchored on the perceived duration of the just-completed task, with insufficient upward adjustment for the greater demands of the current task.

An example of the insufficient adjustment component of such a judgement strategy is the erroneous assumption that each stage of the long duration task (i.e., each of the first eight pages of the Playmobil® instruction booklet) entailed an equal amount of work. Making this assumption should result in time predictions being insufficiently scaled up from the anchor value (i.e., the shorter duration of the previous task) because the last three stages of the task involve assembling many small plastic components, which takes more time than is the case with the earlier stages (e.g., when building the castle walls). Hence, temporal underestimation would be expected to occur on the long duration Playmobil® task if participants made such an assumption.

The anchoring component of such a judgement strategy may have taken the form of an estimate of the duration of the previous task or the recall of the time prediction given on the first trial. Whilst both of these sources of information qualify as potential numerical anchor values, anecdotal evidence suggests that many participants were aware that they had overestimated the duration of the first task. That is, at the end of the first trial, several individuals commented that the just-completed task had taken them less time than they predicted. Such awareness of temporal misestimation indicates that participants might have estimated the duration of the first task retrospectively. Thus, time predictions on the second task were presumably anchored on the perceived duration of the previous task rather than on the recall of the temporal estimate given before the first trial.

Given that the short duration Playmobil® task involved building only one wall of the castle (i.e., one wall of the long task), a lack of prior task experience might explain why time prediction accuracy was not greater when this related task was performed beforehand (compared to when the unrelated three-disk task was completed initially). As participants acquired little information about how to complete the long duration task whilst performing the short sub-component task (e.g., how certain plastic components slotted together), they may have used an anchoring and adjustment judgement strategy involving the perceived duration of the just-completed task. Likewise, in the absence of any prior experience, participants who had just completed the three-disk task might also use such a heuristic information processing strategy when estimating the duration of the long Playmobil® task. Consistent with these suggestions, it has been shown that people who have little or no prior task experience rely on the anchoring and adjustment heuristics when judging personal task performance (e.g., Mussweiler & Strack, 2000; Wilson et al., 1996).

As anticipated, temporal predictions exceeded completion times on the three-disk Tower of Hanoi task and the short duration Playmobil® task. However, the extent of this temporal overestimation differed between these tasks, with time predictions being significantly more accurate on the short duration Playmobil® task. A plausible explanation for this finding is that, when making an initial time estimate, participants in the short duration Playmobil® task condition possessed greater task-related information, which they used to good effect. That is, as the steps involved in performing this task were specified in the instruction booklet, participants were presumably aware of the completion process, and may have based their time estimate on task-related information (e.g., how the plastic task components slotted together).

In contrast, as the three-disk Tower of Hanoi task instruction sheet contained only a pictorial representation of the end state of the task, participants may not have been aware of the steps involved in task completion (e.g., the disk move pattern) when estimating its duration. Hence, due to a lack of task-related information, participants could have given an 'off-the-cuff' guess (Byram, 1997) concerning the duration of this task. Given that greater prior task experience has been shown to reduce temporal misestimation (Josephs & Hahn, 1995), it is unsurprising that time predictions were more accurate among participants who viewed the Playmobil® instruction booklet beforehand.

The present study suggests that the perceived duration of the just-completed task formed the basis of time estimates on the long duration Playmobil® task. However, it is not known whether time predictions on this task will be based on such information when an unrelated task that is of longer duration is performed initially. It could be that previous task duration only forms the basis of time predictions on current tasks that are of longer duration than previous ones. Since longer duration tasks tend to be more complex than shorter ones (Buckhalt & Oates, 2002), performing them is likely to require greater cognitive resources.

In the light of the findings of Zakay (1989), completing more complex (and longer duration) tasks may well leave few cognitive resources available to monitor temporal information. Thus, if individuals have little knowledge of the duration of a previous longer task, they might use information about the nature of the current task (e.g., the number of major components) when estimating its duration. In view of these suggestions, a key aim of Experiment 9 was to determine whether time predictions on the long duration Playmobil® task were based on the perceived duration of an unrelated longer task, which had just been completed.

Given the nature of the short duration Playmobil® task, it was suggested that previous task duration might only form the basis of time predictions when people possess little or no experience of an upcoming task. That is, due to a lack of prior task experience, time predictions might be anchored on the perceived duration of the previous task, but insufficiently adjusted according to the demands of the current task.

As prior task experience has been shown to improve time prediction accuracy (e.g., Experiment 1), having it may well enable people to use information about the nature of a previous task when estimating the duration of a current task. Indeed, there was no evidence of anchoring and adjustment when half of the long Playmobil® task was performed beforehand (Experiment 7), suggesting that predictions on the long task were not based on the shorter duration of the just-completed medium sub-component task. In order to ascertain whether previous task duration forms the basis of time predictions when people possess considerable prior task experience, some participants performed the long duration Playmobil® task twice in succession in Experiment 9.

Consistent with the previous experiments, the tendency to estimate task duration using whole minutes rather than seconds was apparent in the present study. For example, nearly all of the participants used whole minutes when judging the duration of the three-disk Tower of Hanoi task, whereas half of those individuals used whole minutes when estimating their completion time on the short Playmobil® task (see Figures 5.3 and 5.4, respectively). There was also evidence that the majority of participants used temporal units ranging from five to 15 minutes when judging their completion time on the long duration Playmobil® task (see Figure 5.5). Taken together, these findings indicate that in general participants did not estimate task duration in seconds, but preferred instead to use whole minutes or longer temporal units.

5.7 Experiment 9

The aims of this study were twofold: to further address the issue of the relevance of prior task experience; and to ascertain whether time prediction bias on the long duration Playmobil® task differed according to the relative duration of a just-completed unrelated task. The findings of Experiment 8 suggest that time predictions on the long duration Playmobil® task were based on the perceived duration of a previous shorter task. However, it is not known whether this kind of task-related information will be used when an unrelated task of longer duration is performed initially. Likewise, it is unclear whether such information will serve as a basis for subsequent time estimates when participants possess greater prior task experience than was the case in the previous study. It could be that people who have completed the same task beforehand are able to use information about the nature of the previous task when estimating the duration of the upcoming task.

In order to further address the issue of the relevance of prior task experience, a task that was either related or unrelated to the long duration Playmobil® task was performed beforehand in this study. The issue of relative task duration was addressed by employing tasks that were either shorter, longer or of similar duration to the long Playmobil® task. In fact, participants completed one of four tasks beforehand, three of which were unrelated (i.e., different) to the long duration Playmobil® task. Performing any one of these tasks initially would provide no information about the nature of the long duration Playmobil® task, whereas relevant prior experience would be acquired whilst completing the related task.

The unrelated tasks were the three-disk, six-disk and seven-disk versions of the Tower of Hanoi task. Pilot testing revealed that the six-disk task was of similar duration to the long duration Playmobil® task from Experiment 8 (Ms = 584.91 and 606.55 seconds, respectively), whereas the seven-disk task took longer to complete (M = 1013.23 seconds). As was the case in Experiment 8, the three-disk task was selected as the shorter duration unrelated task. The related task was the long duration Playmobil® task itself, which some participants thus performed twice in succession.

Since none of the Tower of Hanoi tasks provided any information about the nature of the long duration Playmobil® task, it was anticipated that time predictions on the second task would be based on the perceived duration of the unrelated first task. Given the findings of Experiment 8, it was hypothesised that performing the three-disk task initially would result in temporal underestimation on the long duration Playmobil® task. Conversely, temporal overestimation was expected to occur on the long duration Playmobil® task when it was preceded by the seven-disk task. That is, time predictions would be pessimistically biased because they were based on the longer duration of the just-completed unrelated task.

As the long duration Playmobil® task and the six-disk task were of similar duration, no prediction was made about the direction of time estimation bias on the second trial when these tasks were performed consecutively. Likewise, no prediction was made about the direction of time estimation bias when the long duration Playmobil® task was performed for a second time. However, as prior experience of performing the same task has been found to reduce time prediction bias (e.g., Experiment 4), temporal estimates were expected to be more accurate when the long duration Playmobil® task was performed for a second time rather than when one of the unrelated tasks had just been completed. Hence, it was anticipated that, when making a second time prediction, participants who had just completed the same task would use information about the nature of this task to good effect.

5.8 Method

5.8.1 Participants

Ninety (77 female and 13 male) students at the University of Plymouth participated voluntarily. Fifty participants took part in partial fulfilment of a psychology course requirement whilst the remainder were paid £4 each. No biographical information other than gender was recorded.

5.8.2 Materials

A word association checklist (Berkowitz & Troccoli, 1990) comprising 40 words was used. The other materials were identical to those used in Experiment 8 except that the Tower of Hanoi apparatus contained seven different-sized disks. The largest disk was removed to form the six-disk task, whereas the four largest disks were removed to form the three-disk task.

5.8.3 Design and Procedure

A 2 (task: first vs. second) x 2 (time: estimated vs. actual) x 4 (prior task experience: three-disk vs. six-disk vs. seven-disk vs. long duration Playmobil® task) mixed factorial design was used. The time and task factors were repeated-measures, whereas prior task experience was manipulated between groups. All participants performed two tasks as determined by random assignment to one of the four prior task experience conditions.

The procedure was similar to that of Experiment 8. Before estimating the duration of each trial, participants who performed the Playmobil® task twice were given 80 seconds to view the plastic task components and the instruction booklet. Participants who performed the six-disk Tower of Hanoi task were also given 80 seconds to view the task apparatus and the instruction sheet before making a time estimate. A 100-second planning period was operative on the seven-disk task. As was the case in Experiment 8, participants were given 20 seconds to preview the three-disk task and its instruction sheet. All participants were informed when 10 seconds of each pre-exposure period remained. In order to give the researcher time to dismantle the apparatus of the long duration Playmobil® task, participants who performed this task twice were required to complete a word association checklist (Berkowitz & Troccoli, 1990) for a period of two minutes at the end of the first trial. Each testing session lasted between 30 and 50 minutes.

5.9 Results

Table 5.2 below contains basic descriptive statistics of estimated and actual duration per task and prior task experience condition. All test statistics used in Experiment 9 are contained in Appendix 9.

Task		Prior Task Experience Condition			
		Three-disk Tower of Hanoi Task	Six-disk Tower of Hanoi Task	Seven-disk Tower of Hanoi Task	Long Duration Playmobil® Task
Second Task (Playmobil®)	Time Estimate	M = 508.70 Mdn = 480.00 SD = 200.53 N = 23	M = 575.22 Mdn = 600.00 SD = 119.54 N = 23	M = 683.18 Mdn = 720.00 SD = 300.28 N = 22	M = 462.73 Mdn = 450.00 SD = 132.42 N = 22
	Completion Time	M = 696.87 Mdn = 677.00 SD = 187.21 N = 23	M = 646.83 Mdn = 642.00 SD = 126.74 N = 23	M = 632.50 Mdn = 568.50 SD = 199.33 N = 22	M = 461.14 Mdn = 436.00 SD = 131.19 N = 22
First Task	Time Estimate	M = 112.38 Mdn = 120.00 SD = 78.94 N = 23	M = 456.52 Mdn = 420.00 SD = 224.31 N = 23	M = 598.64 Mdn = 600.00 SD = 298.35 N = 22	M = 715.91 Mdn = 720.00 SD = 219.93 N = 22
	Completion Time	M = 35.83 Mdn = 36.00 SD = 22.18 N = 23	M = 510.35 Mdn = 501.00 SD = 144.60 N = 23	M = 940.00 Mdn = 953.00 SD = 221.62 N = 22	M = 667.27 Mdn = 680.00 SD = 173.47 N = 22

Table 5.2

Time estimates and completion times per task and prior task experience condition (in seconds)

Table 5.2 shows that, on the second trial, temporal estimates exceeded completion times in the seven-disk task and the Playmobil® task twice conditions, whereas completion times exceeded temporal estimates in the three-disk and six-disk task conditions. On the first trial, Table 5.2 shows that temporal overestimation was evident on the three-disk task and the long duration Playmobil® task, whereas temporal underestimation occurred on the six-disk and seven-disk tasks. Bar graphs of predicted and actual completion times on the second and first tasks are presented below in Figures 5.6 and 5.7, respectively.

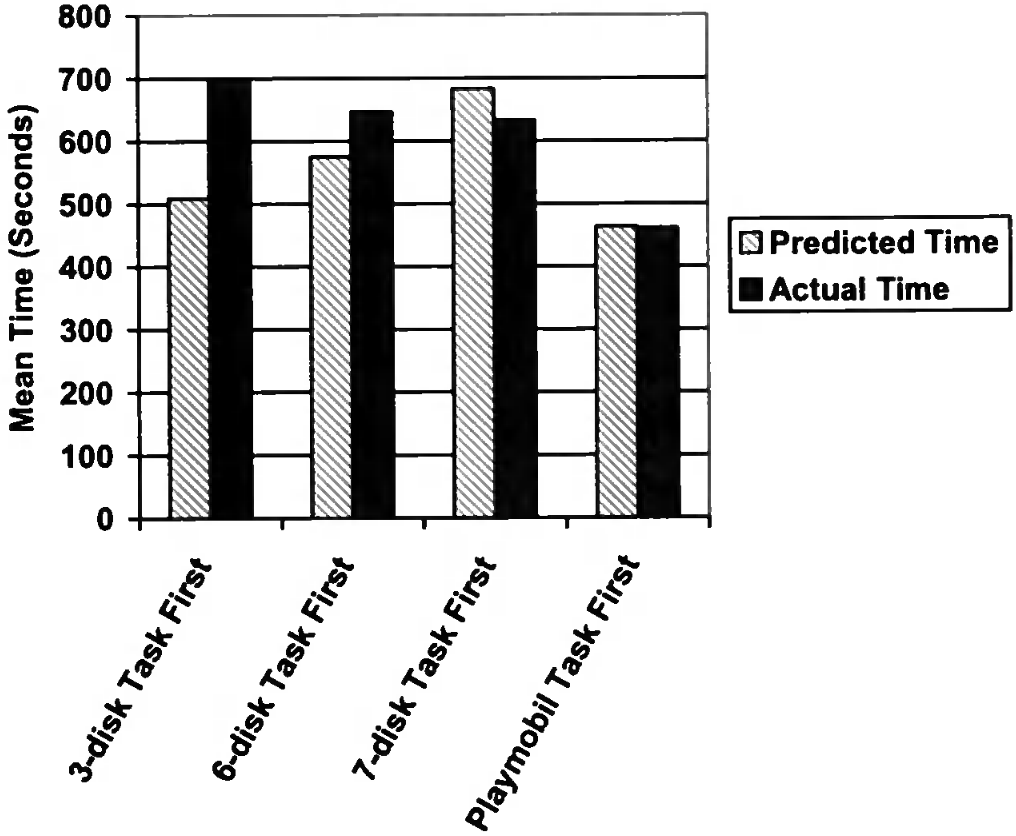

Figure 5.6

Bar graph of predicted and actual completion times per prior task experience condition (data from the second task)

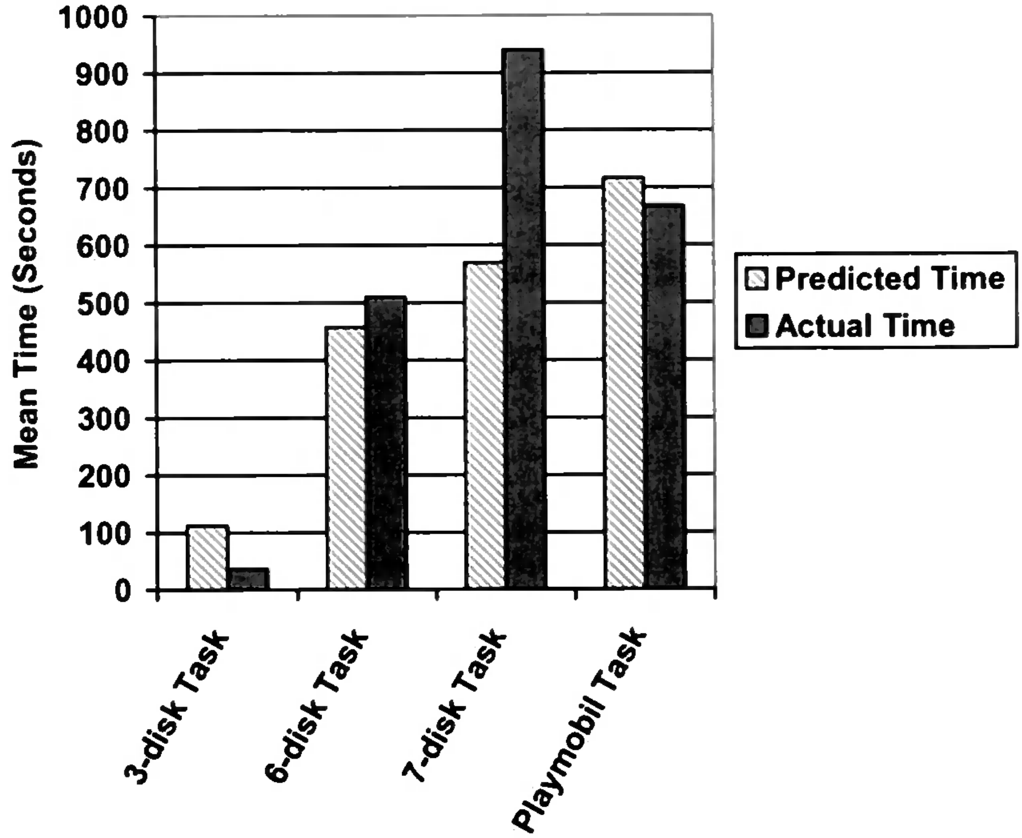

Figure 5.7

Bar graph of predicted and actual completion times per prior task experience condition (data from the first task)

In order to determine how many participants predicted task duration using whole minutes, frequency distributions of time estimates were calculated. Histograms of time estimates on the three-disk, six-disk, seven-disk and long duration Playmobil® tasks (regardless of prior experience condition) are presented below in Figures 5.8, 5.9, 5.10 and 5.11, respectively.

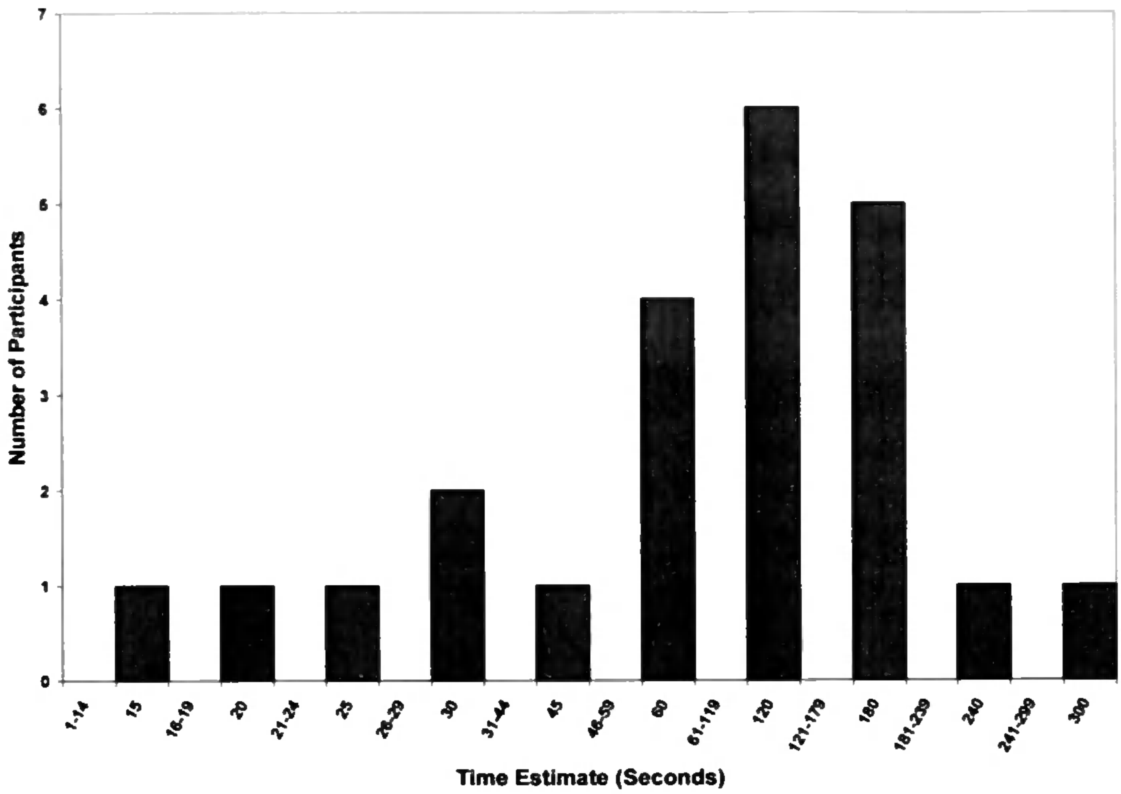

Figure 5.8

Histogram of time estimates on the three-disk Tower of Hanoi task

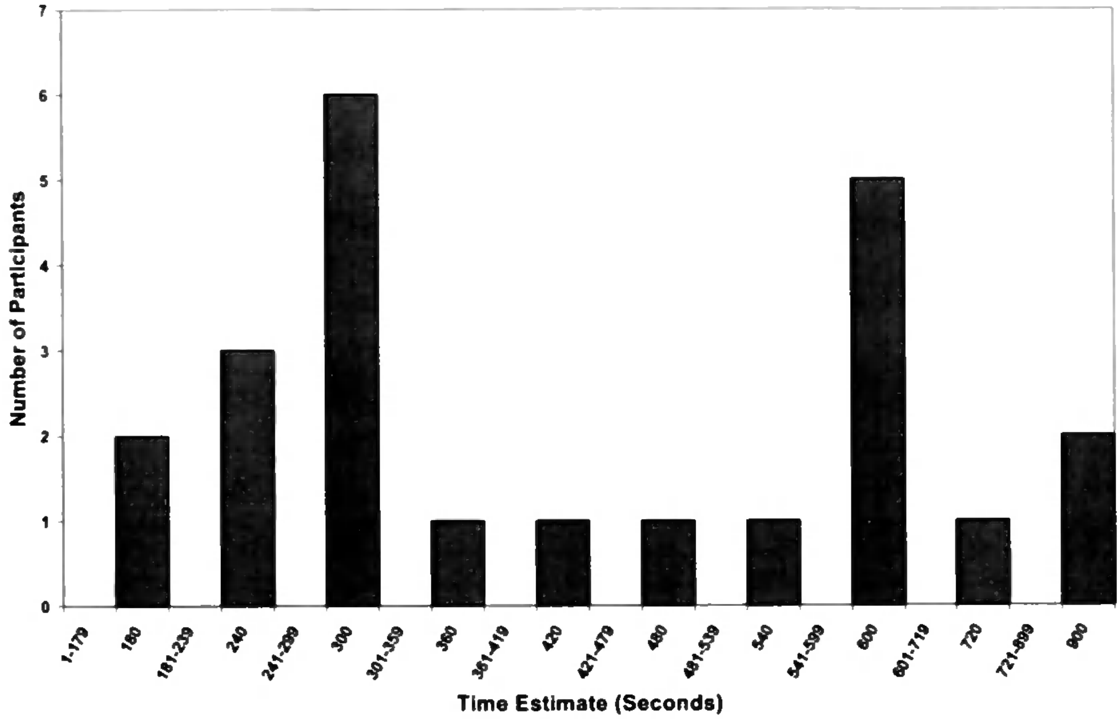

Figure 5.9

Histogram of time estimates on the six-disk Tower of Hanoi task

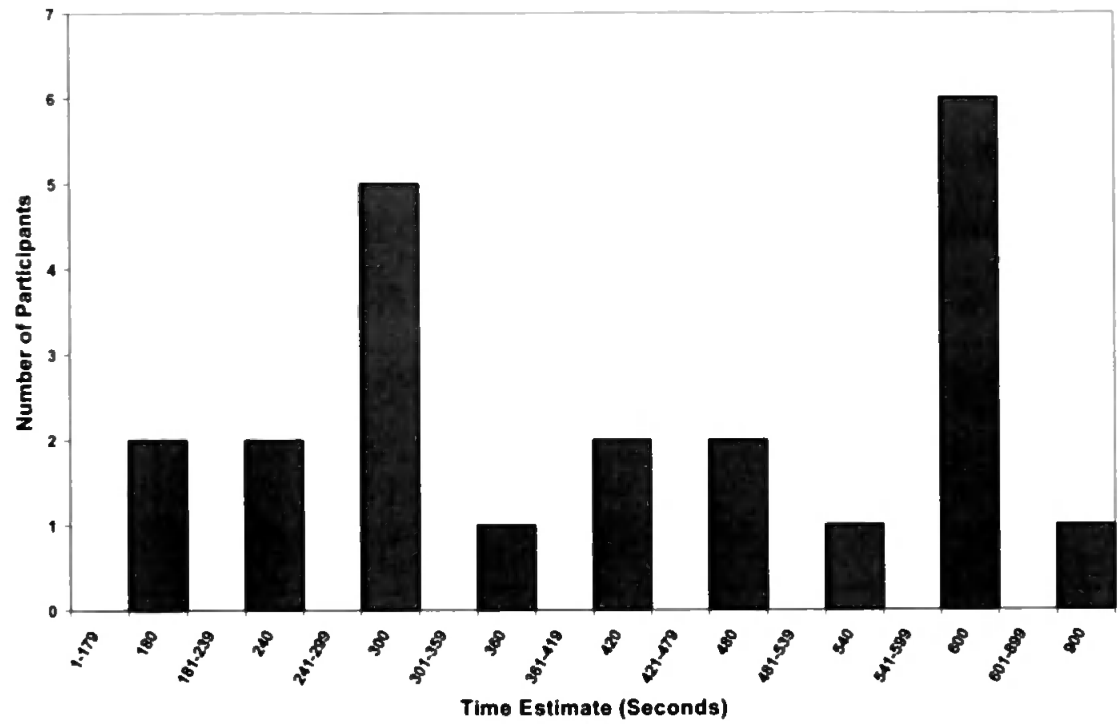

Figure 5.10

Histogram of time estimates on the seven-disk Tower of Hanoi task

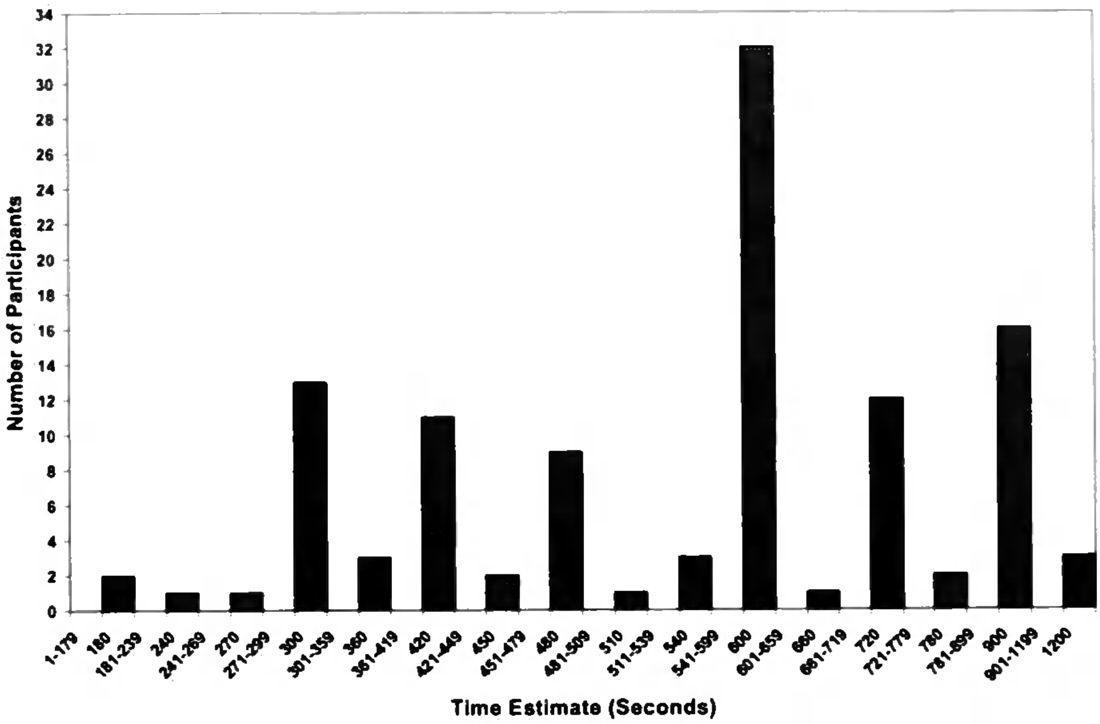

Figure 5.11

Histogram of time estimates on the long duration Playmobil® task

For the same reasons that were specified in Experiment 1, the estimated and actual completion time data were subjected to a logarithmic transformation before being statistically analysed. In order to address the issues of prior task experience and relative task duration, the log-transformed data from the second trial were analysed. A 2 (time) x 4 (prior task experience) mixed design ANOVA produced a main effect of time, $F(1,86) = 9.92$, $MSE = .06$, $p < .01$, with completion times exceeding estimates (Ms = 609.34 and 557.46 seconds, respectively). Hence, on the second trial, general temporal underestimation was evident on the long duration Playmobil® task. There was a significant main effect of prior task experience, $F(3,86) = 6.49$, $MSE = .15$, $p < .01$, which revealed that overall time was shortest in the Playmobil® task twice condition. Scheffé pairwise comparisons revealed that the mean of the Playmobil® task twice condition differed significantly from the means of the other conditions (ps < .05). No other condition mean differences were significant (ps > .10).

The interaction was also significant, F(3,86) = 5.41, MSE = .06, p < .01 (see Figure 5.6 above). This indicated that temporal underestimation occurred in the three-disk and six-disk task conditions, whereas temporal overestimation was evident in the seven-disk task and the Playmobil® task twice conditions. Planned comparisons (LSD t-tests, all two-tailed) revealed that estimates and completion times differed significantly in the three-disk and six-disk task conditions (ps < .05), but not in the seven-disk task and the Playmobil® task twice conditions (ps > .10).

In order to ascertain whether time prediction accuracy differed between the four tasks, the log-transformed data from the first trial were analysed. A 2 (time) x 4 (task) mixed design ANOVA produced a main effect of task, F(3, 86) = 164.60, MSE = .44, p < .001, which revealed that overall time was shortest on the three-disk task. Scheffé pairwise comparisons revealed that the mean of the three-disk task differed significantly from the means of the other tasks (ps < .05). All other condition mean differences were significant (ps < .05) except those of the Playmobil® task and the six-disk task (p > .05) and the Playmobil® task and the seven-disk task (p > .10). The main effect of time was not significant (F < 3, p > .10).

There was also a significant interaction, F(3, 86) = 40.54, MSE = .13, p < .001 (see Figure 5.7 above). This revealed that temporal underestimation was evident on the six-disk and seven-disk tasks, whereas temporal overestimation occurred on the three-disk task and the long duration Playmobil® task. Follow-up analyses (LSD t-tests, all two-tailed) revealed that estimates and completion times differed significantly on the three-disk and seven-disk tasks (ps < .05), but not on the six-disk task and the long duration Playmobil®

task (ps > .10). This finding indicates that the extent of temporal overestimation was only significant on the three-disk task, whereas the magnitude of temporal underestimation was only significant on the seven-disk task.

5.10 Discussion

There was evidence of temporal underestimation on the long duration Playmobil® task when it was preceded by one of two unrelated tasks that were of shorter duration. Specifically, participants who had just performed the three-disk or six-disk versions of the Tower of Hanoi task tended to make optimistically biased time predictions on the long duration Playmobil® task. Whilst pilot testing revealed that the six-disk and Playmobil® tasks were of similar duration, the latter task (when it was performed first in the experiment) took over two minutes longer to complete than the former task (see Table 5.2).

The presence of an optimistic time prediction bias on the long duration Playmobil® task suggests that participants who performed either the three-disk task or the six-disk task initially based their second time estimate on the perceived duration of the just-completed shorter task. Moreover, it could be that these participants used the anchoring and adjustment heuristics (Tversky & Kahneman, 1982a) when judging the duration of the long Playmobil® task. That is, time predictions may have been anchored on the perceived duration of the previous task, but insufficiently scaled up according to the greater demands of the long duration Playmobil® task (e.g., the greater number of task components). The use of these cognitive heuristics provides a plausible explanation for the presence of temporal underestimation on the Playmobil® task when it was preceded by either of these shorter duration tasks.

In the seven-disk task and the Playmobil® task twice conditions, there a hint of time predictions being pessimistically biased on the second trial. Since the first trial took more time to complete than the second trial in these conditions, there was some evidence of temporal overestimation on a shorter task when it was performed after a longer duration one. However, as the extent of temporal overestimation was not significant in these conditions, time predictions were more accurate when a longer rather than a shorter task had just been completed. That is, relative to those individuals who performed either the three-disk task or the six-disk task initially, the magnitude of temporal misestimation was less among participants who had just completed the seven-disk task or the long duration Playmobil® task.

A plausible explanation for this finding is that participants in the seven-disk task and the Playmobil® task twice conditions used information other than the perceived duration of the just-completed task when making a second time estimate. Whilst information about the nature of the first task (e.g., how the plastic components fitted together) should be expected to form the basis of temporal estimates when the long duration Playmobil® task was performed for a second time, no such information was acquired whilst completing the seven-disk task beforehand. Thus, it is unclear what kind of information served as a basis for time estimates on the long Playmobil® task when this longer duration unrelated task had just been performed. One possibility is that these individuals based their second time estimate on information about the nature of the long duration Playmobil® task, which they acquired whilst previewing the instruction booklet and the plastic task components. For example, they may have calculated the number of large plastic components involved in task completion and estimated how much time it would take to assemble these components.

Whilst this explanation could apply to all participants, those individuals who had just performed the seven-disk task might have used such information because they possessed little knowledge of the duration of the previous task. That is, they were unable to devote sufficient cognitive resources to monitoring the duration of the seven-disk task because of its complexity. As the completion of the seven-disk task involves making at least 127 disk moves, it is a complex task that is likely to require considerable cognitive resources. On the basis of Zakay's (1989) research, few cognitive resources should be available to process temporal information on such a task because non-temporal information (i.e., how to complete the task) becomes the focus of attention. Hence, given Zakay's findings, it could well be that participants were unaware of the duration of the seven-disk task, and may have based their second time estimate on information about the nature of the long duration Playmobil® task.

The importance of prior task experience in reducing time prediction bias was further highlighted in this study. That is, time predictions on the second task were most accurate among participants who had just performed the Playmobil® task. In fact, on average, these individuals overestimated the duration of the second task by less than two seconds (see Table 5.2), suggesting that their extensive prior experience was used to good effect. A plausible explanation for this finding is that these participants took account of their previous task performance when making a second time estimate. As both tasks were identical, they were presumably aware of factors that delayed the completion of the first task (e.g., fitting plastic components together incorrectly), and may have based their second time estimate on such task-related information.

Alternatively, participants who performed the long duration Playmobil® task twice might have based their second time prediction on information about the nature of the previous task. For example, they may have estimated the number of large plastic components and made a second time prediction on the basis of this information. Given that time estimation bias has been found to be reduced when people recognise the similarities between previous and current tasks (Buehler et al., 1994), the use of both of these judgement strategies should result in greater time prediction accuracy when the long duration Playmobil® task was performed for a second time.

There was further evidence of the tendency to estimate task duration using whole minutes rather than seconds in the present study. The frequency distributions of time estimates on all four tasks revealed that the vast majority of the participants used whole minutes when judging their completion times (see Figures 5.8 to 5.11). There was also evidence that most participants used temporal units ranging from five to 15 minutes when judging the duration of the six-disk, seven-disk and long duration Playmobil® tasks. Consistent with previous studies (e.g., Experiment 8), these findings indicate that in general participants preferred to use whole minutes or longer temporal units when judging the duration of the present tasks.

5.11 General Discussion

The present studies sought to determine the type of task-related information upon which time estimates on the second of two consecutive tasks are based. In Experiment 8, the extent of temporal underestimation on the long duration Playmobil® task did not differ according to the relevance of prior task experience. That is, time predictions were no less optimistically biased when a related rather than an unrelated task of about 30 seconds' duration had just been completed. This finding suggests that the perceived duration of the just-completed task was the type of information upon which time predictions on the long duration Playmobil® task were based. Consistent with this suggestion, Experiment 9 produced further evidence of temporal underestimation on this task when it was preceded by either one of two tasks, which were of shorter duration and unrelated to it.

Although there was a hint of temporal overestimation on the long duration Playmobil® task when it was performed after an unrelated task that was of longer duration, the extent of this pessimistic time prediction bias was not significant in the seven-disk task condition in Experiment 9. Likewise, the magnitude of temporal overestimation on the long duration Playmobil® task was not significant when it was performed for a second time in succession in that study. These findings suggest that participants who performed an identical or an unrelated longer task initially may not have based their second time estimate on the perceived duration of the just-completed task.

The complexity of the seven-disk task might explain why the perceived duration of this task did not form the basis of time estimates on the long duration Playmobil® task. That is, participants who performed the seven-disk task initially possessed little knowledge of its duration because they had to devote considerable cognitive resources to processing non-temporal information in order to complete this task. Instead of using previous task duration, these individuals may have based their second time estimate on information about the nature of the long duration Playmobil® task, which they acquired during the pre-exposure period. Conversely, it was suggested that performing the long duration Playmobil® task beforehand enabled participants to base their second time estimate on information about the nature of the previous task (e.g., the number of large plastic task components). Since time predictions were most accurate when this task was performed for a second time, these participants presumably used their extensive prior task experience to good effect.

In contrast to research supporting the planning fallacy (e.g., Buehler et al., 1994), the present studies suggest that taking account of distributional information such as personal performance on previous tasks can result in temporal underestimation. That is, when a related or an unrelated shorter duration task had just been completed, time predictions were optimistically biased. This finding indicates that, under such conditions, people use the perceived duration of the previous task as a mechanism for temporal estimation. Moreover, it suggests that erroneous distributional information such as how much time a previous unrelated task took to complete is used when estimating task duration. Given the findings of Experiments 3 to 9, it could be that the temporal underestimation indicative of the planning fallacy is a consequence of time predictions being based on the shorter duration of a just-completed task.

Whilst it is not known whether these findings will apply to other tasks, temporal underestimation may well be a consequence of people's use of the anchoring and adjustment judgemental heuristics. Specifically, time predictions might be anchored on the perceived duration of a previous shorter task, but insufficiently scaled up or adjusted according to the greater demands of an upcoming task. Conversely, temporal overestimation could be due to time predictions being anchored on the perceived duration of a just-completed longer task, with insufficient downward adjustment for the lesser demands of a current task. However, given the lack of a significant pessimistic time prediction bias when the long duration Playmobil® task was preceded by the seven-disk task or the same task (Experiment 9), anchoring and adjustment in the context of time estimation could be uni-directional. That is, these cognitive heuristics might only used when people estimate the duration of a task that takes more time to complete than a previous one.

It could be that people who have prior experience of performing a longer or more complex task engage in more thorough information processing when making a subsequent time estimate. For example, instead of relying on previous task duration, individuals might think about the steps involved in completing the current task and base their time prediction on this information. On the basis of previous research (e.g., Sniezek et al., 1990), judgements of task performance would be expected to be more accurate as a consequence of using such task-related information. Likewise, having prior task experience should enable people to use more in-depth judgement strategies instead of relying on cognitive heuristics (e.g., Smith & Kida, 1991; Wilson et al., 1996) when estimating task duration.

There was further evidence of the benefits of prior task experience in Experiment 9, where time predictions were very accurate when the Playmobil® task was performed for a second time. This finding suggests that people do take account of their performance on previous tasks, and use correct distributional information to good effect. That is, people make good use of information about personal performance on a previous task when this task is directly relevant to the upcoming task.

The present research indicates that, in order to improve time prediction accuracy, all (e.g., Experiment 4) or a substantial part of a current task must have been performed beforehand (Experiment 7). However, when people lack extensive prior task experience, it could well be that they rely on erroneous distributional information (e.g., the perceived duration of a previous unrelated task) when estimating their completion time on a task that is longer or more complex than a previous one. The use of such incorrect information should lead to temporal underestimation if people rely on the anchoring and adjustment heuristics when judging task duration (Buehler et al., 1995).

A similar kind of judgement strategy has in fact been found to be used by people who possess little prior task experience. Josephs and Hahn (1995) report that participants used an anchoring judgement strategy when estimating the duration of tasks such as proof-reading a manuscript. Specifically, time predictions were based on superficial task characteristics such as the number of pages, at the expense of information that would affect task duration such as the amount of text on each page. Josephs and Hahn found that this low cognitive effort judgement strategy resulted in temporal underestimation, and suggest that it was used because participants were unsure of the amount of work involved in completing the upcoming task. Hence, due to a lack of prior task experience, temporal underestimation was a consequence of these individuals engaging in heuristic information processing.

The amount of prior task experience could explain why time predictions were very accurate when the long duration Playmobil® task was completed for a second time (Experiment 9), whereas the extent of temporal underestimation on this task was not reduced when a shorter duration sub-component version of it was performed beforehand (Experiment 8). That is, greater information about the nature of the second task was acquired whilst performing the long (Experiment 9) rather than the short duration Playmobil® task initially (Experiment 8). Having just performed an identical task (i.e., the long duration task), participants in Experiment 9 presumably took account of information about the task completion process, which they used as a basis for their second time estimate.

In contrast, having built only one wall of the Playmobil® castle beforehand (i.e., the short duration sub-component task), participants in Experiment 8 possessed much less task-related information when estimating the duration of the long Playmobil® task. Hence, due to limited knowledge of the upcoming task, these individuals probably relied on the perceived duration of the previous task when making a second time estimate. Consistent with the work of Josephs and Hahn (1995), time estimates would be expected to be less biased when people have greater prior experience of performing a task. Thus, it is unsurprising that time predictions were more accurate when all rather than part of the long duration Playmobil® task had just been completed.

The present studies provide further evidence of the impact of the duration of a current task on the direction in which time predictions on it are biased. Specifically, there was evidence of temporal underestimation on the seven-disk Tower of Hanoi task (Experiment 9). Moreover, temporal underestimation occurred on the long duration Playmobil® task, but only when it was preceded by a related (Experiment 8) or an unrelated task that was of shorter duration (Experiments 8 and 9). Conversely, temporal overestimation was evident on the three-disk Tower of Hanoi task (Experiments 8 and 9) and there was a hint of a pessimistic time prediction bias on the short duration Playmobil® task (Experiment 8).

These findings indicate that people tend to make pessimistically biased time predictions on laboratory tasks that take less than one minute to complete. Conversely, the general temporal underestimation indicative of the planning fallacy may only be evident on laboratory tasks that take in excess of nine minutes to complete. That is, time predictions are optimistically biased on tasks that are of longer duration than the six-disk version of the Tower of Hanoi task, which took about eight and a half minutes to finish (see Table 5.2). It could be that a pessimistic time prediction bias occurs on shorter tasks because of the tendency to estimate task duration using whole minutes or temporal units such as 10 or 15 minutes (Fraisse, 1984).

The frequency distributions of time estimates provide support for the notion that task duration tends to be judged using whole minutes or temporal units such as 10 or 15 minutes. For example, in Experiment 8, all participants predicted the duration of the long Playmobil® task in whole minutes, and several of them used temporal units of 10 and 15 minutes (see Figure 5.5). Similarly, over half of the participants estimated their completion time on the three-disk Tower of Hanoi task and the short duration Playmobil® task in whole minutes rather than seconds in Experiment 8 (see Figures 5.3 and 5.4, respectively).

A similar pattern of time estimation was observed in Experiment 9, where the majority of participants in the three-disk task condition judged the duration of the three-disk task in whole minutes (see Figure 5.8). Likewise, all individuals used whole minutes when predicting the duration of the six-disk, seven-disk and long duration Playmobil® tasks in that study, and several of them used temporal units such as 10 and 15 minutes on these longer duration tasks (see Figures 5.9, 5.10 and 5.11, respectively). Taken together, the findings of the present studies indicate that whole minutes or temporal units such as 10 or 15 minutes tend to be used when estimating the time needed to complete tasks that range in duration from about 30 seconds to 16 minutes.

5.12 Conclusions

The research presented in this chapter provides further clarification of the role of prior task experience in the time estimation process. It has been shown that the extent of temporal underestimation on the long duration Playmobil® task does not differ when a related or an unrelated task, which was of much shorter duration was performed beforehand (Experiment 8). Similarly, prior experience of performing either one of two unrelated shorter tasks (i.e., the three-disk and six-disk versions of the Tower of Hanoi task) also led to temporal underestimation on the long duration Playmobil® task (Experiment 9).

These findings indicate that the perceived duration of a just-completed shorter task may have been the type of information upon which time estimates on the long duration Playmobil® task were based. It was suggested that an anchoring and adjustment judgement strategy might have been responsible for temporal underestimation on this task. Specifically, the perceived duration of the just-completed shorter task served as an anchor for time predictions, which were insufficiently adjusted according to the greater demands of the long duration Playmobil® task.

In contrast, there was little evidence of this kind of heuristic information processing when a longer duration task was performed before the long Playmobil® task (Experiment 9). Whilst there was a hint of temporal overestimation when the same or an unrelated task that was of longer duration was completed initially, the magnitude of this pessimistic time prediction bias was not significant. Thus, relative to participants who performed a shorter task beforehand, individuals who had just completed a longer task were more accurate at estimating the duration of the long duration Playmobil® task (Experiment 9). It was suggested that, when both tasks are identical, information about the nature of the first task served as a basis for time estimates on the second task. However, when an unrelated complex task had just been completed, it was proposed that time estimates may have been based on information about the nature of the upcoming task (i.e., the long duration Playmobil® task).

The present findings indicate that people can and do take account of their performance on previous tasks when estimating the duration of a current task. However, this research suggests that the use of such distributional information (Kahneman &

Tversky, 1979) does not necessarily improve time prediction accuracy, but can lead to temporal underestimation when it is erroneous. That is, when irrelevant information such as the perceived duration of a previous unrelated task forms the basis of time predictions on a longer task.

On the other hand, the present studies suggest that such distributional information can improve time prediction accuracy, but only when people have substantial experience of performing the same task before estimating its duration. Hence, in addition to highlighting the type of information that is used when estimating the duration of subsequent consecutive tasks, the present research indicates that the use of distributional information (i.e., personal performance on previous tasks) does not always lead to more accurate time predictions.

Chapter Six

General Discussion

6.1 Overview

The research presented in this thesis investigated a number of factors that might influence the accuracy of people's estimates of their task completion times. The nine experiments have produced several potentially important findings, and have highlighted two distinct factors that influence the direction (i.e., under or overestimation) and/or extent of time prediction bias. One factor relates to the task itself, namely, its duration, whereas the other factor concerns the person making the time estimate, that is, their prior experience of the to-be-completed task, or of other related or unrelated tasks.

In the final chapter of this thesis, the findings of the present studies will be discussed and potential theoretical and practical implications highlighted. The other aims of this chapter are to evaluate the strengths and limitations of the nine studies and make recommendations for future research. The chapter begins with a section that summarises the findings of the nine experiments. In the next section, potential explanations for the present findings are proposed. Some strengths and limitations of this research are discussed in the next section, which is followed by a section containing recommendations for future research. In the final two sections of the chapter, practical and theoretical implications of the present findings are emphasised.

6.2 Summary of the Findings

The present research provides a remarkably consistent picture concerning the process of estimating the duration of the three-disk version of the Tower of Hanoi task. In all six of the studies where this task was employed, there was no evidence of an optimistic time prediction bias on it. Indeed, the temporal underestimation indicative of the planning fallacy (Kahneman & Tversky, 1979) was not only absent but was reversed on this task. That is, in Experiments 1 to 4, and 8 and 9 (and in every cell in every condition in these studies), there was evidence of temporal overestimation on this simple well structured task.

This simple finding is of considerable importance since the planning fallacy has previously been found in virtually all published studies (e.g., Buehler et al., 1994; Newby-Clark et al., 2000), and with a variety of tasks including origami (Byram, 1997) and college coursework assignments (e.g., Koole & Van't Spijker, 2000). Moreover, this finding indicates that the planning fallacy is not as prevalent a cognitive judgement phenomenon as previously suspected, and that there is at least one task on which it definitely does not occur.

Given the simplicity of the three-disk task, the issue of task difficulty was addressed in Experiments 2 to 4, where the cognitive complexity of the Tower of Hanoi task was increased. In these studies, the four-disk (Experiments 2 and 3) and five-disk tasks (Experiment 4) were employed alongside the three-disk task. There was evidence that, relative to the three-disk task, time predictions were less pessimistically biased on the four-disk and five-disk tasks when they were performed first in Experiments 2 and 4, respectively. Specifically, the difference between estimated and actual duration was significant on the three-disk task, but not on the two more complex versions of the Tower of Hanoi.

In Experiments 2 to 4, there was some evidence of an optimistic time prediction bias on the four-disk and five-disk tasks, but only when they were preceded by the three-disk task. The simplicity of the three-disk, four-disk and five-disk tasks relative to the laboratory tasks used in previous research (e.g., Byram, 1997) was proposed as an explanation for the absence of a general optimistic time prediction bias in Experiments 1 to 4. However, as the laboratory tasks used in previous research also took more time to complete, task duration could provide an alternative explanation for the findings of Experiments 1 to 4. That is, the shorter duration of the three-disk, four-disk and five-disk tasks could be responsible for the general temporal overestimation that was observed on these tasks.

The principal aim of Experiment 5 was to ascertain whether task duration or cognitive complexity was responsible for the findings of the previous four studies. In order to achieve this goal, task duration was manipulated whilst cognitive complexity was held constant. Experiment 5 revealed that general temporal overestimation was evident on two versions of a simple repetitive disk movement task, which were of similar duration to the three-disk and five-disk versions of the Tower of Hanoi task. This finding was consistent with the notion that the shorter duration of the present tasks was responsible for the absence of a general optimistic time prediction bias in Experiments 1 to 5. Hence, the temporal underestimation that was observed in previous research (e.g., Francis-Smythe & Robertson, 1999) could be due to the longer duration of the laboratory tasks used in those studies.

Further support for this notion was evident in Experiment 6, where the direction in which time predictions were biased differed according to the duration of laboratory tasks that are less artificial than the Tower of Hanoi. That is, temporal underestimation was evident on a Playmobil® construction task that was of similar duration to some of the laboratory tasks used previously (e.g., Buehler et al., 1997), whereas temporal overestimation prevailed on another Playmobil® construction task that was of similar duration to the five-disk version of the Tower of Hanoi task.

The impact of the duration of a current task on the direction of time prediction bias was also highlighted in Experiment 9. Here, an optimistic time prediction bias was evident on a different type of laboratory task that was of similar duration to some of the ones used previous research (e.g., see Byram, 1997). Specifically, there was evidence of temporal underestimation on the seven-disk task version of the Tower of Hanoi, a task that took approximately 16 minutes to complete. Likewise, there was a hint of an optimistic time prediction bias on the six-disk version of the Tower of Hanoi, a task that took between eight and nine minutes to complete. Given these findings, it could be that laboratory tasks need to be of at least the duration of the six-disk task in order for the general temporal underestimation indicative of the planning fallacy to occur. Whilst it is not known whether an optimistic time prediction bias will prevail on other tasks of this duration, the present research suggests that task duration is a factor that influences the direction in which temporal judgements are biased (i.e., under or overestimation).

In contrast to previous studies (e.g., Koole & Van't Spijker, 2000; Newby-Clark et al., 2000), the present research emphasises the directional nature of time estimation bias. Temporal overestimation was observed on tasks such as the three-disk Tower of Hanoi, whereas underestimation prevailed on tasks such as the seven-disk Tower of Hanoi. Hence, in relation to well structured laboratory tasks at least, there is evidence that the direction in which time estimates are biased differs according to task duration. That is, temporal overestimation was evident on the tasks that took less than one minute to complete, whereas temporal underestimation occurred on the tasks that were in excess of eight minutes' duration. Given this finding, the longer duration of the laboratory tasks used in previous research (e.g., Byram, 1997) may well have been responsible for the general temporal underestimation that was observed in those studies.

The present research demonstrates that prior task experience is an important determinant of time prediction bias. In fact, three distinct types of prior task experience have been found to influence temporal misestimation. Firstly, there was evidence of the benefits of mental planning on time prediction accuracy on the three-disk and four-disk versions of the Tower of Hanoi task. That is, in Experiments 1 and 2, previewing the three-disk task and its instructions for two minutes resulted in a significant reduction in temporal overestimation. Likewise, this method of pre-exposure led to a significant reduction in time prediction bias on the four-disk task in Experiment 2. Here, the extent of temporal underestimation in the mental planning condition was less than the magnitude of temporal overestimation in the no experience condition. These findings indicate that time predictions were more accurate among individuals who previewed the three-disk or four-disk tasks beforehand.

It was suggested that these findings occurred because participants formed some kind of mental representation of the task completion process during the two-minute pre-exposure period. For example, they might have worked out the correct disk move pattern, and calculated the amount of time that would be needed to move the disks accordingly. Given the findings of Davies (2000a), it could be that mental preplanning is precluded only on the five-disk task and more complex versions of the Tower of Hanoi. If this were the case, then having the opportunity to preview the three-disk and four-disk tasks may well have provided participants with pertinent information upon which to make more accurate time estimates.

Active task practice was the second type of prior experience to be highlighted in the present research. This method of pre-exposure was studied initially in Experiment 1, which produced some evidence of temporal overestimation being reduced among participants who performed the three-disk task at least once beforehand. Similarly, time prediction accuracy was found to be greater on the three-disk and five-disk tasks (Experiment 4) and the long duration Playmobil® task (Experiment 9) when they were performed for a second time in succession. These findings indicate that having prior experience of performing an upcoming task enabled participants to make a more accurate single (Experiment 1) or second time prediction (Experiments 4 and 9). Moreover, these findings suggest that people can and do take account of their performance on previous tasks, and use such distributional information to good effect.

Prior performance of a longer or a shorter duration task was the third type of task experience that was found to influence time prediction bias. Specifically, there was evidence that time prediction bias on the second task differed according to the relative duration of the first task in Experiments 3 to 9. In general, temporal underestimation occurred on longer duration tasks when they were preceded by shorter ones, whereas there was some evidence of temporal overestimation being greater when shorter duration tasks were performed after longer ones.

Since different versions of the same task were performed in the majority of the present studies, it could be that time predictions on the current task were based on information about the nature of the just-completed task. For example, the disk moves needed to complete the three-disk task may have served as a basis for time predictions on the four-disk task (Experiment 3). However, the effects of prior task performance were also observed when no specific information about the nature of the second task was acquired whilst performing the first task. That is, when the first and second experimental trials were not different versions of the same task (Experiment 6).

Temporal overestimation was greater when the short duration Playmobil® task was preceded by the long Playmobil® task, whereas there was a hint of temporal underestimation being greater when the long duration Playmobil® task was performed second rather than first in Experiment 6. Given that two different tasks were performed consecutively in this study, it could be that time predictions on the second task were based on information such as the perceived duration of the first task. Specifically, participants may have estimated the duration of the just-completed task, and used this figure as a basis for making a time estimate on the current task.

In order to determine whether previous task duration or information about the nature of the just-completed task formed the basis of time predictions on the current task, the relevance of prior task experience was manipulated in Experiments 8 and 9. To this end, tasks that were either related or unrelated to each other were performed in succession. If time estimates were based on information about the nature of the previous task, then predictions should be more accurate when a related task had just been completed. Conversely, if time estimates were based on previous task duration then temporal prediction bias should not differ according to the relevance of prior task experience. Experiment 8 provided support for the latter hypothesis, as the extent of temporal underestimation on the long duration Playmobil® task did not differ when a related or an unrelated shorter task had just been performed. Likewise, Experiment 9 revealed that time predictions on the long duration Playmobil® task may have been based on the perceived duration of one of two shorter unrelated tasks, which had just been completed.

In contrast, there was no evidence that time prediction bias differed significantly according to the relative duration of a previous task when one of two shorter tasks was performed after a longer task (Experiment 9). Specifically, the extent of temporal overestimation on the long duration Playmobil® task was not significant when the seven-disk Tower of Hanoi task or the long duration Playmobil® task itself had just been completed. Given this finding, it was suggested that individuals who performed these tasks initially may not have based their second time estimate on the perceived duration of the previous task. Instead, information about the nature of the previous task may have served as a basis for temporal predictions on the long duration Playmobil® task when it was performed for a second time. Conversely, it was suggested that individuals possessed little knowledge of the duration of the seven-disk task because of its complexity, and so used information concerning the nature of the long duration Playmobil® task when estimating its duration.

The findings of Experiment 9 indicate that when a related task of longer duration was performed beforehand, time predictions on the second trial (i.e., the long duration Playmobil® task) may have been based on information about the nature of the previous task. Conversely, when a longer duration unrelated task (i.e., the seven-disk Tower of Hanoi) was completed before the long duration Playmobil® task, it was suggested that time predictions on the second trial might have been based on information about the nature of the upcoming task, which was acquired during the pre-exposure period (Experiment 9).

There was also evidence that time predictions on the long duration Playmobil® task might have been based on the perceived duration of a just-completed shorter task, which was either unrelated or related to it (Experiments 8 and 9). Given that temporal underestimation occurred when only one part of the long duration Playmobil® task (i.e., the short duration sub-component task) had just been completed in Experiments 7 and 8, it could be that people need to possess considerable prior task experience if information about the nature of a previous shorter duration task is to be used to improve time prediction accuracy.

Consistent with this suggestion, there was evidence of information about the nature of the previous task being used to good effect when half of the long duration Playmobil® task was performed initially (Experiment 7). That is, time predictions on the long duration task were more accurate when the medium rather than the short sub-component task had just been completed. Indeed, a pessimistic rather than an optimistic time prediction bias occurred when the long task was preceded by the medium task in this study, suggesting that the shorter duration of the first task did not form the basis of temporal estimates on the second task. Given the present findings, it could be that previous task duration only forms the basis of time estimates when people possess little or no prior task experience. That is, when an unrelated or a related task of much shorter duration than the upcoming task has just been performed.

6.3 Explaining the Findings

The role of distributional information in determining the planning fallacy (Kahneman & Tversky, 1979) implies that information about previous task performance tends to be neglected when estimating task duration. However, the present research highlights conditions under which such information is not only considered, but is also used to good effect. That is, in Experiments 1, 4 and 9, time prediction accuracy was greater among participants who had some prior experience of performing the same task.

This finding is consistent with the work of Buehler et al. (1994), which revealed that temporal underestimation was reduced (and time prediction accuracy improved) when individuals recognised the similarities between previous tasks and the one that they were about to perform. Buehler et al.'s experimental manipulation presumably gave participants the opportunity to recall distributional information, which they used effectively. Since prior task experience was acquired immediately in the present research, participants probably recognised the similarities between the previous and upcoming tasks, and may well have drawn upon their prior experience when making a single (Experiment 1) or a second time prediction (Experiments 4 and 9).

The immediacy with which prior task experience was acquired in the present research might thus explain why distributional information was presumably considered and used to good effect. That is, as the temporal interval between the acquisition of task experience and time estimation was only a matter of minutes, participants may well have recalled information about the just-completed task when making a prediction. For example, when estimating the duration of the three-disk task in Experiment 1, some participants had just completed the task at least once during the two-minute pre-exposure period. Hence, greater time prediction accuracy may have been a consequence of these participants recalling pertinent information such as the correct disk movement pattern, which they used as a basis for their temporal estimate.

In contrast, in many real world situations, the temporal interval between repeated performance of similar or identical tasks tends to be much longer than was the case in the present studies. For example, marking students' examination scripts is a task that is usually performed at the end of each academic year, meaning that there is a period of around nine months until this activity is undertaken again. This kind of lengthy temporal interval would be likely to impede the recall of distributional information such as the duration of previous examination script marking tasks. Thus, in such instances, the person can only really estimate their completion time using information about the current task such as the number of examination scripts to be marked. On the basis of previous research (e.g., Buehler et al., 1994), time predictions would be expected to be over-optimistic in such instances because of the person's reliance on singular information (Kahneman & Tversky, 1979).

The temporal interval between previous and current tasks could thus explain the well established link between the planning fallacy and the failure to use distributional information. Similarly, the temporal interval between time estimation and task performance might explain the absence of a general optimistic time prediction bias in the present research. That is, the imminence of task performance may have forced participants to engage in defensive pessimism (Norem & Cantor, 1986b), which resulted in them erring on the side of caution when making a time estimate.

Indeed, it has been shown that non-temporal judgements are less optimistically biased when task performance is imminent rather than temporally distant (e.g., Sheppard et al., 1996; Lerner & Tetlock, 1999). Likewise, research into temporal construal theory (e.g., Liberman & Trope, 1998) has found that people consider different types of task-related information when judging their performance on tasks that are temporally distant or imminent. For example, factors that might delay task completion are overlooked when making judgements about future activities, whereas such information becomes the focus of attention when judging performance on tasks that are imminent (Trope & Liberman, 2000).

On the basis of such research, it could be that time estimates were pessimistically biased in the present studies because they were based on information such as potential impediments to optimal task completion. Moreover, given the presence of the researcher in the laboratory, participants were presumably aware that their task performance would be objectively evaluated. Thus, participants may have been over-cautious when estimating task duration in order to appear competent and maintain self-esteem (Gilovich et al., 1993).

However, as an optimistic time prediction bias has been observed when task performance occurred immediately after time estimation (e.g., Byram, 1997), it could be that a brief temporal interval does not lead to pessimistically biased temporal estimates on tasks that are of longer duration than the ones used here. That is, the tasks used in such previous research took more time to complete than those used in the present studies.

In research where a lengthy temporal interval was operative (e.g., Buehler et al., 1994), an optimistic time prediction bias would be expected to prevail because participants' task performance was not assessed until some time after task duration was predicted. Indeed, as completion times were self-reported in some studies (e.g., Newby-Clark et al., 2000), task performance was not objectively evaluated, and so participants had no obvious need to maintain self-esteem. Given the findings of previous research (e.g., Savitsky et al., 1998), optimistically biased predictions should occur in such circumstances because attention is focused on information concerning successful task completion (i.e., singular information). In view of these suggestions, the temporal interval between time estimate and task performance might well influence the direction in which predictions of task duration are biased (i.e., under or overestimation).

The use of the anchoring and adjustment judgemental heuristics (Tversky & Kahneman, 1982a) could explain the effects of prior performance of a shorter version of the same task or a shorter similar task, which were observed in the present research. Support for this suggestion came from Experiments 8 and 9, which revealed that time predictions on longer tasks may have been based on the perceived duration of a just-completed shorter task. Thus, temporal underestimation might have been a consequence of using an anchoring and adjustment judgement strategy whereby the shorter duration of the previous task served as an anchor for time predictions, which were insufficiently adjusted according to the greater demands of the upcoming task.

Given the prevalence and robustness of the anchoring and adjustment heuristics (George et al., 2000), it is likely that temporal as well as non-temporal judgements would be affected by these cognitive heuristics. As the planning fallacy has been attributed to heuristic information processing (Kahneman & Tversky, 1979), temporal underestimation could be a consequence of using cognitive heuristics. In fact, Buehler et al. (1995) report a study in which previously-presented numerical values served as anchors for time predictions on a college coursework assignment.

Whilst the methodology reported by Buehler et al. is consistent with the standard anchoring paradigm (Chapman & Johnson, 2002), the present research suggests that anchoring and adjustment can occur in the absence of numerical values that are presented by the experimenter. That is, as no outcome feedback was made available to participants in the present studies, they presumably generated their own estimate of the duration of the just-completed task. Indeed, at the end of the first trial in all of the studies, a number of participants mentioned that they had misestimated the duration of the just-completed task. Although this evidence is purely anecdotal, it seems to suggest that participants may have made implicit retrospective estimates of task duration.

Self-generated anchor values have been found to lead to bias in non-temporal judgements of task performance (Epley & Gilovich, 2002; Kruger, 1999; Mussweiler & Strack, 1999), suggesting that externally-presented numerical values are not a prerequisite for anchoring effects to occur. Such research provides support for the notion that, when a shorter duration task had just been performed, time predictions on a longer task were anchored on the perceived duration of the previous task in Experiments 3 to 9. That is, participants in these studies estimated the duration of the just-completed task, and used this figure as a basis for their next time prediction, which was insufficiently scaled up according to the greater demands of the upcoming task. Engaging in this kind of heuristic information processing would be expected to result in temporal underestimation on a longer duration task when a shorter one had just been performed.

There was limited evidence of this type strategy being used when a longer duration task had just been completed (Experiments 3 to 9), prompting the suggestion that anchoring and adjustment in the context of time estimation could be uni-directional. That is, an anchoring and adjustment judgement strategy might only be used when a shorter duration task has just been performed. In fact, temporal underestimation rather than temporal overestimation was evident in previous research where anchoring effects occurred (Buehler et al., 1995). Given that anchoring and adjustment was not evident when all (e.g., Experiment 9) or half of the same task had just been completed (Experiment 7), it could be that having substantial prior task experience enables people to engage in more thorough information processing when judging task duration.

Additionally, there was evidence that possessing such extensive task-related knowledge resulted in greater time prediction accuracy, suggesting that distributional information was used to good effect. Conversely, when a shorter duration task had just been performed and prior task experience was minimal (e.g., Experiment 8) or absent (e.g., Experiment 9), there was evidence that time predictions were anchored on erroneous distributional information such as the perceived duration of a previous unrelated task. On the basis of previous research (e.g., Wilson et al., 1996), judgements of task performance would be expected to be influenced by the anchoring and adjustment heuristics in the absence of prior task experience.

The directional nature of time estimation bias was highlighted in the present research. That is, the direction in which time predictions were biased was found to differ according to the duration of the upcoming task. There was evidence of general temporal overestimation on tasks with a duration of less than one minute, whereas general temporal underestimation prevailed on tasks that took between about eight and 16 minutes to complete. The latter finding indicates that, in relation to well structured laboratory tasks at least, the planning fallacy tends only to be evident when task duration is greater than eight minutes. Indeed, the laboratory tasks used in previous research where an optimistic time prediction bias prevailed were of at least eight minutes' duration (Buehler et al., 1997), suggesting that such tasks may need to take this amount of time to complete before the planning fallacy occurs.

A possible explanation for temporal overestimation on the short duration tasks used in the present studies is that participants may have estimated their completion times using whole minutes or longer temporal units such as 10 minutes. By the same token, on the long tasks, an optimistic time prediction bias could have been due to participants using temporal units such as five minutes. For example, on the seven-disk Tower of Hanoi task (Experiment 9), temporal underestimation may have been a consequence of participants giving a prediction of five minutes when the task itself took about 16 minutes to complete. Consistent with this suggestion, the frequency distribution of time estimates on the seven-disk task revealed that almost half of the participants in the seven-disk task condition judged the duration of the first trial as being either five or 10 minutes.

Given that the estimation of everyday activities is often made using temporal units such as five and 10 minutes (Fraisse, 1984), the finding from Experiment 9 indicates that several participants used such units when judging the duration of a laboratory task that took about 16 minutes to complete. In a similar vein, there was evidence that a number of individuals estimated task duration using whole minutes rather than seconds in Experiments 1 to 9. This tendency was observed on tasks ranging in duration from less than 30 seconds (e.g., the three-disk Tower of Hanoi task) to about 16 minutes (i.e., the seven-disk Tower of Hanoi task). Hence, in all of the present studies, there was evidence of participants estimating task duration in whole minutes rather than seconds.

Although task duration has received little empirical treatment previously, it has been suggested that a closely related factor (i.e., task complexity) might influence the length of temporal estimates (Zakay, 1989). Since more complex tasks are generally of longer duration than simpler ones (Buckhalt & Oates, 2002), with the exception of the repetitive disk movement task (Experiment 5), the longer tasks used here were presumably more difficult than the shorter ones. Given that time estimates have been found to be shorter when task performance requires greater attentional resources (Zakay, 1989), the over-optimism that tended to occur on the longer duration tasks could be due to cognitive processing limitations. For example, participants in Experiment 9 may have underestimated the duration of the seven-disk Tower of Hanoi task because they could not mentally represent the 127 disk moves needed for optimal task completion during the pre-exposure period

It has been proposed that judgement overconfidence is a consequence of insufficient cognitive processing of task-related information (Sniezek et al., 1990). Hence, temporal underestimation could be due to people's inability to cognitively process information concerning the completion of an upcoming task. In fact, greater cognitive processing of task-related information has been found to reduce over-optimism in non-temporal judgements of task performance (e.g., Hoch, 1985; Koriat et al., 1980). On the basis of such research, it could be that time estimation accuracy improves as a consequence of engaging in more in-depth cognitive processing. Given the role of heuristic information processing in determining the planning fallacy (i.e., the neglect of distributional information), a sound theoretical basis exists for further investigation of this issue.

The absence of a general optimistic time prediction bias in the present research could be due to the fact that the present tasks were well structured. That is, since the Tower of Hanoi task, the repetitive disk movement task, and the Playmobil® tasks all comprised sequential components, participants might have been able to envisage the task completion process at the outset. Indeed, Lesgold (1988) found that it was possible to mentally represent the steps involved in the completion of well structured problem solving tasks. Hence, as the present tasks were well structured, participants may well have based their time predictions on information about the nature of the to-be-completed task. For example, in Experiments 1 to 5, participants might have estimated the amount of time needed to place one disk on a peg, and based their task duration predictions on such information. Similarly, as the steps involved in completing the Playmobil® tasks were clearly specified in the instruction booklets, it is likely that individuals would have recognised the task components when estimating the duration of these tasks.

In contrast, as less well structured tasks such as anagrams have been employed in some previous research (e.g., Buehler et al., 1997), it could be that participants in those studies based their time estimates on some kind of non-task information. For example, as tasks such as completing a college thesis (Buehler et al., 1994) were of considerable personal importance, participants might have given over-optimistic time predictions in order to bolster self-esteem (Norem & Cantor, 1986a). Likewise, as tasks with a deadline were employed in some studies (e.g., Pychyl et al., 2000), over-optimism may have been due to participants trying to enhance feelings of self-efficacy (Taylor & Brown, 1988) regarding their ability to complete such tasks on time. Since the temporal underestimation indicative of the planning fallacy has been observed on many real world tasks (Buehler et al., 1994), it may prove fruitful for future research to determine whether time prediction bias differs according to the type of task that is about to be performed.

6.4 Strengths and Limitations of the Present Research

A major strength of the present research is the experimental control of prior task experience. Given the key role of distributional information in determining the planning fallacy, it is surprising that many of the tasks employed in previous research were ones that people had little prior experience of performing. For example, it is unlikely that Buehler et al.'s (1994) participants would have possessed prior experience of completing a final year college dissertation. Likewise, as income tax forms tend to be completed annually, participants in Buehler et al.'s (1997) research may not have been able to recall information about their previous task performance because it was not readily available to them and thus could not be brought to mind (Tversky & Kahneman, 1973).

In the present studies however, prior task experience was acquired within a few minutes of performing a target trial. Moreover, by using different versions of the same task, it was possible to vary the amount of prior task experience that participants possessed. Taking such steps not only ensured that participants had some prior experience of an upcoming task when estimating its duration, but also meant that the amount and type of such distributional information was experimentally manipulated.

Another major strength of this research is that objective measures of task duration were obtained in all studies. In order to examine the accuracy of participants' time estimates, it was decided not to employ self-report measures of task duration, which have been used in much previous research (e.g., Koole & Van't Spijker, 2000; Newby-Clark et al., 2000). The principal reason for not using self-report measures was that participants might misremember their completion time or provide inaccurate information regarding the duration of the just-completed task. Indeed, it has been shown that retrospective self-reported judgements of task performance are often characterised by the inaccurate recall of information (Ross, 1989), suggesting that self-reported estimates of task duration might also be erroneous. As there was no verification of task duration in previous studies were self-report measures were used, the possibility of participants giving incorrect information cannot be ruled out. Given this possibility, objective measures of task duration should be obtained when conducting research into time estimation.

The nature of the tasks used here is a potential limitation of the present research. Whilst the Tower of Hanoi is undoubtedly an artificial task, it may well also be a novel one. Specifically, it seems likely that participants had little or no pre-experimental experience of performing this task. Conversely, the Playmobil® tasks comprised many of the characteristics of certain everyday activities, suggesting that participants might have possessed some prior experience of performing similar kinds of task. However, as the Playmobil® tasks were unlikely to have been performed on a regular basis, any prior task experience that participants had may have been obtained quite some time before the experiment. Thus, in addition to the Tower of Hanoi task, individuals might not have possessed any readily available distributional information (Kahneman & Tversky, 1979) concerning the Playmobil® tasks prior to their participation in the present research.

A further potential limitation of the present research is that laboratory tasks were used in the nine studies. That is, many factors that can influence task performance and time estimation in everyday life do not occur in a laboratory environment. Support for this suggestion comes from Buehler et al. (1997), who found that the extent of temporal underestimation was less on laboratory tasks compared to everyday activities. Specifically, these authors found that, relative to the real world tasks used in previous research (Buehler et al., 1994), the magnitude of temporal underestimation was less on an anagram task that they employed.

Buehler et al. (1997) explain this finding by suggesting that various factors that can impede or delay the completion of real world tasks are absent from laboratory settings. For example, the task of collating weekly sales figures could be interrupted by telephone enquiries from clients, which may well cause the person to complete this activity later than anticipated. Conversely, such distractions rarely occur in the laboratory and the person is thus better able to complete the task within the estimated time (Buehler et al., 1997). In a similar vein, the present tasks were discrete, that is, participants did not perform other activities simultaneously. Hence, in contrast to most real world situations where the planning and completion of multiple activities is prevalent (Smith, 1996), participants estimated the duration of a single task, which they then performed in isolation.

6.5 Recommendations for Future Research

Given that time predictions have been shown to be more accurate when people have prior experience of a task (Josephs & Hahn, 1995; see also Experiments 1, 4 and 9), the issue of expertise should be addressed in future research. Since experts are involved in performing many organisational tasks such as project management (Markham, 1997), the study of expertise has potential practical implications for everyday time estimation. For example, a greater understanding of the conditions under which civil engineers take account of their performance on previous similar tasks may be useful in developing strategies for reducing costs and overheads on large scale construction projects. Whilst Kahneman and Tversky (1979) suggest that experts tend to be over-optimistic when estimating task duration, research in this domain has yet to investigate expert-novice differences. However, expertise has been found to improve task performance in areas as diverse as sporting prowess (McPherson, 2000; Ste-Marie, 1999) and business acumen (Baron, 1998).

In the domain of specialist problem solving, the impact of expertise on task performance was investigated by Chi et al. (1982). These authors found that experts used their more extensive domain-specific knowledge to complete tasks more effectively (e.g., to produce faster completion times). Specifically, Chi et al. found that experts (i.e., physics professors) solved mechanical physics problems using abstract principles of physics derived from implicit knowledge, whereas novices (i.e., physics undergraduate students) used explicit information, which was presented in the problem statement. Chi et al. suggest that experts and novices use different information processing strategies, which are derived from the amount of knowledge accumulated in memory.

The effects of expertise on the accuracy of judgements of task performance were studied by Smith and Kida (1991). These authors found that, compared to novices (i.e., students), experts (i.e., auditors) relied less on cognitive heuristics such as representativeness when judging their performance on job-related tasks. Smith and Kida suggest that novices' judgements are more susceptible to cognitive heuristics because they often perform experimental tasks of which they have no experience. Given that the planning fallacy has been attributed to heuristic information processing, Smith and Kida's research provides an empirical basis for studying the role of expertise in relation to judgements of task duration. Moreover, given the present findings, expertise may well prove to be a fruitful topic for future research in this domain.

Task deadlines is another issue that could be addressed in future research. Whilst many real world tasks are accompanied by deadlines (e.g., college assignments), this issue has received little empirical treatment in relation to time estimation. One exception is the work of Buehler et al. (1994), which revealed that the extent of temporal underestimation was less among participants who were set a deadline for task completion. In a similar vein, the type of task deadline was investigated by Ariely and Wertenbroch (2002), who found that procrastination was eliminated when deadlines were self-set rather than externally-imposed. That is, student participants who set their own deadline for completing a college coursework assignment tended to finish the task within the amount of time that they predicted.

As many activities have self-set (e.g., dieting) or externally-imposed deadlines (e.g., college coursework assignments), research into this issue has potential implications for time estimation in everyday life. For example, the elimination of procrastination observed by Ariely and Wertenbroch could translate into a reduction in time prediction bias. Thus, temporal misestimation could feasibly be overcome by ensuring that individuals set their own deadlines for task completion. Given that financial penalties are associated with the late completion of a growing number of organisational projects (Grundy & Brown, 2002), future research could investigate whether disincentives for failing to meet a task deadline mediate the accuracy of people's time predictions.

Another potentially fruitful area for future research is the type of task upon which time estimates are based. Whilst previous research has employed various tasks ranging from essays (e.g., Josephs & Hahn, 1995) to self-assembly furniture (Byram, 1997), it is unclear whether time estimation accuracy differs according to the type of the task that is about to be performed. However, time predictions were found to be more accurate on the short duration Playmobil® task relative to the three-disk Tower of Hanoi task (Experiment 8), prompting the suggestion that greater task-related information was available to participants on the former task. Given this suggestion, the kind of information upon which temporal estimates are based might differ between tasks as diverse as essay writing and assembling flat-packed furniture. For example, the step-by-step instructions that accompany self-assembly furniture products provide people with task-related information (Byram, 1997), which may well form the basis of their time estimates.

In contrast, on less well structured activities like essays, information about the nature of the task completion process (e.g., the number of major task components) may not be available to people when estimating task duration. Thus, temporal estimates might be based on information that is extrinsic to the task at hand, such as the number of other college coursework assignments that must be completed within a certain period of time. In order to address the issue of task type, future research could employ a single task that can be completed either with or without the aid of a set of sequential instructions.

As tasks of varying levels of difficulty are performed in everyday life, the issue of task complexity might also prove to be a fruitful area for future research. Although there was some evidence of time predictions being less biased on more complex versions of the Tower of Hanoi task (e.g., Experiment 3), it is not known whether such findings are applicable to real world tasks. Given that the impact of task complexity on bias in non-temporal judgements of task performance is well supported empirically (e.g., Fischhoff & MacGregor, 1982; Pulford & Colman, 1997), it could be that this factor also influences the accuracy of temporal estimates. Indeed, the existence of the difficulty effect (Griffin & Tversky, 1992) provides a sound theoretical basis for studying the role of task complexity in relation to temporal judgements. By employing real world tasks that differ in cognitive complexity, future research should be well placed to provide information concerning the applicability of the difficulty effect to estimates of task duration.

Given the paucity of previous research into factors that influence time estimation bias, it is unsurprising that several recommendations for future work should follow from the present studies. Topics such as task complexity, task deadlines and expertise all have potentially important implications for time estimation in everyday life. Moreover, given the prevalence of the planning fallacy (Buehler et al., 2002), and the costs associated with temporal underestimation on many real world projects (Grundy & Brown, 2002), there are sound practical and theoretical reasons for conducting further research in this domain.

In order to gain further insight into the kind of task-related information that forms the basis of time estimates on the current task, verbal protocols should be obtained in future research. Using verbal protocol analyses should also make it possible to determine the components of an anchoring and adjustment time estimation strategy. For example, having participants verbalise their thoughts would help determine the type of task-related information that is responsible for insufficient upward adjustment of time estimates on longer duration and more complex tasks.

6.6 Practical Implications of the Present Research

Whilst the present research suggests that prior task experience leads to greater time prediction accuracy (Experiments 1, 4 and 9), it is not known whether this finding is applicable to more everyday kinds of task than the ones used here. If this finding were to generalise to tasks such as college coursework assignments, then the benefits of prior experience on time prediction accuracy could be highlighted in study skills courses taught within educational institutions. For example, encouraging students to draw upon their previous task experience when preparing for examinations and planning the completion of assignments might result in better academic performance. Given the changing face of higher education (e.g., the introduction of tuition fees), teaching such task planning skills might also result in a reduction in student attrition rates, thus benefiting universities as well as students. Hence, further research into the effects of prior task experience can potentially aid the development of strategies for improving the accuracy of time predictions in everyday life.

If the findings of Experiments 3 to 9 are applicable to more everyday kinds of task, then people's use of the anchoring and adjustment judgemental heuristics could explain why the planning fallacy is evident on many large scale projects. That is, people who perform large scale projects will typically have prior experience of undertaking similar but less complex tasks. Moreover, as large scale organisational tasks such as major construction projects tend to be undertaken only infrequently (Grundy & Brown, 2002), judgements of their duration can only really be based on the shorter duration of previous less complex tasks. If anchoring on the shorter duration of previous less complex tasks occurs, then such time estimates would be expected to be optimistically biased.

Extrapolating from the present research, organisations might benefit from training employees to recognise (and avoid using) the cognitive heuristics that contaminate judgement and decision making processes. For example, emphasising the prevalence of the anchoring and adjustment heuristics might lead to less biased time predictions and the attenuation or elimination of the planning fallacy. Whilst the somewhat artificial nature of the present tasks renders this suggestion rather speculative, the fact that many everyday tasks are also well structured provides support for the potentially important practical implications of the present research. Moreover, given that serious consequences are often associated with temporal underestimation (e.g., project overrun costs), raising awareness of the adverse effects of using cognitive heuristics may well be beneficial in terms of improved judgement accuracy.

6.7 Theoretical Implications of the Present Research

Kahneman and Tversky's (1979) theory of the planning fallacy states that temporal underestimation is a consequence of attention being focused on singular information at the expense of distributional information, which is neglected. In order to identify why the planning fallacy occurs, subsequent research has employed various potential debiasing techniques including the mode in which time estimates are made (Griffin & Buehler, 1999) and the generation of less optimistic mental scenarios of the task completion process (Newby-Clark et al., 2000). However, aside from the use of information about personal performance on previous tasks (e.g., Buehler et al., 1994), these debiasing techniques have not resulted in greater time prediction accuracy (Buehler et al., 2002).

The present research provides further evidence that the use of information about previous personal task performance is a successful technique for debiasing estimates of task duration. Specifically, time predictions were found to be more accurate when participants completed the same task beforehand, suggesting that such distributional information was used to good effect. However, the present findings indicate that, under certain conditions, people consider erroneous information (e.g., the perceived duration of a previous unrelated task), which they use when estimating task duration.

Unsurprisingly, the use of such incorrect information does not improve time prediction accuracy, but leads to temporal underestimation when a current task is of longer duration than a previous one. Thus, in contrast to Kahneman and Tversky's (1979) theory of the planning fallacy, the present research implies that the neglect of distributional information is not a prerequisite for temporal underestimation. Instead, an optimistic time prediction bias might be a consequence of engaging in an anchoring and adjustment judgement strategy whereby attention is focused on the perceived duration of a just-completed shorter task.

Given the present findings, a possible alternative interpretation of the planning fallacy suggests itself. That is, an optimistic time prediction bias occurs because people use the anchoring and adjustment judgemental heuristics when estimating the duration of a longer duration or more complex task. Hence, rather than ignoring distributional information, individuals take account of their previous performance on shorter duration or simpler tasks, but engage in heuristic information processing when making time predictions.

Whilst it is for future research to determine whether the present findings apply to other types of task, temporal underestimation is evident when time predictions are based on information concerning personal performance on a previous task. In the light of the present research, some modification of Kahneman and Tversky's (1979) theory of the planning fallacy may be necessary. Specifically, it is the amount of prior task experience that people possess rather than their use of distributional information per se, which determines the extent to which their predictions of task duration are biased.

By experimentally controlling prior task experience, the present research has provided a more thorough treatment of the issue of distributional information than has been the case in previous studies (e.g., Buehler et al., 1994; Byram, 1997). In addition to emphasising the benefits of prior experience of the same task, the consequences of using different types of task-related information have been demonstrated. For example, basing time predictions on irrelevant information about just-completed shorter duration tasks has been shown to lead to temporal underestimation on longer tasks. Conversely, when a substantial part of a current task has just been performed, distributional information can be used to improve time prediction accuracy.

Since the neglect of base rate data is a key determinant of over-optimistic judgements of task performance (Lichtenstein et al., 1982), an in-depth analysis of the issue of distributional information is likely to enhance our understanding of the planning fallacy. However, given the prevalence of temporal underestimation both inside and outside of the laboratory, it is surprising that the components of the planning fallacy are somewhat ill-specified. That is, the nature of singular and distributional information is poorly understood. Although there has been considerable empirical evidence of the planning fallacy (e.g., Newby-Clark et al., 2000), the present research indicates that temporal underestimation is not just a matter of neglecting distributional information, but may also be a consequence of using the wrong kind of information about personal performance on previous tasks.

In the light of the present research, it could be that no single explanation of the planning fallacy exists, and that contextual factors such as the amount and type of prior task experience are responsible for the occurrence of optimistically biased time predictions. Whilst it is for future research to determine the exact components of this cognitive judgement phenomenon, the present studies have gone some way to identifying the conditions under which people use information about their previous task performance to good effect (i.e., to improve time prediction accuracy).

As it has been shown that temporal misestimation is reduced when people have prior experience of performing a longer duration but unrelated task (i.e., the seven-disk Tower of Hanoi task), the role of singular information in determining the planning fallacy may need to be reconsidered in the light of the present research. If the findings of Experiment 9 generalise to tasks where temporal underestimation prevails, then a further amendment to Kahneman and Tversky's (1979) theory of the planning fallacy suggests itself. Specifically, using singular information does not necessarily lead to an optimistic time prediction bias, but can improve judgement accuracy when information about the task at hand is acquired immediately before a temporal estimate is given and task performance occurs.

Alternatively, singular information might only be used to good effect when people possess little or no knowledge of the duration of the previous task, that is, when a more complex task has just been performed. In such instances, people may think about the amount of work involved in completing the current task, and base their time estimates on such singular information. On the basis of previous research (e.g., Sniezek et al., 1990), such cognitive processing of task-related information would be expected to result in more accurate predictions of task performance. Whatever the potential modifications to Kahneman and Tversky's theory of the planning fallacy, it is clear that this cognitive judgement phenomenon is not merely a consequence of attention being focused on case-specific information at the expense of distributional information, which is neglected.

6.8 Conclusion

The present research programme has perhaps raised many more questions than it has answered: Will the same findings occur with other tasks? Does the temporal interval between the performance of previous and current tasks mediate time prediction bias? Is the planning fallacy a consequence of using an anchoring and adjustment temporal judgement strategy?

Whilst it is for future research to answer these questions, the present studies provide compelling evidence that the planning fallacy is not as pervasive a cognitive judgement phenomenon as previously thought. Hence, the sources of its absence and indeed reversal on laboratory (and perhaps, real world) tasks are in need of further study. Moreover, the potentially important role of task duration rather than cognitive complexity in explaining the general overestimation of time on very short duration tasks is a topic worthy of further investigation.

This research programme has also demonstrated the importance of prior task experience in mediating time estimation bias, thus emphasising the need to understand the way in which such distributional information can be used to improve temporal prediction accuracy. Given the costs of misestimating the duration of many major projects and the potential gains of accurately estimating task completion times, there is an obvious need to more fully understand these effects.

Although the present studies have gone some way to enhancing our understanding of factors that can influence time prediction accuracy, the fact that this thesis was completed almost two months later than was initially anticipated provides an apposite example of temporal underestimation on complex and long duration tasks. Given the present findings, a plausible explanation for this over-optimism is that, having only undertaken similar but less complex tasks previously (e.g., writing journal articles), the author had little prior knowledge of the amount of work involved in writing a doctoral thesis, and thus possessed insufficient information about the nature of this task when estimating its duration.

Appendix 1: Test Statistics From Experiment 1

Analysis of Variance (ANOVA)

	Sum of Squares	df	Mean Square	F	Sig.
Time	9.360	1	9.360	38.367	.000
Time* Task Experience	2.631	2	1.316	5.392	.007
Error (Time)	13.906	57	.244		
Task Experience	13.355	2	6.667	13.106	.000
Error (Task Experience)	29.041	57	.509		

Levene Tests

	F	df 1	df 2	Sig.
Log Time Prediction	1.417	2	57	.251
Log Completion Time	4.067	2	57	.022

Appendix 1 (Continued):

Scheffé Tests

Task Experience Conditions		Mean Difference	Sig.
No Experience	Mental Planning	.802	.000
	Active Practice	.537	.006
Mental Planning	No Experience	-.802	.000
	Active Practice	-.265	.260
Active Practice	No Experience	-.537	.006
	Mental Planning	.265	.260

LSD t-tests (Following Time*Experience Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
No Experience Condition	.275	.611	1.109	19	.302
Mental Planning Condition	.967	.857	5.049	19	.000
Active Practice Condition	.434	.597	3.252	19	.004

Appendix 2: Test Statistics From Experiment 2

Analysis of Variance (ANOVA) on the Data From the First Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	1.584	1	1.584	5.265	.025
Time*Experience	2.225	1	2.225	7.395	.008
Time*Task	1.662	1	1.662	5.525	.022
Time*Task*Experience	1.271	1	1.271	4.223	.041
Error (Time)	18.652	62	.301		
Experience	6.770	1	6.770	10.477	.002
Task	62.229	1	62.229	96.308	.000
Task*Experience	.060	1	.060	.093	.762
Error (Task*Experience)	40.061	62	.646		

Levene Tests

	F	df 1	df 2	Sig.
Log Time Prediction	.228	3	62	.876
Log Completion Time	.943	3	62	.425

Appendix 2 (Continued):

LSD t-tests (Following Task*Time Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task	4.402	.682	3.704	32	.001
4-disk Task	.007	.936	.936	32	.968

LSD t-tests (Following Task*Time*Experience Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task (Mental Planning Condition)	.2158	.689	1.739	16	.103
4-disk Task (Mental Planning)	-.2642	.913	-1.554	16	.157
3-disk Task (No Experience Condition)	.6372	.671	3.408	15	.004
4-disk Task (No Experience Condition)	.4362	.808	2.324	15	.031

Appendix 2 (Continued):

Analysis of Variance (ANOVA) on the Data From the Second Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	4.015	1	4.015	12.631	.001
Time*Experience	.199	1	.199	.627	.431
Time*Task	8.079	1	8.079	25.414	.000
Time*Task*Experience	.401	1	.401	1.261	.266
Error (Time)	19.710	62	.318		
Experience	5.431	1	5.431	15.316	.000
Task	34.031	1	34.031	95.972	.000
Task*Experience	1.114	1	1.114	3.141	.081
Error (Task*Experience)	21.985	62	.355		

Levene Tests

	F	df 1	df 2	Sig.
Log Time Prediction	.970	3	62	.413
Log Completion Time	7.237	3	62	.000

Appendix 2 (Continued):

LSD t-tests (Following Task*Time Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task	.850	.748	6.529	32	.000
4-disk Task	-.145	.843	-.989	32	.330

Appendix 3: Test Statistics From Experiment 3

Analysis of Variance (ANOVA) on the Data From the First Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	.687	1	.687	2.111	.150
Time*Incentive	.011	1	.011	.149	.861
Time*Task	.584	1	.584	1.795	.184
Time*Task*Incentive	1.073	2	.573	1.649	.198
Error (Time)	30.272	93	.326		
Incentive	2.322	2	1.161	1.649	.198
Task	115.447	1	115.447	163.943	.000
Task*Incentive	.045	2	.023	.032	.968
Error (Task*Incentive)	65.489	93	.704		

Levene Tests

	F	df 1	df 2	Sig.
Log Time Prediction	.713	5	93	.615
Log Completion Time	5.305	5	93	.000

Appendix 3 (Continued):

Analysis of Variance (ANOVA) on the Data From the Second Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	.031	1	.031	.129	.720
Time*Incentive	.573	2	.286	1.186	.310
Time*Task	13.136	1	13.136	54.408	.000
Time*Task* Incentive	.636	2	.318	1.317	.273
Error (Time)	22.454	93	.241		
Incentive	.457	2	.228	.429	.653
Task	92.782	1	92.782	174.165	.000
Task* Incentive	1.426	2	.713	1.338	.267
Error (Task*Incentive)	49.543	93	.533		

Levene Tests

	F	df 1	df 2	Sig.
Log of Time Prediction	.667	5	93	.642
Log of Completion Time	6.964	5	93	.000

Appendix 3 (Continued):

LSD t-tests (Following Task*Time Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task	-.575	.612	-6.576	48	.033
4-disk Task	.494	.832	4.244	49	.000

Analysis of Variance (ANOVA) on the Completion Time Data

	Sum of Squares	df	Mean Square	F	Sig.
Task	156.179	1	156.179	723.953	.000
Task*Incentive	.134	2	.007	.310	.735
Error (Task)	20.710	96	.216		
Incentive	2.545	2	1.272	3.041	.052
Error (Incentive)	22.454	93	.418		

Appendix 3 (Continued):

Levene Tests

	F	df 1	df 2	Sig.
Log Completion Time (3-disk Task)	.224	2	96	.800
Log Completion Time (4-disk Task)	.983	2	96	.378

Appendix 4: Test Statistics From Experiment 4

Analysis of Variance (ANOVA) on the Data From the First Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	8.903	1	8.903	27.886	.000
Time*Task	12.134	1	12.134	38.004	.000
Error (Time)	27.374	92	.319		
Task	120.897	1	120.897	226.697	.000
Error (Task)	49.063	92	.533		

Levene Tests

	F	df 1	df 2	Sig.
Log Time Prediction	.663	1	92	.418
Log Completion Time	1.889	1	92	.173

Appendix 4 Continued):

LSD t-tests (Following Time*Task Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task	.943	.795	8.069	46	.000
4-disk Task	-.007	.802	-.630	46	.532

Analysis of Variance (ANOVA) on the Data From the Second Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	6.042	1	6.042	20.920	.000
Time*Experience	12.134	1	12.134	25.204	.000
Error (Time)	25.993	92	.289		
Experience	148.269	3	49.423	120.643	.000
Error (Experience)	36.870	92	.410		

Appendix 4 Continued):

Levene Tests

	F	df 1	df 2	Sig.
Log Time Prediction	1.102	3	90	.353
Log Completion Time	.925	3	90	.432

Scheffé Tests

Task Experience Conditions		Mean Difference	Sig.
3-disk Task Twice	5- Then 3-disk Task	-.4751	.006
	5-disk Task Twice	-2.144	.000
	3- Then 5-disk Task	-1.782	.000
5- Then 3-disk Task	3-disk Task Twice	.457	.006
	5-disk Task Twice	-1.669	.000
	3- Then 5-disk Task	-1.306	.000
5-disk Task Twice	3-disk Task Twice	2.144	.000
	5- Then3-disk Task	1.669	.000
	3- Then 5-disk Task	.363	.068
3- Then 5-disk Task	3-disk Task Twice	1.782	.000
	5- Then 3-disk Task	1.310	.000
	5-disk Task Twice	-.363	.068

Appendix 4 Continued):

LSD t-tests (Following Time*Experience Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task Twice Condition	.434	.793	2.683	23	.013
5- Then 3-disk Task Condition	1.417	.678	10.250	23	.000
5-disk Task Twice Condition	-.001	.784	-.053	22	.959
3- Then 5-disk Task Condition	-.409	.781	-2.509	22	.020

Appendix 5: Test Statistics From Experiment 5

Analysis of Variance (ANOVA) on the Data From the First Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	2.333	1	2.333	21.391	.000
Time*Task	1.530	1	1.530	14.033	.000
Error (Time)	5.235	48	.109		
Task	75.884	1	75.884	501.718	.000
Error (Task)	7.260	48	.151		

Levene Tests

	F	df 1	df 2	Sig.
Log Time Prediction	.492	1	48	.486
Log Completion Time	.412	1	48	.524

Appendix 5 (Continued):

LSD t-tests (Following Time*Task Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task	.553	.494	5.593	24	.000
5-disk Task	.004	.442	.479	24	.637

Analysis of Variance (ANOVA) on the Data From the Second Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	1.969	1	1.969	16.597	.000
Time*Task	5.795	1	5.795	48.838	.000
Error (Time)	5.682	48	.119		
Task	65.624	1	65.624	424.125	.000
Error (Task)	7.4327	48	.155		

Appendix 5 (Continued):

Levene Tests

	F	df 1	df 2	Sig.
Log Time Prediction	.764	1	48	.386
Log Completion Time	1.306	1	48	.259

LSD t-tests (Following Time*Task Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task	.762	.517	7.368	24	.000
5-disk Task	-.201	.454	-2.212	24	.037

Independent-Groups t-test on Time Predictions (Log Transformed) From the 3-disk Task

	Mean	Standard Deviation	t	df	Sig.
First Trial	3.960	.521	.239	48	.812
Second Trial	3.995	.534			

Appendix 6: Test Statistics From Experiment 6

Analysis of Variance (ANOVA)

	Sum of Squares	df	Mean Square	F	Sig.
Time	38.164	1	38.164	625.010	.000
Time*Order	.139	1	.139	2.280	.136
Error (Time)	3.603	59	.061		
Task	.166	1	.166	1.353	.249
Task*Order	.362	1	.362	2.947	.091
Error (Task)	7.237	59	.123		
Time*Task	3.731	1	3.731	86.654	.000
Time*Task*Order	.197	1	.197	4.575	.037
Error (Task*Time)	2.540	59	.043		
Order	1.217	1	1.217	3.228	.077
Error (Order)	22.245	59	.377		

Appendix 6 (Continued):

Levene Tests

	F	df 1	df 2	Sig.
Log Prediction (Short Task)	5.291	1	59	.025
Log Completion Time (Short Task)	.016	1	59	.900
Log Prediction (Long Task)	.341	1	59	.561
Log Completion Time (Long Task)	.013	1	59	.910

Appendix 6 (Continued):

LSD t-tests (Following Time*Task Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
Short Task	.193	.457	3.295	60	.002
Long Task	.300	.368	-6.366	60	.000

LSD t-tests (Following Time*Task*Order Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
Short Task First	.061	.509	.671	30	.508
Short Task Second	.329	.356	5.058	29	.000
Long Task First	-.279	.404	-5.321	30	.000
Long Task Second	-.320	.356	-3.786	29	.001

Appendix 7: Test Statistics From Experiment 7

Analysis of Variance (ANOVA) on the Data From the Second Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	.284	1	.284	2.517	.117
Time*Experience	2.076	3	.692	6.122	.001
Error (Time)	8.591	76	.113		
Experience	269.354	1	89.785	637.662	.000
Error (Experience)	10.701	76	.141		

Levene Tests

	F	df 1	df 2	Sig.
Log Prediction	6.413	3	76	.001
Log Completion Time	1.025	3	76	.386

Appendix 7 (Continued):

Scheffé Tests

Task Experience Conditions		Mean Difference	Sig.
Short Then Long	Long Then Short	3.126	.000
	Medium Then Long	-.091	.758
	Long Then Medium	.826	.000
Long Then Short	Short Then Long	-3.126	.000
	Medium Then Long	-3.217	.000
	Long Then Medium	-2.230	.000
Medium Then Long	Short Then Long	.091	.758
	Long Then Short	3.217	.000
	Long Then Medium	.917	.000
Long Then Medium	Short Then Long	-.826	.000
	Long Then Short	2.230	.000
	Medium Then Long	-.917	.000

Appendix 7 (Continued):

LSD t-tests (Following Time*Experience Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
Short Task	.321	.651	1.590	19	.128
Medium Task	.294	.339	3.886	19	.001
Long Task (Short First)	-.292	.529	-2.466	19	.023
Long Task (Medium First)	.104	.293	1.583	19	.130

Appendix 7 (Continued):

Independent-Groups t-tests on Time Predictions (Log Transformed) From the Short and Medium Tasks

	Mean	Standard Deviation	t	df	Sig.
Short Task First	3.470	.628	1.674	38	.102
Short Task Second	3.132	.650			
Medium Task First	5.793	.407	2.612	38	.013
Medium Task Second	5.463	.392			

Analysis of Variance (ANOVA) on the Data From the First Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	1.212	1	1.212	13.786	.000
Time*Experience	.453	3	.151	1.674	.216
Error (Time)	6.683	76	.088		
Experience	271.886	3	90.629	464.122	.000
Error (Experience)	14.840	76	.195		

Appendix 7 (Continued):

Levene Tests

	F	df 1	df 2	Sig.
Log Prediction	2.526	3	76	.064
Log Completion Time	1.001	3	76	.397

Scheffé Tests

Task Experience Conditions		Mean Difference	Sig.
Short Then Long	Long Then Short	-3.205	.000
	Medium Then Long	-2.456	.000
	Long Then Medium	-3.140	.000
Long Then Short	Short Then Long	3.205	.000
	Medium Then Long	.749	.000
	Long Then Medium	.065	.933
Medium Then Long	Short Then Long	2.456	.000
	Long Then Short	-.749	.000
	Long Then Medium	-.684	.000
Long Then Medium	Short Then Long	3.140	.000
	Long Then Short	-.065	.933
	Medium Then Long	.684	.000

Appendix 8: Test Statistics From Experiment 8

Analysis of Variance (ANOVA) on the Data From the Long Playmobil® Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	4.532	1	4.532	61.814	.000
Time*Experience	.022	1	.022	.302	.585
Error (Time)	3.959	54	.073		
Experience	.000	1	.000	.003	.958
Error (Experience)	5.965	54	.110		

Levene Tests

	F	df 1	df 2	Sig.
Log Prediction	.061	1	54	.806
Log Completion Time	.191	1	54	.664

Appendix 8 (Continued):

Analysis of Variance (ANOVA) on the Data From the Short Tasks

	Sum of Squares	df	Mean Square	F	Sig.
Time	18.064	1	18.064	45.731	.000
Time*Task	7.746	1	7.746	19.609	.000
Error (Time)	21.331	54	.395		
Task	10.039	1	10.039	15.556	.000
Error (Task)	34.851	54	.645		

Levene Tests

	F	df 1	df 2	Sig.
Log Prediction	.663	1	54	.419
Log Completion Time	15.562	1	54	.000

Appendix 8 (Continued):

LSD t-tests (Following Task*Time Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
Short Playmobil® Task	.277	.914	1.605	27	.120
3-disk Tower of Hanoi Task	1.329	.863	8.153	27	.001

Appendix 9: Test Statistics From Experiment 9

Analysis of Variance (ANOVA) on the Data From the Second Task

	Sum of Squares	df	Mean Square	F	Sig.
Time	.623	1	.623	9.918	.002
Time*Experience	1.020	3	.340	5.412	.002
Error (Time)	5.404	86	.063		
Experience	2.861	3	.954	6.489	.001
Error (Experience)	12.640	86	.147		

Levene Tests

	F	df 1	df 2	Sig.
Log Prediction	1.268	3	86	.290
Log Completion Time	5.611	3	86	.001

Appendix 9 (Continued):

Scheffé Tests

Task Experience Conditions		Mean Difference	Sig.
3-disk Task	6-disk Task	.064	.886
	7-disk Task	-.085	.776
	Playmobil® Task	.235	.044
6-disk Task	3-disk Task	.064	.886
	7-disk Task	-.021	.996
	3-disk Task	.299	.005
7-disk Task	Short Then Long	.085	.776
	6-disk Task	.021	.996
	Playmobil® Task	.320	.003
Playmobil® Task	3-disk Task	-.235	.044
	6-disk Task	-.299	.005
	7-disk Task	-.320	.003

Appendix 9 (Continued):

LSD t-tests (Following Time*Experience Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task First Condition	-.362	.410	-4.237	22	.000
6-disk Task First Condition	-.120	.238	-2.417	22	.024
7-disk Task First Condition	.012	.409	.133	21	.895
Playmobil® Task Twice Condition	-.001	.335	-.008	21	.993

Appendix 9 (Continued):

Analysis of Variance (ANOVA) on the Data From the First Tasks

	Sum of Squares	df	Mean Square	F	Sig.
Time	.347	1	.347	2.705	.104
Time*Task	15.623	3	5.208	40.541	.000
Error (Time)	11.047	86	.128		
Task	216.925	3	72.308	164.596	.000
Error (Task)	37.780	86	.439		

Levene Tests

	F	df 1	df 2	Sig.
Log Prediction	10.434	3	86	.000
Log Completion Time	15.072	3	86	.000

Appendix 9 (Continued):

Scheffé Tests

Task Experience Conditions		Mean Difference	Sig.
3-disk Task	6-disk Task	-2.210	.000
	7-disk Task	-2.650	.000
	Playmobil® Task	-2.610	.000
6-disk Task	3-disk Task	2.210	.000
	7-disk Task	-.440	.024
	3-disk Task	-.399	.050
7-disk Task	Short Then Long	2.650	.000
	6-disk Task	.440	.024
	Playmobil® Task	.041	.994
Playmobil® Task	3-disk Task	2.610	.000
	6-disk Task	.399	.050
	7-disk Task	.041	.994

Appendix 9 (Continued):

LSD t-tests (Following Time*Task Interaction)

Predicted vs. Actual Time (Log Data)	Mean	Standard Deviation	t	df	Sig.
3-disk Task	1.030	.529	9.333	22	.000
6-disk Task	-.174	.577	-1.443	22	.163
7-disk Task	-.561	.511	-5.145	21	.000
Playmobil® Task	-.060	.385	.690	21	.498

Appendix 10: Copy of Published Journal Article

Applied Cognitive Psychology, 17, 6, 655-673 (2003)

Exploring the Time Prediction Process: The Effects of Task Experience and Complexity on Prediction Accuracy

Kevin E. Thomas, Stephen E. Newstead, and Simon J. Handley

University of Plymouth, Devon. UK.

Abstract

Whilst considerable research shows that people tend to underestimate their task completion times, there is little research concerning factors that mediate the time prediction process. In Experiments 1 to 3 a simple, well structured task, the 3-disk Tower of Hanoi, showed no evidence of underestimation; in fact, participants consistently overestimated the duration of this task. However, predictions were more accurate among participants who acquired some task experience beforehand. Task complexity was also found to be an important factor since the more cognitively complex 4- and 5-disk versions produced less biased predictions. Using a cognitively undemanding disk movement task, we found a general temporal overestimation in Experiment 4, thus suggesting that task duration might be responsible for the general lack of underestimation in the present studies. These results have implications for the planning of tasks in everyday life, and also suggest conditions under which time prediction accuracy can be improved.

Introduction

At the time of writing, one of the present authors is having a house extension built. The initial builder went bankrupt with the project half completed as a result of having underestimated the time a previous very large job would take to complete. The current builder is now four weeks over the scheduled completion date with no sign of things coming to an end. This scenario is a commonplace example of the planning fallacy, which is the tendency to underestimate how long a task will take to complete (Kahneman & Tversky, 1979). There is no shortage of other examples of temporal underestimation, ranging from students failing to complete college assignments on schedule to massive delays (and overspend) on government projects such as the Strategic Defence Initiative and the Channel Tunnel.

The consequences of such underestimation can be enormous. At a personal level, the ability to effectively plan daily activities and make accurate predictions about task performance is of considerable importance in terms of the enhancement of feelings of self-efficacy and professional and personal achievement (Taylor & Brown, 1988). In the workplace, failure to complete a project as planned might lead to loss of rewards (e.g., bonus payments) or even job losses and business closure. At governmental level, there may be electoral and financial implications of delays. Thus, there are sound practical reasons for investigating the time prediction process: Such research not only provides important information concerning why so many tasks are completed later than anticipated, but can also aid in the development of strategies for improving the accuracy of time predictions.

The present research focuses on several potentially important factors that might mediate the accuracy of people's time predictions. These factors are the cognitive complexity, duration and prior experience of tasks that are well structured (i.e., comprise sequential components), and on which objective measures of completion time can be obtained (e.g., performed in a laboratory environment).

The planning fallacy was identified by Kahneman and Tversky (1979), who found that experts made over-optimistic time predictions despite being aware that previous similar tasks had taken longer than anticipated. They suggest that two types of data are available to task planners: Information that is specific to the task at hand (i.e., singular information); and information about previous similar tasks (i.e., distributional information). Kahneman and Tversky propose that, when planning a task, people tend to focus attention on factors concerning the current task (e.g., amount of work involved) rather than on information about why previous tasks were completed later than predicted (e.g., unexpected setbacks). Thus people tend to ignore distributional information (e.g., performance on previous tasks) and concentrate instead on singular information (e.g., unique aspects of the current task), which leads to the construction of mental scenarios about how task performance will proceed smoothly.

The process of predicting task duration has been the focus of much research (e.g., Buehler, Griffin, & Ross, 1994; 1995; Byram, 1997; Koole & Van't Spijker, 2000; Newby-Clark, Ross, Buehler, Koehler, & Griffin, 2000). In these studies participants have predicted their completion time on various tasks ranging from assembling furniture (Byram, 1997) to submitting a college thesis (Buehler *et al.*, 1994). Findings have consistently revealed that predictions are optimistically biased, leading to the conclusion that the planning fallacy is a robust cognitive bias applicable to both short (Buehler, Griffin, & MacDonald, 1997; Byram, 1997) and long duration tasks (Buehler *et al.*, 1994). However, much previous research has not directly addressed the issue of distributional information, a major component of the planning fallacy. Specifically, much research (e.g., Buehler *et al.*, 1994) has employed tasks that participants had no previous experience of performing (e.g., completing a college thesis), meaning that participants possessed no distributional information when predicting task duration.

To explore the impact of distributional information on prediction accuracy it seems prudent for research to employ tasks on which participants have (or can acquire) some experience. This issue is addressed in the present research where participants performed different versions of the same task, or experienced a task before making a prediction. Such measures helped ensure that participants had some kind of information about their performance on previous similar tasks when they predicted their completion time.

Various techniques aimed at debiasing people's time predictions have been examined previously, but have not in general been successful in improving prediction accuracy. For example, Byram (1997) found that task experience did not reduce prediction bias, as participants who viewed task instructions twice before making a prediction were no less optimistic than participants who read the instructions once. Byram employed other debiasing techniques such as encouraging participants to divide a furniture assembly task into discrete portions and predict their completion time for each part, and encouraging the listing of potential pitfalls. Neither of these techniques resulted in less biased predictions.

Byram suggests that people may inaccurately predict task duration because of their use of the availability heuristic (Tversky & Kahneman, 1982). There is empirical support for this assertion as it has been demonstrated that predictions tend to be based on information about how a task will proceed without impediments (Buehler *et al.*, 1994). Thus it seems feasible that people attend to such information because it is available to them, and use it as a basis for making a time prediction. Byram also suggests that the use of the anchoring and adjustment heuristics (Tversky & Kahneman, 1982) might result in a failure to recognise possible setbacks because predictions are anchored on plans of success and adjusted accordingly.

Buehler, MacDonald, and Griffin (1994, as cited in Buehler *et al.*, 1995) found that participants anchored on a pre-determined initial prediction and failed to adjust their own time estimate accordingly. In the present studies, the impact of the anchoring and adjustment heuristics on the time prediction process is explored by manipulating the order in which participants perform different versions of the same task. For example, performing a longer duration task after a shorter one might result in over-optimism if the duration of the shorter task serves as an anchor for predictions concerning the longer task.

Although there is compelling empirical evidence that people tend to be over-optimistic when predicting task duration (e.g., Byram, 1997), several factors that might mediate the accuracy of time predictions have not been fully investigated. The present research aims to explore several of these factors. Perhaps the most important factors concern the effects of the nature of the task, including its cognitive complexity and duration. Other important factors might concern people's previous experience of a task. These factors will be considered in turn.

Various factors concerning the task itself have been studied, but seem to have little effect in reducing the extent to which time predictions are biased. Previous research has used both real life tasks such as writing an essay or completing a tax form, and artificial laboratory tasks such as solving anagrams (e.g., Buehler *et al.*, 1997). As we have seen, over-optimism seems to be prevalent in both these contexts. Likewise, the planning fallacy appears to be present on tasks ranging in duration from a few minutes (Byram, 1997, Study 1) to several weeks (e.g., Buehler *et al.*, 1994, Study 1). However, no attempt has been made to directly assess the extent of time prediction bias as a function of task duration.

Similarly, the impact of the cognitive complexity of a task on time predictions has not been investigated previously.

Interestingly, considerable research (e.g., Fischhoff & MacGregor, 1982; Suantak, Bolger, & Ferrell, 1996) has found that people are more overconfident when judging their performance on complex (rather than easy) general knowledge test questions, a phenomenon known as the difficulty effect (Griffin & Tversky, 1992). Perhaps over-optimism is more pronounced on cognitively complex tasks because people are unable to mentally represent the components necessary for successful completion when making a time prediction. Whatever the reasons for any potential relationship between time prediction bias and the cognitive complexity of a task, the existence of the difficulty effect emphasises the need to investigate this issue further.

Many of the tasks used in previous research do not seem to lend themselves to the study of the time prediction process. That is, many of the tasks employed do not consist of sequential components, thus making it difficult for participants to construct a plan before making a prediction (Hayes-Roth, 1980). Whilst some research has employed tasks with well defined components that can be performed sequentially (e.g., Byram, 1997, Study 1), most other studies have used tasks that are less well structured. For example, Buehler *et al.* (1997, Study 2) used an anagram task, but research by Kelley and Jacoby (1996) suggests that such tasks are ill structured and solved using the intuitive mode of cognition associated with insight problems. Thus it seems unlikely that participants in Buehler *et al.'s* study would perceive the components necessary for successful completion when predicting task duration. If, as seems feasible with certain tasks, people cannot construct a plan, it is not clear what information they use as a basis for predicting their completion time.

It is therefore important for research to use tasks with well defined components to investigate whether over-optimistic predictions prevail on such tasks. The need to employ well structured tasks is further emphasised by the fact that people perform them in life (e.g., cooking a meal from a recipe). Thus, investigating factors that might mediate the accuracy of predictions on well structured tasks has potentially important implications for improving the calibration of people's everyday temporal judgements. For example, if prior task experience is found to improve the accuracy of predictions on subsequent tasks then such findings could be used to help task planners use distributional information to reduce prediction bias (e.g., recognising the elements that are common to both previous and current tasks). However, if research is to accurately measure the extent of time prediction bias, the need to obtain objective measures of task duration is also paramount. Thus it seems prudent to employ a well structured task that is performed in a laboratory environment where objective measures of duration can be obtained.

A classic example of a well structured laboratory task, which has also been the subject of research concerning planning processes (e.g., Davies, 2000; Kotovsky, Hayes, & Simon, 1985), is the Tower of Hanoi. This task comprises different-sized disks that are moved across three vertical pegs to complete a specified pattern using only permitted moves. On the basis of previous research (e.g., Byram, 1997) one would expect that predictions on this task would be optimistically biased, but this is not yet known. Indeed, Lesgold (1988) found that, on well defined problems (i.e., ones with clearly defined initial and end states), people were able to form a mental representation of the problem structure and the actions needed for successful completion. Thus we might expect predictions on the Tower of Hanoi to be more accurate as people have a mental representation of what the task entails, whereas on ill structured tasks predictions might be more biased as people cannot construct such a representation. A key aim of Experiment 1 was to investigate the direction (i.e., optimistic or pessimistic) and extent of prediction bias on the Tower of Hanoi task.

The impact of task experience on the time prediction process has also been explored. Byram (1997) found no evidence that task experience reduced over-optimism (or improved prediction accuracy), but the manipulation used (i.e., briefly viewing the task instructions on two separate occasions) may not have been adequate to induce familiarity with the origami task. Research into other judgement and decision making processes supports the notion that experience (or familiarity) might be an important factor in improving time prediction accuracy.

Research by King, Zechmeister, and Shaughnessy (1980) found that predictions concerning the recall of memory test items were more accurate as task experience increased. Likewise, Smith and Kida (1991) report that experts' (auditors') judgements were fairly accurate and less biased than those of novices (students). Smith and Kida attribute this finding to the experts' greater familiarity with the job-related tasks they performed, and suggest that increased task experience yields judgements that are less contaminated by cognitive biases and heuristics. In relation to time predictions it seems feasible that task experience might equate to distributional information, which could form the basis of temporal estimates that are less biased. Thus task experience would appear to be a potentially important determinant of time prediction bias and was investigated in Experiment 1.

In Experiment 1 task experience was manipulated by permitting some participants to mentally plan or actively practise the Tower of Hanoi task before making a prediction and performing a target trial. Consistent with Smith and Kida's findings it was hypothesised that task experience would be inversely related to prediction bias. Thus we anticipate that predictions will be more accurate among participants who have some immediate experience of the task.

Experiment 1

This study examined the accuracy of time predictions on a well structured laboratory task, the 3-disk Tower of Hanoi. Task experience was investigated by using three different types of experience with the task before completion times were predicted.

Method

Participants. Sixty (53 female and 7 male) psychology undergraduates at the University of Plymouth participated in partial fulfilment of a course requirement.

Materials. The Tower of Hanoi task consists of three wooden disks of different sizes (with a hole in the middle of each one) and a flat wooden board containing three equidistantly-spaced vertical pegs of equal length. In the starting position the three disks were stacked in descending order of size on the left-hand peg and the aim of the task was to transfer them to the same position on the right-hand peg by moving one disk at a time, but without placing a larger disk on top of a smaller disk.

Design and Procedure. A 2 (time: predictions vs. task completion times) x 3 (experience: active practice vs. mental planning vs. no experience) mixed factorial design was used. Participants were tested individually, and after being briefed about the experimental rationale, were asked to remove their watches and place them out of sight. The task instructions were presented, and the researcher demonstrated the permissible and nonpermissible disk moves. There then followed a two-minute period that was different for the three groups.

For the no experience group the Tower of Hanoi apparatus was placed out of sight and a word association task (Berkowitz & Troccoli, 1990) was presented. The mental planning group were told that they had a short period of time to think about how to complete the task, but were instructed not to move any disks during the planning period. The active practice group were informed that they had a short period of time to practise performing the task before the target trial. All participants in this group completed the task at least once during the practice period. Following this two-minute period, the Tower of Hanoi apparatus was placed in front of all participants and they were asked to predict how much time it would take to complete. Participants then performed the task. Task completion time was measured using a stopwatch. Each testing session lasted approximately 15 minutes.

Results

As a measure of prediction accuracy, prediction discrepancy scores were calculated for participants in all experiments. These scores were calculated by subtracting actual from predicted time, thus a positive discrepancy score reflected a temporal overestimation whereas a negative score reflected a temporal underestimation. Descriptive statistics can be seen in Table 1.

The mean discrepancy scores were all positive, suggesting that there was a general tendency to overestimate time. The standard deviations indicated considerable variability within the data, and Levene tests were significant ($p < .05$), suggesting a lack of homogeneity of variance. Histograms revealed that the prediction and completion time data were positively skewed. Thus, these data were subjected to a logarithmic transformation, which improved homogeneity of variance and normality of distributions.

Table 1

Predicted and Actual Completion Times, and Discrepancy Scores Per Task Experience Condition (in Seconds)

	Actual Time	Predicted Time	Discrepancy Scores
No Experience	21.70 (14.15) n = 20	61.65 (42.77) n = 20	39.95 (39.78) n = 20
Active Practice	14.80 (56.41) n = 20	23.90 (11.09) n = 20	9.10 (10.81) n = 20
Mental Planning	12.25 (5.48) n = 20	19.20 (14.78) n = 20	6.95 (12.02) n = 20

Note: The figures given here and in Tables 2 - 4 are means with SDs in parentheses.

A time (predicted vs. actual) x experience (active practice vs. mental planning vs. no experience) mixed design ANOVA produced a main effect of time, $F(1,57) = 38.37$, $MSE = .24$, $p < .001$, with predictions ($M = 34.92$ seconds) exceeding completion times ($M = 16.25$ seconds). There was also a main effect of experience, $F(1,57) = 13.11$, $MSE = .51$, $p < .001$, with overall time being higher for the no experience group. The interaction was also significant, $F(2,57) = 5.39$, $MSE = .24$, $p < .01$, and suggested that the difference between actual and predicted time was higher for the no experience group than for the other two groups. Planned comparisons (LSD) revealed a significant difference between predicted and actual times for the active practice and no experience groups (ps < .05). However, there was no significant difference between predicted and actual times for the mental planning group ($p > .10$). These findings suggest that having some experience of the task beforehand resulted in predictions that were less pessimistically biased (i.e., more accurate).

<u>Discussion</u>

The present results support the hypothesis that task experience is positively related to prediction accuracy, a finding consistent with Smith and Kida's (1991) research. Participants who mentally planned or actively practised the task before making a prediction overestimated their completion time to a lesser extent than participants who undertook an unrelated task beforehand. Presumably, participants in the mental planning and active practice groups had the opportunity to mentally represent the correct solution of the task during the pre-exposure period (Lesgold, 1988), which may have enabled them to make predictions that were less biased.

The pre-exposure period might have given these participants time to consider (and reject) potential setbacks that could impede optimal task performance such as incorrectly moving a disk. However, those participants without prior experience might have engaged in defensive pessimism in order to maintain self-esteem (Norem & Cantor, 1986). Indeed, the high variability within the prediction data of the no experience group might suggest that these participants provided 'off-the-cuff' guesses (Byram, 1997) about task duration.

There was no evidence of the planning fallacy in participants' time predictions on the 3-disk Tower of Hanoi task. In fact, there was a general tendency to overestimate time, a finding that contrasts with much previous research (e.g., Buehler *et al.*, 1994). A possible explanation for this finding would appear to be that a relatively simple task was used in the present study. The 3-disk Tower of Hanoi task can be correctly completed by making a minimum of only seven disk moves, and took on average less than 20 seconds to finish.

Thus it was not only a simple task, but was also of much shorter duration than the laboratory tasks used previously. For example, Buehler *et al.* (1997) used laboratory tasks that took in excess of five minutes to complete, whereas the duration of the tasks used by Byram (1997) ranged from 15 minutes to one hour. As such tasks took more time they presumably comprised more components and may have involved greater cognitive effort. Participants may have been overconfident (i.e., over-optimistic) on these previous tasks because of cognitive processing limitations.

Support for this assertion comes from Sniezek, Paese, and Switzer (1990), who found that overconfidence was reduced when participants were encouraged to consider all possible alternative answers to complex multiple choice questions, thus increasing the amount of information being cognitively processed. Sniezek *et al.* suggest that overconfidence is positively related to task complexity, and occurs because people are unable or unwilling to devote sufficient cognitive resources to predicting their performance on more complex test questions. Such research gives credence to the notion that the greater complexity of the tasks used in previous studies could be responsible for the general over-optimism that has been observed previously.

Experiment 2

This study further explored the relationship between task experience and prediction accuracy on the Tower of Hanoi task. Specifically, all participants performed the 4-disk task in addition to the 3-disk, which enabled us to investigate whether experience with a different version affects the accuracy of predictions on the second task. Moreover, by using the 3- and 4-disk versions we were able to manipulate the cognitive complexity of the task. The Tower of Hanoi is well suited to the manipulation of cognitive complexity as different versions vary in the number of subgoals, which are established whenever a disk cannot be placed in its final location or goal state (cf. Goel & Grafman, 1995).

Cognitive complexity also differs because the number of disks that must be moved varies between different versions of the task (Davies, 2000). That is, as the number of disks increases, people's ability to mentally represent all relevant disk moves is constrained by limited working memory capacity, thus increasing the cognitive complexity of the task. For example, optimal performance on the 3-disk task involves a minimum of seven disk moves whereas the 4-disk version requires a minimum of 15 moves. Thus, the minimum number of moves on the 3-disk task can feasibly be retained in working memory whereas this is unlikely to be the case with the 4-disk version.

Consistent with the difficulty effect (Griffin & Tversky, 1992), we predicted a relationship between cognitive complexity and the calibration of time predictions. That is, time predictions were expected to be less pessimistic (or more accurate) on the more complex 4-disk task. Task performance order was manipulated to examine whether the anchoring and adjustment heuristics (Tversky & Kahneman, 1982) are applicable to the time prediction process. It was anticipated that participants would anchor their prediction concerning the second task on the duration of the preceding task. Thus, predictions on the 4-disk task should be over-optimistic when preceded by the 3-disk, whereas predictions on the 3-disk should be pessimistic when preceded by the 4-disk.

Method

Participants. Thirty-three (29 female and 4 male) students and staff at the University of Plymouth participated either in partial fulfilment of a course requirement or for payment of £1.50 each.

Materials. These were the same as those used previously except that the Tower of Hanoi apparatus contained four different-sized disks, with the largest one being removed to form the three-disk task.

Design and Procedure. A 2 (time: predicted vs. actual) x 2 (task: 3-disk vs. 4-disk) x 2 (order: 3-disk first vs. 4-disk first) mixed factorial design was used. Time and task were repeated measures. Participants performed both versions of the task in a fully counterbalanced order. Task performance order was manipulated between groups with participants being randomly assigned to one of two groups. As the group who mentally planned how to complete the 3-disk task made the most accurate predictions in Experiment 1, this method of task pre-exposure was used for all participants.

Participants were tested individually, and the instructions and procedure were almost identical to those for the mental planning group in Experiment 1. The only difference was that two versions of the Tower of Hanoi task were presented. Each testing session lasted approximately 15 minutes.

Results

Descriptive statistics can be seen in Table 2. The discrepancy score means on the 3-disk task were positive, suggesting that time was overestimated. However, on the 4-disk task discrepancy scores differed according to task performance order, with time being overestimated when this task was performed first, but underestimated when it was performed second.

Table 2

Predicted and Actual Completion Times, and Discrepancy Scores Per Task and Task Performance Order Condition (in Seconds)

	Actual Time	Predicted Time	Discrepancy Scores
3-disk task first	14.84 (7.46) n = 17	21.24 (19.79) n = 17	6.35 (16.22) n = 17
3-disk task second	10.06 (2.98) n = 16	26.31 (26.56) n = 16	16.25 (26.17) n = 16
4-disk task first	70.25 (51.78) n = 16	109.25 (80.54) n = 16	39.00 (76.45) n = 16
4-disk task second	86.59 (45.10) n = 17	44.29 (26.27) n = 17	-42.29 (42.76) n = 17

Because the prediction and completion time data were skewed with high variability, they were subjected to a logarithmic transformation. A 2 (task) x 2 (order) x 2 (time) mixed design ANOVA produced a main effect of task, $F(1,31) = 225.77$, $MSE = .31$, $p < .001$, with overall time being higher on the 4-disk task. The time by order interaction was also significant, $F(1,31) = 15.01$, $MSE = .33$, $p < .01$. Planned comparisons (LSD) revealed a significant difference between predicted and actual times for the group who performed the 4-disk task first ($p < .05$), but not for the group who performed the 3-disk first ($p > .10$).

The task by time interaction was also significant, $F(1,31) = 10.99$, $MSE = .25$, $p < .01$, and revealed that the difference between predicted and actual times was greater on the 3-disk (Ms = 47.55 vs. 24.90 seconds) than on the 4-disk task (Ms = 153.54 vs. 156.84 seconds). Planned comparisons (LSD) revealed that the difference between predicted and actual times was significant on the 3-disk task only ($p < .05$). Thus it seems that predictions were more accurate on the more cognitively complex 4-disk task. Specifically, a significant temporal overestimation was evident on the 3-disk task, whereas the tendency to underestimate the duration of the 4-disk was not significant. The ANOVA produced no other significant interactions or main effects (Fs < 2.50, ps > .10).

To investigate the effect of task experience on time predictions we analysed the log-transformed data from the second task only. A 2 (time) x 2 (task) mixed design ANOVA produced no significant main effect of time or task (Fs < 2, ps > .10). However, the interaction approached significance, $F(1,31) = 3.15$, $MSE = .87$, $p < .09$. This suggested that predictions exceeded completion times on the 3-disk task (Ms = 26.31 vs. 10.06 seconds), whereas completion times exceeded predictions on the 4-disk (Ms = 86.59 vs. 44.29 seconds). Planned comparisons (LSD) revealed that the difference between predicted and actual times was significant on both tasks ($p < .05$). These findings provide some evidence of anchoring effects as predictions were over-optimistic on the 4-disk task when preceded by the 3-disk, but were pessimistically biased on the 3-disk when preceded by the 4-disk.

Discussion

Consistent with the findings of Experiment 1, there was evidence of a general tendency to overestimate the duration of the 3-disk task in the present study. This effect was independent of task performance order, and further supports the notion that time predictions on such simple tasks do not seem to be subject to the planning fallacy. There was also evidence that participants were more accurate at predicting the duration of the 4-disk task, a finding that is consistent with the expected positive relationship between prediction calibration and the cognitive complexity of the task. This finding also suggests that the difficulty effect (Griffin & Tversky, 1992) might apply to the process of estimating task completion times, as participants made less biased predictions on the more complex task.

There was no evidence that previously performing an easier or more complex version of the Tower of Hanoi differentially affected the accuracy of time predictions on the second task. However, previous experience seemed to influence the direction in which predictions were biased on the second task. Specifically, the planning fallacy was evident on the 4-disk task, whereas completion times were generally overestimated on the 3-disk. This finding suggests that participants used the anchoring and adjustment heuristics when estimating the duration of the second task. That is, participants seem to have used the duration of the first task as the basis for their predictions concerning the second task.

The present findings provide some support for the notion that the planning fallacy might be only be evident on tasks that are more cognitively complex than the 3-disk Tower of Hanoi. However, the order in which the more complex 4-disk version was performed seemed to affect whether or not participants made over-optimistic predictions about the duration of this task. The lack of general over-optimism on the 4-disk task could be explained by the fact that it still took less time to complete than the laboratory tasks used in previous research. Specifically, the 4-disk task took less than two minutes to complete, whereas the tasks used previously have taken at least five minutes (e.g., Buehler *et al.*, 1997). Thus the 4-disk task may have been too simple to induce overconfidence in participants, a notion that is consistent with the difficulty effect. A key aim of Experiment

3 was to investigate this issue by increasing the cognitive complexity of the Tower of Hanoi task.

Experiment 3

This experiment further explored the relationship between cognitive complexity and prediction accuracy by employing the 5-disk Tower of Hanoi task alongside the 3-disk version. Following the findings of Experiment 2, we anticipated that there would be a positive relationship between prediction accuracy and the cognitive complexity of the task. Thus the difference between predicted and actual completion times was expected to be reduced on the 5-disk task.

Task performance order was also manipulated, and we hypothesised that predictions on the second task would be anchored on the duration of the preceding task. Thus predictions should be over-optimistic on the 5-disk task when preceded by the 3-disk, whereas predictions should be more pessimistic on the 3-disk task when preceded by the 5-disk. Task experience was also manipulated, and involved some participants performing the same Tower of Hanoi task twice, whilst others performed both the 3- and 5-disk versions. It was predicted that experience with the same (as opposed to a different) version would improve prediction accuracy on the second task. Thus predictions on the 5-disk task should be more accurate when performed for a second time rather than when preceded by the 3-disk. Likewise, predictions should be more accurate on the 3-disk task when preceded by the 3-disk rather than the 5-disk.

Method

Participants.

Ninety-four (88 female and 6 male) psychology undergraduates at the University of Plymouth participated voluntarily in partial fulfilment of a course requirement.

Materials.

These were the same as those used previously except that the Tower of Hanoi apparatus contained five different-sized disks, with the two largest being removed to form the 3-disk task.

Design and Procedure.

A 2 (time: predicted vs. actual) x 2 (task: first vs. second) x 4 (experience: 3-disk twice vs. 5-disk twice vs. 3- then 5-disk vs. 5- then 3-disk) mixed factorial design was used. Time and task were repeated measures. The order in which tasks were performed was determined by random assignment to a level of the task experience factor. Task experience was manipulated between groups with participants being randomly assigned to one of four groups.

The experimental procedure was identical to that used in Experiment 2 except that participants had only 20 seconds (after reading the task instructions) to make a prediction on each task. Each testing session lasted approximately 20 minutes.

Results

Descriptive statistics can be seen in Table 3. The majority of discrepancy score means were positive, suggesting that there was a general tendency to overestimate time. However, there was evidence that participants used the anchoring and adjustment heuristics when predicting their completion time on the second task. The largest underestimation occurred on the 5-disk task when preceded by the 3-disk, whereas the largest overestimation occurred on the 3-disk task when preceded by the 5-disk.

Table 3

Predicted and Actual Completion Times, and Discrepancy Scores Per Task and Task Experience Condition (in Seconds)

	First Task			Second Task		
	Actual Time	Predicted Time	Discrepancy Scores	Actual Time	Predicted Time	Discrepancy Scores
3-disk task twice	32.71 (23.11) n = 24	93.54 (49.88) n = 24	60.83 (57.17) n = 24	19.17 (13.93) n = 24	30.63 (22.33) n = 24	11.46 (27.51) n = 24
5-disk task then 3-disk task	248.75 (135.14) n = 24	243.96 (169.01) n = 24	-4.97 (189.03) n = 24	18.00 (10.03) n = 24	82.50 (52.29) n = 24	64.50 (50.88) n = 24
5-disk task twice	247.22 (109.18) n = 23	268.70 (160.41) n = 23	21.48 (155.86) n = 23	203.87 (117.57) n = 23	204.57 (121.85) n = 23	0.70 (140.49) n = 23
3-disk task then 5-disk task	31.09 (19.66) n = 23	71.96 (46.65) n = 23	40.87 (44.14) n = 23	169.39 (85.17) n = 23	120.22 (77.00) n = 23	-47.19 (11.28) n = 23

Because the prediction and completion time data were skewed with considerable variability they were subjected to a logarithmic transformation. The task experience hypothesis was tested by analysing the log-transformed data from the second task only. A 2 (time) x 4 (experience) mixed design ANOVA produced a main effect of time, $F(3,90) = 20.92$, $MSE = .29$, $p < .001$, with predictions ($M = 109.48$ seconds) exceeding completion times ($M = 102.61$ seconds). There was also a main effect of experience, $F(3,90) = 120.64$, $MSE = .41$, $p < .001$. LSD (t-test) pairwise comparisons revealed that overall time was significantly shorter for the two groups who performed the 3-disk task (ps $< .05$). The interaction was also significant, $F(3,90) = 25.20$, $MSE = .29$, $p < .001$. This revealed that, on the 5-disk task, there was evidence of underestimation among the 3- then 5-disk group only, whereas completion times tended to be overestimated by both groups who performed the 3-disk task.

Planned comparisons (LSD) on the 5-disk task data revealed a significant difference between predicted and actual times for the 3- then 5-disk group only ($p < .05$). However, on the 3-disk task, the difference between predicted and actual times was greater for the 5- then 3-disk group ($p < .01$) than for the group who performed the task twice ($p < .05$). Taken together, these findings are consistent with the task experience hypothesis, as participants who performed the same task twice tended to predict the duration of the second task more accurately than participants who undertook a different task initially. Likewise, these findings support the anchoring hypothesis, as predictions seem to be anchored on the duration of the preceding task. Specifically, predictions on the 3-disk were more pessimistic when this task was preceded by the 5-disk, whereas predictions on the 5-disk task were over-optimistic only when the 3-disk was performed first.

The cognitive complexity hypothesis was tested by analysing the log-transformed data from the first task only. These data were partially collapsed across task experience and performance order by combining the two groups who performed the 3- or 5-disk task. A 2 (time) x 2 (task) mixed design ANOVA produced a main effect of time, $F(3,90) = 27.87$, $MSE = .32$, $p < .001$, with predictions ($M = 169.54$ seconds) exceeding completion times ($M = 139.94$ seconds). There was also a main effect of task, $F(3,90) = 226.70$, $MSE = .53$, $p < .001$, with overall time being lower on the 3-disk. The interaction was also significant, $F(3,90) = 38.00$, $MSE = .32$, $p < .001$, and revealed that completion times were overestimated to a greater extent on the 3-disk task. Planned comparisons (LSD) revealed a significant difference between predicted and actual times on the 3-disk task ($p < .05$), but not on the 5-disk ($p > .10$). These findings suggest that cognitive complexity is positively related to prediction accuracy, as predictions were more accurate on the 5-disk task.

Discussion

Consistent with the task experience hypothesis, we found that predictions on the second task were more accurate when participants had previously performed the same (rather than a different) version of the Tower of Hanoi. If participants were able to mentally represent the task components whilst performing the first task, it seems likely that such a representation would form the basis of their second time prediction. Thus one would expect participants who performed the same task twice to be fairly accurate when predicting their completion time on the second task. Likewise, if some kind of mental representation was constructed, participants who performed different versions of the Tower of Hanoi should base their second prediction on the duration of the just-completed task.

There is empirical support for this assertion, as we found that the direction in which predictions were biased on the second task differed according to the cognitive complexity of the first task. This finding suggests that participants used the anchoring and adjustment heuristics when predicting the duration of the second task. Consistent with the anchoring hypothesis, we found that predictions on the 5-disk task were over-optimistic only when preceded by the 3-disk, whereas predictions were more pessimistic on the 3-disk task when preceded by the 5-disk rather than the 3-disk.

The present findings suggest that completion times were predicted with greater accuracy on the more complex 5-disk task. However, in relation to the first task, there was a general tendency to overestimate time, which was more pronounced on the 3-disk task. As was the case in Experiment 2 we found no evidence of general over-optimism on the more complex version of the Tower of Hanoi. In fact, the planning fallacy was only evident on the 5-disk task when it was preceded by the 3-disk.

Task duration could explain the disparity between the present findings and those of previous research where the planning fallacy was evident (e.g., Byram, 1997). Specifically, whilst the average duration of the 5-disk task was more than twice that of the 4-disk in Experiment 2 (Ms = 217.31 vs. 78.42 seconds respectively) it still took less time than the laboratory tasks used previously. Thus the 5-disk task might have been too short to induce feelings of overconfidence in participants. Likewise, if duration is a positive function of the cognitive complexity of the task, the absence of over-optimism could be due to participants' ability (or willingness) to devote sufficient cognitive resources to the process of predicting the duration of the 5-disk task (Sniezek *et al.*, 1990). Whatever the reasons there is further evidence that prediction accuracy is positively related to task complexity, a finding that suggests that the difficulty effect might be applicable to the time prediction process.

<u>Experiment 4</u>

Thus far, we have assumed that the difference between the 3-disk task and the 4-disk and 5-disk task is one of cognitive complexity. However, as the latter task also take more time to complete it is possible that duration alone can explain the present findings. This study investigated whether the duration or cognitive complexity of the task is responsible for the general tendency to overestimate time on well structured laboratory tasks such as the Tower of Hanoi. To achieve this goal we examined the accuracy of predictions on a repetitive manual disk movement task, which used the same apparatus as the 3- and 5-disk tasks. Unlike the Tower of Hanoi there were few restrictions on disk movement in the repetitive task, which is thus cognitively undemanding and simple to perform. Moreover, because there are few restrictions on disk movement (e.g., larger disks can be placed on top of smaller disks) there is no need for participants to create subgoals in order to complete either version of this task. Thus the two repetitive tasks do not vary in cognitive complexity but differ only in the amount of time needed for successful completion.

If cognitive complexity is an important factor in determining the time prediction process, participants might be over-optimistic about their ability to complete such a simple, repetitive task. Alternatively, if task duration is an important factor, then the same general temporal overestimation will be found with this task as was found with the Tower of Hanoi in Experiments 1 to 3. Task performance order was also manipulated to explore whether participants used the anchoring and adjustment heuristics when predicting the duration of the second task.

The repetitive task involved moving the disks one at a time from the left-hand peg to the centre peg to the right-hand peg and back again in reverse order. This process was repeated until all disks had been placed on the pegs a certain number of times. The two repetitive tasks were designed to be of similar duration to the 3- and 5-disk Tower of Hanoi tasks (when performed first) in Experiment 3 (Ms = 31.90 and 247.99 seconds respectively). These mean completion times served as a benchmark for the number of pegs on which the disks had to be placed. A pilot study indicated that eight pegs could be covered in 32 seconds on the 3-disk task, whereas 36 pegs could be covered in 248 seconds on the 5-disk. Thus these were the criteria for successful task completion.

Method

Participants.

Fifty (44 female and 6 male) psychology undergraduates at the University of Plymouth participated voluntarily in partial fulfilment of a course requirement.

Materials.

These were identical to those used in Experiment 3.

Design and Procedure.

A 2 (time: predicted vs. actual) x 2 (task: 3-disk vs. 5-disk) x 2 (order: 3-disk first vs. 5-disk first) mixed factorial design was used. Task and time were repeated measures, whereas task performance order was manipulated between groups. The experimental procedure was identical to that of Experiment 3 except that the researcher recorded the number of pegs on which disks were stacked, and told participants when each task was completed. Each testing session lasted approximately 20 minutes.

Results

Descriptive statistics can be seen in Table 4. The majority of discrepancy score means were positive, suggesting that there was a general tendency to overestimate time. This overestimation was greatest on the 3-disk task when preceded by the 5-disk, whereas over-optimism was only evident on the 5-disk task when preceded by the 3-disk. Thus it seems that participants used the anchoring and adjustment heuristics when predicting time on the second task. However, as there was no general evidence of the planning fallacy it seems that task duration (rather than cognitive complexity) is responsible for the general temporal overestimation on such laboratory tasks.

Because the prediction and completion time data were skewed with high variability they were subjected to a logarithmic transformation. To explore the issue of cognitive complexity versus task duration we analysed the log-transformed data from the first task only. A 2 (time) x 2 (task) mixed design ANOVA produced a main effect of time, $F(1,48) = 21.39$, $MSE = .11$, $p < .001$, with predictions ($M = 157.60$ seconds) exceeding completion times (M = 129.42 seconds). There was also a main effect of task, $F(1,48) = 501.72$, MSE = .16, $p < .001$, with overall time being higher on the 5-disk task. The interaction was also significant, $F(1,48) = 14.04$, MSE = .11, $p < .001$, and suggested that completion times were overestimated to a greater extent on the 3-disk task. Planned comparisons (LSD) revealed that predictions were significantly higher than completion times on the 3-disk task ($p < .05$), but not on the 5-disk ($p > .10$).

Table 4

Predicted and Actual Completion Times, and Discrepancy Scores Per Task and Task Performance Order Condition (in Seconds)

	Actual Time	Predicted Time	Discrepancy Scores
3-disk task first	30.48 (4.62) n = 25	59.60 (30.92) n = 25	29.12 (29.92) n = 25
3-disk task second	26.08 (6.72) n = 25	63.20 (41.05) n = 25	37.12 (40.61) n = 25
5-disk task first	228.36 (47.44) n = 25	255.60 (106.46) n = 25	27.24 (108.62) n = 25
5-disk task second	210.68 (38.89) n = 25	184.00 (74.89) n = 25	-26.68 (83.12) n = 25

To investigate the impact of task performance order on prediction accuracy we analysed the log-transformed data from the second task only. A 2 (time) x 2 (task) mixed design ANOVA produced a main effect of time, $F(1,48) = 16.60$, $MSE = .12$, $p < .001$, with predictions ($M = 123.60$ seconds) exceeding completion times ($M = 118.38$ seconds). There was also a main effect of task, $F(1,48) = 424.13$, $MSE = .16$, $p < .001$, with overall time being higher on the 5-disk task. The interaction was also significant, $F(1,48) = 48.84$, $MSE = .12$, $p < .001$, and suggested that predictions exceeded completion times on the 3-disk task, whereas predictions were lower than completion times on the 5-disk. Planned comparisons (LSD) revealed a significant difference between predicted and actual times on both tasks ($ps < .05$). These findings are consistent with the anchoring hypothesis, as predictions on the 5-disk task should be over-optimistic if based on the shorter duration of the just-completed 3-disk. Likewise, predictions on the 3-disk task should be pessimistically biased if anchored on the longer duration of the just-completed 5-disk.

Discussion

The present findings suggest that task duration rather than cognitive complexity is responsible for the general overestimation of time on such laboratory tasks. In general, participants tended to make pessimistically biased predictions about the duration of this repetitive disk movement task, which suggests that the absence of the planning fallacy here could be due to the shorter duration of the tasks used in the present studies.

These findings also indicate that participants used the anchoring and adjustment heuristics when predicting the duration of the second task. Specifically, predictions on the 3-disk task were pessimistically biased when preceded by the 5-disk, whereas predictions on the 5-disk task were over-optimistic when preceded by the 3-disk. The latter finding suggests that the planning fallacy was evident on the 5-disk task, but is mediated by the order in which this task is performed. Thus there is evidence that completion times were underestimated only when participants had just performed a shorter duration version of the repetitive disk movement task.

General Discussion

The present findings provide a remarkably consistent picture concerning the process of predicting the duration of the Tower of Hanoi task. It is clear that, on the 3-disk version, the planning fallacy does not occur and is in fact reversed. In Experiments 1, 2, and 3 (and indeed in every cell in every condition in these experiments) participants overestimated the amount of time that this task would take to complete. This simple finding is of considerable importance since the planning fallacy has previously been found in virtually all published studies and with a variety of different tasks. It indicates that the planning fallacy is not as prevalent as previously suspected, and that there is at least one task on which it definitely does not occur.

We suspect that what makes the 3-disk task different to those tasks that have been used in previous research is its simplicity. It can be performed in a few seconds, involves a minimum of only seven disk moves, and has a very well defined structure. Support for this notion comes from Experiments 2 and 3, which used the 4- and 5-disk versions of the Tower of Hanoi. Although these tasks are still well structured they involve more than twice (4-disk) or four times (5-disk) as many minimum disk moves as the 3-disk version. In Experiments 2 and 3 we found consistent evidence that the amount of temporal overestimation was significantly reduced on these more cognitively complex tasks. Thus it seems that the cognitive complexity of the task is an important factor in the time prediction process, and there is evidence that the difficulty effect (Griffin & Tversky, 1992) generalises to people's judgements of their task completion times.

There was also evidence that in general participants tended to overestimate the duration of all versions of the Tower of Hanoi task. Likewise, there was a tendency for participants to overestimate the duration of the 3- and 5-disk versions of the simple repetitive task used in Experiment 4. Taken together, these findings suggest that people tend to make pessimistically biased predictions on well structured tasks, which range in duration from less than 15 seconds to four minutes. However, there was a hint of the planning fallacy in Experiments 2, 3, and 4, but participants' over-optimism seemed to be mediated by the order in which they performed the 4- and 5-disk tasks. That is, participants tended to underestimate their completion times on these tasks only when they had performed either the 3-disk Tower of Hanoi or the 3-disk repetitive task beforehand.

Our experiments also demonstrate that task experience is an important determinant of the time prediction process. In fact, we have investigated two distinct types of task experience. Firstly, we examined the effects of task practice on prediction accuracy (Experiment 1), and found that two minutes mentally planning or actively practising the 3-disk Tower of Hanoi led to less temporal overestimation. In other words, having the opportunity to acquire some kind of experience of this task increased the accuracy of participants' time predictions.

Secondly, in Experiment 3 we examined the effects of previous experience with the same or a different version of the Tower of Hanoi, and found that predictions were more accurate on the second task when it was identical to the first task. It seems feasible that acquiring prior task experience enabled these participants to mentally represent what the task entailed, which resulted in them making a more realistic single (Experiment 1) or second prediction (Experiment 3). Similarly, in Experiments 2 and 3 we found that previous experience of an easier or more difficult version of the Tower of Hanoi affected the direction in which predictions were biased on the second task. This effect was also evident in Experiment 4, suggesting that the duration of the first repetitive task affected the direction of prediction bias on the second version of this task.

The latter findings suggest that participants used the anchoring and adjustment heuristics (Tversky & Kahneman, 1982) when predicting the duration of the second task. In relation to the time prediction process, evidence of anchoring effects implies that participants based their predictions about their task performance on what happened on the just-completed task. Thus those participants who performed the 3-disk Tower of Hanoi or the 3-disk repetitive task second tended to overestimate their completion times even more, and participants who performed the 4- and 5-disk tasks second tended to underestimate time. In fact, the only real evidence of the planning fallacy in the present studies occurred on the 4- and 5-disk tasks when they were preceded by the 3-disk tasks. Specifically, in Experiments 2, 3, and 4, prior performance of the 3-disk tasks led to an underestimation of how long the 4- and 5-disk tasks would take to complete.

If the present findings are applicable to more everyday tasks then people's use of the anchoring and adjustment heuristics could explain why the planning fallacy is evident on many large scale projects. That is, people who perform large scale projects will typically have previous experience of undertaking similar but less complex tasks. Moreover, as large scale projects (e.g., the Channel Tunnel) are inevitably undertaken only infrequently, estimates of their duration can only really be based on the duration of similar but less complex (or shorter duration) tasks. If anchoring on the duration of less complex tasks occurs, such estimates will inevitably err on the side of optimism.

Extrapolating from the present findings, organisations might thus benefit from training employees to overcome the various cognitive biases that contaminate the decision making process. For example, highlighting the prevalence of the anchoring and adjustment heuristics might lead to less biased predictions and the attenuation or elimination of the planning fallacy. Although the somewhat artificial nature of the present tasks makes this assertion rather speculative, the fact that many everyday tasks are also well structured provides support for the potentially important practical implications of the present findings.

Our experiments also seem to highlight the importance of distributional information in the time prediction process; that is, the effect of task experience on prediction accuracy. We found that predictions were more accurate when participants had some prior experience of the same task (Experiments 1 and 3), suggesting that these participants might have used distributional information (e.g., previous performance on the same task) to good effect. However, the immediacy with which task experience was acquired could explain why these participants made more accurate predictions. For example, when predicting the duration of the second task in Experiment 3, some participants had just completed the same task. Thus they might have recalled information about the just-completed task (i.e., distributional information), which they then used as a basis for their second time prediction. Support for this assertion comes from Buehler *et al.* (1994), who found that prediction bias was reduced when participants were encouraged to recognise the similarities between previous tasks and the task they were about to undertake.

The temporal distance between previous and current tasks could thus explain the well established link between the planning fallacy and the failure to use distributional information. Specifically, in many everyday situations the temporal distance between previous and current tasks is likely to be considerably longer than was the case in the present studies. For example, marking students' examination scripts is a task that many academics perform at the end of each semester, meaning that there is a period of at least five months until this activity is undertaken again. It seems feasible that a lengthy temporal distance is likely to impede the recall of distributional information (e.g., duration of previous examination script marking tasks). Thus in such instances the person can only really estimate their completion time using information about the current task (e.g., the number of examination scripts to be marked). On the basis of previous research (e.g., Buehler *et al.*, 1994) one would expect such predictions to be over-optimistic because of the person's reliance on singular information (Kahneman & Tversky, 1979).

The fact that the present tasks were well structured could explain the general absence of the planning fallacy in our experiments. Specifically, as these tasks comprised sequential components it seems feasible that participants could have based their time predictions on some kind of task-related information (e.g., the amount of time needed to cover one peg with three disks).

Consistent with this assertion we found evidence of a positive relationship between task experience and prediction accuracy, suggesting that participants might have used task-related information as the basis for estimating their completion times. However, as previous research has tended to use less well structured tasks it is possible that participants in previous studies had to base their time estimates on some kind of non-task information. For example, as several of the previous tasks were of considerable importance to participants (e.g., completing a college thesis), they might have given over-optimistic predictions in order to bolster self-esteem (Norem & Cantor, 1986). Likewise, as some studies used tasks with a deadline (e.g., Buehler *et al.*, 1994), over-optimism could be due to participants trying to enhance feelings of self-efficacy (Taylor & Brown, 1988) regarding their ability to complete such tasks on time.

The present experiments have perhaps raised more questions than they have answered: Will the same findings occur with other simple tasks? Is the planning fallacy only evident on longer duration tasks? And how cognitively complex must a task be in order to produce optimistically biased predictions? Nonetheless, we have provided evidence that the planning fallacy is not as pervasive as previously thought, and the sources of its absence (and indeed reversal) on simple laboratory tasks are thus in need of further study. Moreover, the potentially important role of task duration (rather than cognitive complexity) in explaining the present general overestimation of time is also worthy of further investigation. We have also highlighted the importance of task experience in mediating the extent to which people's time predictions are biased, thus emphasising the need to understand the way in which distributional information can be used to improve prediction accuracy. Given the cost of underestimating the duration of major projects and the potential gains of accurately predicting task completion times, we need to understand more fully these effects.

References

Berkowitz, L. & Troccoli, B. T. (1990). Feelings, direction of attention, and expressed evaluation of others. *Cognition & Emotion, 4*, 305-325.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67, 3*, 366-381.

Buehler, R., Griffin, D., & Ross, M. (1995). It's about time: Optimistic predictions in work and love. In W. Stroebe and M. Hewstone (eds.). *European Review of Social Psychology. Volume 6*, John Wiley & Sons Ltd: Oxford, (pp.1-32).

Buehler, R., Griffin, D. & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin, 23, 3*, 238-247.

Byram, S. J. (1997). Cognitive and motivational factors influencing time predictions. *Journal of Experimental Psychology: Applied, 3, 3*, 216-239.

Davies, S. P. (2000). Memory and planning processes in solutions to well-structured problems. *The Quarterly Journal of Experimental Psychology, 53A, 3*, 896-927.

Fischhoff, B. & MacGregor, D. (1982). Subjective confidence in forecasts. *Journal of Forecasting, 1*, 155-172.

Goel, V. & Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions?: Interpreting data from the Tower of Hanoi. *Neuropsychologica, 33, 5*, 623-642.

Griffin, D. & Tversky, A. (1992). The weighting of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411-435.

Hayes-Roth, B. (1980). *Human planning processes*. Report prepared for the Office of Naval Research, R-2670-ONR.

Kahneman, D. & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in the Management Sciences, 12,* 313-327.

Kelley, C. M. & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35,* 157-175.

King, J. F., Zechmeister, E. B. & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval. *American Journal of Psychology, 93, 2,* 329-343.

Koole, S. & Van't Spijker, M. (2000). Overcoming the planning fallacy through willpower: Effects of implementation intentions on actual and predicted task-completion times. *European Journal of Social Psychology, 30,* 873-888.

Kotovsky, K., Hayes, J. R. & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*, 248-294.

Lesgold, A. (1988). Problem solving. In R. J. Sternberg and E. E. Smith (eds.). *The Psychology of Human Thought,* Cambridge University Press: New York (pp.188-213).

Newby-Clark, I. R., Ross, M., Buehler, R., Koehler, D. J. & Griffin, D. (2000). People focus on optimistic scenarios and disregard pessimistic scenarios when predicting task completion times. *Journal of Experimental Psychology: Applied, 6, 3*, 171-182.

Norem, J. K., & Cantor, N. (1986). Anticipatory and post hoc cushioning strategies: Optimism and defensive pessimism in "risky" situations. *Cognitive Therapy and Research, 10, 3,* 347-362.

Smith, J. F., & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin,* 109, 472-489.

Sniezek, J. A., Paese, P. W. & Switzer, F. S. III. (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46,* 264-282.

Suantak, L., Bolger, F., & Ferrell, W. R. (1996). The hard-easy effect in subjective probability calibration. *Organizational Behavior and Human Decision Processes, 67, 2,* 201-221.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103, 2,* 193-210.

Tversky, A. & Kahneman, D. (1982). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic and A. Tversky (eds.). *Judgment Under Uncertainty: Heuristics and Biases,* Cambridge University Press: Cambridge (pp. 3-20).

Author Notes

This research was funded by a research grant from the Economic and Social Research Council of the United Kingdom (R42200034413) to Kevin Thomas.

Correspondence and requests for reprints of this article should be addressed to: Kevin Thomas, Department of Psychology, University of Plymouth, Drake Circus, Plymouth. Devon. UK. PL4 8AA. Electronic mail may be sent via the Internet to: k.e.thomas@plymouth.ac.uk.

References

Abernathy, B., Neal, R. J. & Koning, P. (1994). Visual-Perceptual and Cognitive Differences Between Expert, Intermediate, and Novice Snooker Players. *Applied Cognitive Psychology, 8,* 185-211.

Allan, L. G. (1979). The Perception of Time. *Perception & Psychophysics, 26, 5*, 340-354.

Ariely, D. & Loewenstein, G. F. (2000). When Does Duration Matter in Judgment and Decision Making? *Journal of Experimental Psychology: General, 129, 4*, 508-523.

Ariely, D., Loewenstein, G. F. & Kahneman, D. (2000). Comment on "When Does Duration Matter in Judgment and Decision Making?" (Ariely & Loewenstein, 2000). *Journal of Experimental Psychology: General, 129, 4*, 524-529.

Ariely, D. & Wertenbroch, K. (2002). Procrastination, Deadlines, and Performance: Self-Control by Precommitment. *Psychological Science, 13, 3*, 219-224.

Ariely, D. & Zakay, D. (2001). A Timely Account of Duration in Decision Making. *Acta Psychologica, 108, 2*, 187-207.

Arkes, H. R., Faust, D., Guilmette, T. J. & Hart, K. (1988). Eliminating the Hindsight Bias. *Journal of Applied Psychology, 73, 2*, 305-307.

Armor, D. A. & Taylor, S. E. (2002). When Predictions Fail: The Dilemma of Unrealistic Optimism. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*, (pp. 334-347). New York: Cambridge University Press.

References (Continued)

Aschoff, J. (1985). On the Perception of Time During Prolonged Temporal Isolation. *Human Neurobiology, 4,* 41-52.

Ashford, S. J. (1989). Self-Assessments in Organizations: A Literature Review and Integrative Model. In L. L. Cummings & B. M. Staw (Eds.). *Research in Organizational Behavior: An Annual Series of Analytical Essays and Critical Reviews. Volume 11,* (pp. 133-174). Geenwich, CT: JAI Press Inc.

Balzer, W. K., Doherty, M. E. & O'Connor, R., Jr. (1989). Effects of Cognitive Feedback on Performance. *Psychological Bulletin, 106, 3,* 410-433.

Baron, R. A. (1998). Cognitive Mechanisms in Entrepreneurship: Why and When Entrepreneurs Think Differently Than Other People. *Journal of Business Venturing, 13,* 275-294.

Bénabou, R. & Tirole, J. (2003). Self-Knowledge and Self-Regulation: An Economic Approach. In I. Brocas & J. D. Carrillo (Eds.). *The Psychology of Economic Decisions,* (pp. 137-167). Oxford: Oxford University Press.

Berkowitz, L. & Troccoli, B. T. (1990). Feelings, Direction of Attention, and Expressed Evaluation of Others. *Cognition & Emotion, 4,* 305-325.

Bernsden, M. & van der Pligt, J. (2001). Time is on My Side: Optimism in Intertemporal Choice. *Acta Psychologica, 108, 2,* 173-186.

Block, R. A. (1989). Experiencing and Remembering Time: Affordances, Context, and Cognition. In I. Levin & D. Zakay (Eds.). *Time and Human Cognition: A Lifespan Perspective,* (pp. 333-363). Amsterdam: North Holland.

References (Continued)

Block, R. A. & Zakay, D. (1997). Prospective and Retrospective Duration Judgments: A Meta-Analytic Review. *Psychonomic Bulletin and Review, 4, 2*, 184-197.

Bornstein, B. H. & Zickafoose, D. J. (1999). "I Know I Know it, I Know I Saw it": The Stability of the Confidence-Accuracy Relationship Across Domains. *Journal of Experimental Psychology: Applied, 5, 1*, 76-88.

Buehler, R., Griffin, D. & MacDonald, H. (1997). The Role of Motivated Reasoning in Optimistic Time Predictions. *Personality and Social Psychology Bulletin, 23, 3*, 238-247.

Buehler, R., Griffin, D. & Ross, M. (1994). Exploring the "Planning Fallacy": Why People Underestimate Their Task Completion Times. *Journal of Personality and Social Psychology, 67, 3*, 366-381.

Buehler, R., Griffin, D. & Ross, M. (1995). It's About Time: Optimistic Predictions in Work and Love. In W. Stroebe & M. Hewstone (Eds.). *European Review of Social Psychology. Volume 6*, (pp. 1-32). Oxford: John Wiley & Sons Ltd.

Buehler, R., Griffin, D. & Ross, M. (2002). Inside the Planning Fallacy: The Causes and Consequences of Optimistic Time Predictions. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*, (pp. 250-270). New York: Cambridge University Press.

Buckhalt, J. A. & Oates, D. F. (2002). Sensation Seeking and Performance on Divided Attention Tasks Varying in Cognitive Complexity. *Personality and Individual Differences, 32*, 67-78.

References (Continued)

Burt, C. D. B. (1993). The Effect of Actual Event Duration and Event Memory on the Reconstruction of Duration Information. *Applied Cognitive Psychology, 7,* 61-73.

Burt, C. D. B. & Kemp, S. (1994). Construction of Activity Duration and Time Management Potential. *Applied Cognitive Psychology, 8,* 155-168.

Busenitz, L. W. & Murphy, G. B. (1996). New Evidence in the Pursuit of Locating New Business. *Journal of Business Venturing, 11, 3,* 221-231.

Byram, S. J. (1997). Cognitive and Motivational Factors Influencing Time Predictions. *Journal of Experimental Psychology: Applied, 3, 3,* 216-239.

Byrne, R. (1977). Planning Meals: Problem-Solving on a Real Data-Base. *Cognition, 5,* 287-332.

Carlson, B. W. (1990). Anchoring and Adjustment in Judgments Under Risk. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 4,* 665-676.

Chapman, G. B. & Johnson, E. J. (1994). The Limits of Anchoring. *Journal of Behavioral Decision Making, 7,* 223-242.

Chapman, G. B. & Johnson, E. J. (1999). Anchoring, Activation, and the Construction of Values. *Organizational Behavior and Human Decision Processes, 79, 2,* 115-153.

Chapman, G. B. & Johnson, E. J. (2002). Incorporating the Irrelevant: Anchors in Judgments of Belief and Value. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment,* (pp. 120-138). New York: Cambridge University Press.

Chi, M. T. H., Glaser, R. & Rees, E. (1982). Expertise in Problem Solving. In R. J. Sternberg (Eds.). *Advances in the Psychology of Human Intelligence. Volume 1*, (pp. 7-75). Hillside, NJ: Lawrence Erlbaum Associates.

Cooper, A. C., Woo, C. Y. & Dunkelberg, W. C. (1988). Entrepreneurs' Perceived Chances for Success. *Journal of Business Venturing, 3*, 97-108.

Curton, E. D. & Lordahl, D. S. (1974). Effects of Attentional Focus and Arousal on Time Estimation. *Journal of Experimental Psychology, 103, 5*, 861-867.

Das, T. K. (1991). Time: The Hidden Dimension in Strategic Planning. *Long Range Planning, 24, 3*, 49-57.

Das, T. K. & Teng, B-S. (1999). Cognitive Biases and Strategic Decision Processes: An Integrative Perspective. *Journal of Management Studies, 36, 6*, 757-778.

Davies, S. P. (2000a). Memory and Planning Processes in Solutions to Well-Structured Problems. *The Quarterly Journal of Experimental Psychology, 53A, 3*, 896-927.

Davies, S. P. (2000b). Move Evaluation as a Predictor and Moderator of Success in Solutions to Well-Structured Problems. *The Quarterly Journal of Experimental Psychology, 53A, 4*, 1186-1201.

Dougherty, M. R. P., Gettys, C. F. & Thomas, R. P. (1997). The Role of Mental Simulation in Judgments of Likelihood. *Organizational Behavior and Human Decision Processes, 70, 2*, 135-148.

References (Continued)

Einhorn, H. J. & Hogarth, R. M. (1978). Confidence in Judgment: Persistence of the Illusion of Validity. *Psychological Review, 85, 5*, 395-416.

Elstein, A. S., Shulman, L. S. & Sprafka, S. A. (1990). Medical Problem Solving: A Ten-Year Perspective. *Evaluation and the Health Professions, 13, 1*, 5-36.

Epley, N. & Gilovich, T. (2002). Putting Adjustment Back in the Anchoring and Adjustment Heuristic. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*, (pp. 139-149). New York: Cambridge University Press.

Fischhoff, B. & MacGregor, D. (1982). Subjective Confidence in Forecasts. *Journal of Forecasting, 1,* 155-172.

Francis-Smythe, J. A. & Robertson, I. T. (1999). On the Relationship Between Time Management and Time Estimation. *British Journal of Psychology 90, 3,* 333-348.

George, J. F., Duffy, K. & Ahuja, M. (2000). Countering the Anchoring and Adjustment Bias With Decision Support Systems. *Decision Support Systems, 29,* 195-206.

Gilovich, T. & Griffin, D. (2002). Heuristics and Biases: Then and Now. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*, (pp. 1-18). New York: Cambridge University Press.

Gilovich, T., Griffin, D. & Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment.* New York: Cambridge University Press.

References (Continued)

Gilovich, T., Kerr, M. & Medvec, V. H. (1993). Effect of Temporal Perspective on Subjective Confidence. *Journal of Personality and Social Psychology, 64, 4*, 552-560.

Goel, V. & Grafman, J. (1995). Are the Frontal Lobes Implicated in "Planning" Functions? Interpreting Data From the Tower of Hanoi. *Neuropsychologica, 33, 5*, 623-642.

Goldin, S. E. & Hayes-Roth, B. (1980). *Individual Difference in Planning Processes*. Santa Monica, CA: Rand Corporation, Note N-1488 - ONR, June.

Griffin, D. & Buehler, R. (1999). Frequency, Probability, and Prediction: Easy Solutions to Cognitive Illusions? *Cognitive Psychology, 38*, 48-78.

Griffin, D. & Tversky, A. (1992). The Weighting of Evidence and the Determinants of Confidence. *Cognitive Psychology, 24*, 411-435.

Griffin, D., Dunning, D. & Ross, L. (1990). The Role of Construal Processes in Overconfident Predictions About the Self and Others. *Journal of Personality and Social Psychology, 59, 6*, 1128-1139.

Grundy, T. & Brown, L. (2002). *Strategic Project Management: Creating Organizational Breakthroughs*. London: Thomson Learning.

Harries, C., Evans, J. St. B. T., Dennis, I. & Dean, J. (1996). A Clinical Judgement Analysis of Prescribing Decisions in General Practice. *Le Travail Humain, 59*, 87-111.

Hawkins, M. F. & Tedford, W. H., Jr. (1976). Effects of Interest and Relatedness on Estimated Duration of Verbal Material. *Bulletin of the Psychonomic Society, 8, 4*, 301-302.

References (Continued)

Hayes-Roth, B. (1980). *Human Planning Processes*. Santa Monica, CA: Rand Corporation, Note R-2670 - ONR, December.

Hayes-Roth, B. & Hayes-Roth, F. (1979). A Cognitive Model of Planning. *Cognitive Science, 3*, 275-310.

Henry, R. A. (1994). The Effects of Choice and Incentives on the Overestimation of Future Performance. *Organizational Behavior and Human Decision Processes, 57*, 210-225.

Henry, R. A. & Sniezek, J. A. (1993). Situational Factors Affecting Judgments of Future Performance. *Organizational Behavior and Human Decision Processes, 54*, 104-132.

Hicks, R. E. & Brundige, R. M. (1974). Judgments of Temporal Duration While Processing Verbal and Physiognomic Stimuli. *Acta Psychologica, 38*, 447-453.

Hoch, S. J. (1984). Availability and Interference in Predictive Judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 4*, 649-662.

Hoch, S. J. (1985). Counterfactual Reasoning and Accuracy in Predicting Personal Events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 4*, 719-731.

Hoch, S. J. & Loewenstein, G. F. (1989). Outcome Feedback: Hindsight and Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 4*, 605-619.

Jacowitz, K. E. & Kahneman, D. (1995). Measures of Anchoring in Estimation Tasks. *Personality and Social Psychology Bulletin, 21, 11*, 1161-1166.

Johnson, D. S. & Kanfer, R. (1992). Goal-Performance Relations: The Effects of Initial Task Complexity and Task Practice. *Motivation and Emotion, 16, 2,* 117-141.

Josephs, R. A. & Hahn, E. D. (1995). Bias and Accuracy in Estimates of Task Duration. *Organizational Behavior and Human Decision Processes, 61, 2,* 202-213.

Juslin, P., Winman, A. & Olsson, H. (2000). Naïve Empiricism and Dogmatism in Confidence Research: A Critical Examination of the Hard-Easy Effect. *Psychological Review, 107, 2,* 384-396.

Kahneman, D. & Lovallo, D. (1993). Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking. *Management Science, 39, 1,* 17-31.

Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kahneman, D. & Tversky, A. (1973). On the Psychology of Prediction. *Psychological Review, 80, 4,* 237-251.

Kahneman, D. & Tversky, A. (1979). Intuitive Prediction: Biases and Corrective Procedures. *TIMS Studies in the Management Sciences, 12,* 313-327.

Kahneman, D. & Tversky, A. (1982a). Intuitive Prediction: Biases and Corrective Procedures. In D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment Under Uncertainty: Heuristics and Biases*, (pp. 414-421). Cambridge: Cambridge University Press.

References (Continued)

Kahneman, D. & Tversky, A. (1982b). The Simulation Heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment Under Uncertainty: Heuristics and Biases*, (pp. 201-208). New York: Cambridge University Press.

Karat, J. (1982). A Model of Problem Solving With Incomplete Constraint Knowledge. *Cognitive Psychology, 14,* 538-559.

Kawamura, S. (2000). Effect of Phonological Processing on Temporal Processing. *Japanese Psychological Research, 42, 3*, 178-182.

Kelley, C. M. & Jacoby, L. L. (1996). Adult Egocentrism: Subjective Experience Versus Analytic Bases for Judgment. *Journal of Memory and Language, 35,* 157-175.

Kidd, J. B. (1970). The Utilization of Subjective Probabilities in Production Planning. *Acta Psychologica, 34,* 338-347.

King, J. F., Zechmeister, E. B. & Shaughnessy, J. J. (1980). Judgments of Knowing: The Influence of Retrieval. *American Journal of Psychology, 93, 2*, 329-343.

Klayman, J. & Schoemaker, P. J. H. (1993). Thinking About the Future: A Cognitive Perspective. *Journal of Forecasting, 12,* 161-187.

Koehler, D. J. (1991). Explanation, Imagination, and Confidence in Judgment. *Psychological Bulletin, 110, 3*, 499-519.

Koole, S. & Van't Spijker, M. (2000). Overcoming the Planning Fallacy Through Willpower: Effects of Implementation Intentions on Actual and Predicted Task-Completion Times. *European Journal of Social Psychology, 30,* 873-888.

References (Continued)

Koriat, A., Lichtenstein, S. & Fischhoff, B. (1980). Reasons for Confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6, 2*, 107-118.

Kotovsky, K., Hayes, J. R. & Simon, H. A. (1985). Why are Some Problems Hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17,* 248-294.

Kruger, J. (1999). Lake Wobegon Be Gone! The "Below-Average-Effect" and the Egocentric Nature of Comparative Ability Judgments. *Journal of Personality and Social Psychology, 77, 2,* 221-232.

Kuhn, K. M. & Sniezek, J. A. (1996). Confidence and Uncertainty in Judgmental Forecasting: Differential Effects of Scenario Presentation. *Journal of Behavioral Decision Making, 9,* 231-247.

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin, 108, 3,* 480-498.

Lawrence, M. & O'Connor, M. (2000). Sales Forecasting Updates: How Good Are They in Practice? *International Journal of Forecasting, 16,* 369-382.

Lay, C. H. (1986). At Last, My Research Article on Procrastination. *Journal of Research in Personality, 20,* 474-495.

Lerner, J. S. & Tetlock, P. E. (1999). Accounting for the Effects of Accountability. *Journal of Personality and Social Psychology, 125, 2,* 255-275.

Lesgold, A. (1988). Problem Solving. In R. J. Sternberg & E. E. Smith (Eds.). *The Psychology of Human Thought,* (pp. 188-213). New York: Cambridge University Press.

References (Continued)

Lesgold, A., Glaser, R., Rubinson, H., Klopfer, D., Feltovich, P. & Wang, Y. (1988). Expertise in a Complex Skill: Diagnosing X-Ray Pictures. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.). *The Nature of Expertise*, (pp. 311-342). Hillsdale, NJ: Lawrence Erlbaum Associates.

Liberman, N. & Trope, Y. (1998). The Role of Feasibility and Desirability Considerations in Near and Distant Future Decisions: A Test of Temporal Construal Theory. *Journal of Personality and Social Psychology, 75, 1*, 5-18.

Lichtenstein, S., Fischhoff, B. & Phillips, L. D. (1982). Calibration of Probabilities: The State of the Art to 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment Under Uncertainty: Heuristics and Biases*, (pp. 306-344). Cambridge: Cambridge University Press.

Littlepage, G., Robison, W. & Reddington, K. (1997). Effects of Task Experience and Group Experience on Group Performance, Member Ability, and Recognition of Expertise. *Organizational Behavior and Human Decision Processes, 69, 2*, 133-147.

Loffler, G. (1998). Biases in Analyst Forecasts: Cognitive, Strategic or Second-Best? *International Journal of Forecasting, 14,* 261-275.

Macan, T. H., Shahani, C., Dipboye, R. L. & Phillips, A. P. (1990). College Students' Time Management: Correlations With Academic Performance and Stress. *Journal of Educational Psychology, 82, 4*, 760-768.

MacGregor, D. G. & Armstrong, J. S. (1994). Judgmental Decomposition: When Does it Work? *International Journal of Forecasting, 10,* 495-506.

References (Continued)

MacGregor, D. G., Lichtenstein, S. & Slovic, P. (1988). Structuring Knowledge Retrieval: An Analysis of Decomposed Quantitative Judgments. *Organizational Behavior and Human Decision Processes, 42,* 303-323.

Markham, C. (1997). *Practical Management Consultancy.* Third Edition. Milton Keynes: Accountancy Books.

Maynard, D. C. & Hakel, M. D. (1997). Effects of Objective and Subjective Task Complexity on Performance. *Human Performance, 10, 4,* 303-330.

McClain, L. (1983). Interval Estimation: Effect of Processing Demands on Prospective and Retrospective Reports. *Perception & Psychophysics, 34, 2,* 185-189.

McConchie, R. D. & Rutschmann, J. (1971). Human Time Estimation: On Differences Between Methods. *Perceptual and Motor Skills, 32,* 319-336.

McMackin, J. & Slovic, P. (2000). When Does Explicit Justification Impair Decision Making? *Applied Cognitive Psychology, 14, 6,* 527-541.

McPherson, S. L. (2000). Expert-Novice Differences in Planning Strategies During Collegiate Singles Tennis Competition. *Journal of Sport and Exercise Psychology, 22,* 39-62.

Michon, J. A. (1965). Studies on Subjective Duration: Subjective Time Measurement During Tasks With Different Information Content. *Acta Psychologica, 24,* 205-219.

Mumpower, J. L. & Stewart, T. R. (1996). Expert Judgement and Expert Disagreement. *Thinking and Reasoning, 2, 2/3,* 191-211.

Mussweiler, T. & Strack, F. (1999). Hypothesis-Consistent Testing and Semantic Priming in the Anchoring Paradigm: A Selective Accessibility Model. *Journal of Experimental Social Psychology, 35,* 136-164.

Mussweiler, T. & Strack, F. (2000). Numeric Judgments Under Uncertainty: The Role of Knowledge in Anchoring. *Journal of Experimental Social Psychology, 36,* 495-518.

Nembhard, D. A. (2000). The Effects of Task Complexity and Experience on Learning and Forgetting: A Field Study. *Human Factors, 42, 2*, 272-286.

Newby-Clark, I. R., Ross, M., Buehler, R., Koehler, D. J. & Griffin, D. (2000). People Focus on Optimistic Scenarios and Disregard Pessimistic Scenarios When Predicting Task Completion Times. *Journal of Experimental Psychology: Applied, 6, 3*, 171-182.

Nisbett, R. E. & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Engelwood Cliffs, NJ: Prentice-Hall, Inc.

Norem, J. K. & Cantor, N. (1986a). Anticipatory and Post Hoc Cushioning Strategies: Optimism and Defensive Pessimism in "Risky" Situations. *Cognitive Therapy and Research, 10, 3*, 347-362.

Norem, J. K. & Cantor, N. (1986b). Defensive Pessimism: Harnessing Anxiety as Motivation. *Journal of Personality and Social Psychology, 51, 6,* 1208-1217.

Northcraft, G. B. & Neale, M. A. (1987). Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions. *Organizational Behavior and Human Decision Processes, 39, 1*, 84-97.

References (Continued)

Osberg, T. M. & Shrauger, J. S. (1986). Self-Prediction: Exploring the Parameters of Accuracy. *Journal of Personality and Social Psychology, 51, 5*, 1044-1057.

Paese, P. W. & Sniezek, J. A. (1991). Influences on the Appropriateness of Confidence in Judgment: Practice, Effort, Information and Decision Making. *Organizational Behavior and Human Decision Processes, 48, 2*, 100-130.

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993). *The Adaptive Decision Maker*. New York: Cambridge University Press.

Peterson, D. K. & Pitz, G. F. (1988). Confidence, Uncertainty, and the Use of Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 1*, 85-92.

Phillips, L. H., Wynn, V. E., Gilhooly, K., Della Sala, S. & Logie, R. H. (1999). The Role of Memory in the Tower of London Task. *Memory, 7, 2*, 209-231.

Phillips, L. H., Wynn, V. E., McPherson, S. & Gilhooly, K. (2001). Mental Planning and the Tower of London Task. *The Quarterly Journal of Experimental Psychology, 54, 2*, 599-612.

Poynter, D. (1989). Judging the Duration of Time Intervals: A Process of Remembering Segments of Experience. In I. Levin & D. Zakay (Eds.). *Time and Human Cognition: A Lifespan Perspective*, (pp. 305-331). Amsterdam: North Holland.

Price, P. C. (2000). Wishful Thinking in the Prediction of Competitive Outcomes. *Thinking and Reasoning, 6, 2*, 161-172.

References (Continued)

Pulford, B. D. & Colman, A. M. (1997). Overconfidence: Feedback and Item Difficulty Effects. *Personality and Individual Differences, 23, 1*, 125-133.

Pychyl, T. A., Morin, R. W. & Salmon, B. R. (2000). Procrastination and the Planning Fallacy: An Examination of the Study Habits of University Students. *Journal of Social Behavior and Personality, 15, 5*, 135-150.

Radhakrishnan, P., Arrow, H. & Sniezek, J. A. (1996). Hoping, Performing, Learning and Predicting: Changes in the Accuracy of Self-Evaluations of Performance. *Human Performance, 9, 1*, 23-49.

Reeves, T. & Lockhart, R. S. (1993). Distributional Versus Singular Approaches to Probability and Errors in Probabilistic Reasoning. *Journal of Experimental Psychology: General, 122, 2*, 207-236.

Ross, M. (1989). Relation of Implicit Theories to the Construction of Personal Histories. *Psychological Review, 96, 2*, 341-357.

Ross, M. & Buehler, R. (2001). Identity Through Time: Constructing Personal Pasts and Futures. In A. Tesser & N. Schwarz (Eds.). *Blackwell Handbook of Social Psychology: Intraindividual Processes*, (pp. 518-544). Malden, MA: Blackwell Publishers.

Savitsky, K., Medvec, V. H., Charlton, A. E. & Gilovich, T. (1998). "What, Me Worry?": Arousal, Misattribution, and the Effect of Temporal Distance on Confidence. *Personality and Social Psychology Bulletin, 24, 5*, 529-536.

Schnaars, S. P. & Topol, M. T. (1987). The Use of Multiple Scenarios in Sales Forecasting: An Empirical Test. *International Journal of Forecasting, 3*, 405-419.

References (Continued)

Schoemaker, P. J. H. (1993). Multiple Scenario Development: Its Conceptual and Behavioral Foundation. *Strategic Management Journal, 14,* 193-213.

Schouwenburg, H. C. (1992). Procrastinators and Fear of Failure: An Exploration of Reasons for Procrastination. *European Journal of Personality, 6,* 225-236.

Schwarz, N. & Wyer, R. S. Jr. (1985). Effects of Rank Ordering Stimuli on Magnitude Ratings of These and Other Stimuli. *Journal of Experimental Social Psychology, 21,* 30-46.

Shahani, C., Weiner, R. & Streit, M. K. (1993). An Investigation of the Dispositional Nature of the Time Management Construct. *Anxiety, Stress, and Coping, 6,* 231-243.

Shanteau, J. (1992). How Much Information Does an Expert Use? Is it Relevant? *Acta Psychologica, 81,* 75-86.

Shanteau, J. & Stewart, T. R. (1992). Why Study Expert Decision Making? Some Historical Perspectives and Comments. *Organizational Behavior and Human Decision Processes, 53,* 95-106.

Sharp. G. L., Cutler, B. L. & Penrod, S. D. (1988). Performance Feedback Improves the Resolution of Confidence Judgments. *Organizational Behavior and Human Decision Processes, 42,* 271-283.

Sheppard, J. A., Ouellette, J. A. & Fernandez, J. K. (1996). Abandoning Unrealistic Optimism: Performance Estimates and the Temporal Proximity of Self-Relevant Feedback. *Journal of Personality and Social Psychology, 70, 4,* 844-855.

References (Continued)

Simon, H. A. (1975). The Functional Relevance of Problem Solving Skills, *Cognitive Psychology, 7*, 268-288.

Smith, J. (1996). Planning About Life: Towards a Social-Interactive Perspective. In P. B. Baltes & U. M. Staudinger (Eds.). *Interactive Minds: Life-Span Perspectives on the Social Foundation of Cognition,* (pp. 242-275). Cambridge: Cambridge University Press.

Smith, J. F. & Kida, T. (1991). Heuristics and Biases: Expertise and Task Realism in Auditing. *Psychological Bulletin, **109**, 3*, 472-489.

Sniezek, J. A. & Buckley, T. (1991). Confidence Depends on Level of Aggregation. *Journal of Behavioral Decision Making, **4***, 263-272.

Sniezek, J. A., Paese, P. W. & Switzer, F. S. III. (1990). The Effect of Choosing on Confidence in Choice. *Organizational Behavior and Human Decision Processes, **46***, 264-282.

Sohn, A. & Gaudiot, J-L. (1992). Learning Legal Moves in Planning Problems: A Connectionist Approach to Examining Legal Moves in the Tower-of-Hanoi. *Engineering Applications in Artificial Intelligence, **5**, 3*, 239-245.

Soll, J. B. (1996). Determinants of Confidence and Miscalibration: The Roles of Random Error and Ecological Structure. *Organizational Behavior and Human Decision Processes, **65**, 2*, 117-137.

Speroff, T., Connors, A. F., Jr. & Dawson, N. V. (2000). Lens Model Analysis of Hemodynamic Status in the Critically Ill. *Medical Decision Making, **9***, 243-252.

Spitz, H. H., Minsky, S. K. & Bessellieu, C. L. (1984). Subgoal Length Versus Full Solution Length in Predicting Tower of Hanoi Problem-Solving Performance. *Bulletin of the Psychonomic Society, 22, 4*, 301-304.

Ste-Marie, D. M. (1999). Expert-Novice Differences in Gymnastic Judging: An Information-Processing Perspective. *Applied Cognitive Psychology, 13*, 269-281.

Stone, D. N. & Ziebart, D. A. (1995). A Model of Financial Incentive Effects in Decision Making. *Organizational Behavior and Human Decision Processes, 61, 3*, 250-261.

Suantak, L., Bolger, F. & Ferrell, W. R. (1996). The Hard-Easy Effect in Subjective Probability Calibration. *Organizational Behavior and Human Decision Processes, 67, 2*, 201-221.

Switzer, F. S. III. & Sniezek, J. A. (1991). Judgment Processes in Motivation: Anchoring and Adjustment Effects on Judgment and Behavior. *Organizational Behavior and Human Decision Processes, 49*, 208-229.

Taylor, S. E. & Brown, J. D. (1988). Illusion and Well-Being: A Social Psychological Perspective on Mental Health. *Psychological Bulletin, 103, 2*, 193-210.

Taylor, S. E., Pham, L. B., Rivkin, I. D. & Armor, D. A. (1998). Mental Simulation, Self-Regulation, and Coping. *American Psychologist, 53, 4*, 429-439.

Tetlock, P. E. (1985). Accountability: The Neglected Social Context of Judgment and Choice. In L. L. Cummings & B. M. Staw (Eds.). *Research in Organizational Behavior: An Annual Series of Analytical Essays and Critical Reviews. Volume 7*, (pp. 297-332). Geenwich, CT: JAI Press Inc.

References (Continued)

Tetlock, P. E. & Boettger, R. (1994). Accountability Amplifies the Status Quo Effect When Change Creates Victims. *Journal of Behavioral Decision Making, 7,* 1-23.

Trafimow, D. & Sniezek, J. A. (1994). Perceived Expertise and its Effect on Confidence. *Organizational Behavior and Human Decision Processes, 45,* 290-302.

Trope, Y. & Liberman, N. (2000). Temporal Construal and Time-Dependent Changes in Preference. *Journal of Personality and Social Psychology, 79, 6*, 876-889.

Tversky, A. & Kahneman, D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology, 5,* 207-232.

Tversky, A. & Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science, 185,* 1124-1131.

Tversky, A. & Kahneman, D. (1982a). Judgment Under Uncertainty: Heuristics and Biases. In D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment Under Uncertainty: Heuristics and Biases*, (pp. 3-20). Cambridge: Cambridge University Press.

Tversky, A. & Kahneman, D. (1982b). Judgments of and by Representativeness. In D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment Under Uncertainty: Heuristics and Biases*, (pp. 84-98). Cambridge: Cambridge University Press.

Tversky, A. & Kahneman, D. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review, 90, 4*, 293-315.

References (Continued)

Tversky, A. & Kahneman, D. (2002). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and Biases: The Psychology of Intuitive Judgment*, (pp. 19-48). New York: Cambridge University Press.

Vallone, R. P., Griffin, D., Lin, S. & Ross, L. (1990). Overconfident Prediction of Future Actions and Outcomes by Self and Others. *Journal of Personality and Social Psychology, 58, 4*, 582-592.

Wigton, R. S. (1996). Social Judgement Theory and Medical Judgement. *Thinking and Reasoning, 2, 2/3*, 175-190.

Ward, G. & Allport, A. (1997). Planning and Problem-Solving Using the Five-disc Tower of London Task. *The Quarterly Journal of Experimental Psychology, 50A, 1*, 49-78.

Welsh, M. C., Satterlee-Cartmell, T. & Stine, M. (1999). Towers of Hanoi and London: Contribution of Working Memory and Inhibition to Performance, *Brain and Cognition, 41*, 231-242.

Wilson, T. D., Houston, C. E., Etling, K. M. & Brekke, N. (1996). A New Look at Anchoring Effects: Basic Anchoring and Its Antecedents. *Journal of Experimental Psychology: General, 125, 4*, 387-402.

Winman, A. (1999). Cognitive Processes Operating in Hindsight. *Scandinavian Journal of Psychology, 40*, 135-145.

References (Continued)

Wright, W. F. & Aboul-Ezz, M. E. (1988). Effects of Extrinsic Incentives on the Quality of Frequency Assessments. *Organizational Behavior and Human Decision Processes, 41,* 143-152.

Wright, W. F. & Anderson, U. (1989). Effects of Situation Familiarity and Financial Incentives on Use of the Anchoring and Adjustment Heuristic for Probability Assessment. *Organizational Behavior and Human Decision Processes, 44,* 68-82.

Wright, G. & Wisudha, A. (1982). Distribution of Probability Assessments for Almanac and Future Event Questions. *Scandinavian Journal of Psychology, 23,* 219-224.

Zakay, D. (1989). Subjective Time and Attentional Resource Allocation: An Integrated Model of Time Estimation. In I. Levin & D. Zakay (Eds.). *Time and Human Cognition: A Lifespan Perspective*, (pp. 365-397). Amsterdam: North Holland.

Zakay, D. & Block, R. A. (1996). The Role of Attention in Time Estimation Processes. In M. A. Pastor & J. Artieda (Eds.). *Advances in Psychology 115: Time, Internal Clocks and Movement*, (pp. 143-164). Amsterdam: Elsevier, North Holland.

Zakay, D. & Block, R. A. (1997). Temporal Cognition. *Current Directions in Psychological Science, 6, 1,* 12-16.

Zakay, D. & Fallach, E. (1984). Immediate and Remote Time Estimation. *Acta Psychologica, 57,* 69-81.

Zakay, D., Nitzan, D. & Glicksohn, J. (1983). The Influence of Task Difficulty and External Tempo on Subjective Time Estimation. *Perception & Psychophysics, 34, 5*, 451-456.