The gestalt characteristics of experiences that extend over time: A further analysis

by

Mr. Shue Loong CHOW

Submitted in accordance with the requirements for the degree of

Doctor in Philosophy (PhD)

Leeds University Business School,

University of Leeds

March, 2007

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Author: Mr. Shue Loong CHOW
Title of thesis: The gestalt characteristics of sequences: A further analysis
Degree: Submitted for the degree of Doctor of Philosophy (PhD)
Date of submission of final revision: March, 2007
Institution: Leeds University Business School, University of Leeds

Abstract

The literature on how people evaluate their sequential experiences makes two contrasting observations. On the one hand, the literature suggests that some of the key findings, such as the preference for an improving sequence, are persistent effects. On the other hand, an emerging, critical literature has pointed out that such effects are highly conditional on a number of factors, such as the sequence domain. The presence of such factors could moderate the preference for improvement. The objective of the studies in this thesis was to explore the role of such factors on people's evaluations of sequences of delays. The results of the studies suggested that gestalt theories such as the preference for improvement in a sequence of delays are not always applicable, but instead conditional on many factors, such as the duration of delays, the degree to which delays prevented task achievement, and people's expectations of how delays are usually experienced.

Key words: Gestalt characteristics, Sequences, Delays, Preference for improvement.

Acknowledgements

The author would like to acknowledge the supervision received from all his supervisors, namely Professor Kevin Keasey (University of Leeds), Professor Daniel Read (University of Durham), and Professor John Maule (University of Leeds).

The author would also like to acknowledge the substantial financial assistance received from the International Institute of Banking and Financial Services at the University of Leeds. In addition, he also acknowledges the financial assistance received from Professor Maule's research account at the University of Leeds, as well as the from the author's own research account at the University of Manchester.

The author would like to acknowledge the technical assistance on web page development received from the Computer Help Desk at Leeds University Business School. In addition, the author would like to thank the entire administration team at Leeds University Business School for all the administrative help received during the course of the PhD.

Last, but certainly not least, the author would like to thank his family, friends and colleagues who have been patiently waiting for him to complete his thesis. The author would also like to thank those who are left out of this section on acknowledgements, since a comprehensive list of names is all but impossible to compile within the space constraints of this thesis.

The author is solely responsible for any omissions and mistakes made.

Total number of pages: 262
Total number of words (excluding diagrams): 79,181

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.0 Introduction

This brief introductory chapter provides an outline of the structure and chapters in this thesis. A more formal and in-depth introduction to the topic is included in the next two chapters (2 and 3). The initial motivation of this thesis is to explore issues that relate to consumers' experiences with internet delays. There is evidence to suggest that such delays are affecting satisfaction with the burgeoning internet retail sector in a detrimental way, which led to calls for more research on consumers' evaluations of internet delays.

Given that internet surfing experiences are often punctuated by sequences of delays, it is proposed that the literature on how people evaluate sequences would be able to provide a useful theoretical starting point for investigating consumers' evaluations of internet sequences of delays. The literature up to the time of the first study is reported in chapter 2.

The objective of the study in chapter 3 is to develop and test a model of how people evaluated sequences of internet delays. This model, called Model 1, was developed based on the theoretical framework discussed in the literature review in chapter 2. The results of this study, however, did not support the hypothesis made for it. This result meant that a re-examination of the model's assumptions and experimental methods was needed to determine the reason(s) for the unexpected result observed. In addition, the emergence of a body of literature soon after the completion of the study (i.e. post-1999) had questioned some of the earlier literature's findings and models. The findings of this post-1999 literature had showed that the applicability of some of the earlier theories was more limited than previously thought. This newer literature focussed instead on investigating the boundaries of the theories proposed in the earlier literature. The review of this more recent literature was presented in chapter 4.

The objective of this thesis had changed from the application of theory to a re-examination of theory. In particular, the focus of the thesis after the conclusion of the first study was to consider in more detail the boundaries or the limits of the theories on how people evaluate sequences. This was the general theme for the remainder of the studies that followed from the first study reported in chapter 3. The broad focus of the studies in chapter 5 was to further refine and improve Model 1. The results would contribute to the literature by providing knowledge on the boundary conditions of the theories on sequence evaluations. The studies in chapter 6 developed Model 1 further by providing additional explanations of the factors that influence the evaluation of delay sequences.

Finally, chapter 7 presents a summary and conclusion of the main findings of this thesis, as well as future directions for research on sequences. The end section of this thesis included a reference list, as well as an extensive appendix that included examples of the actual stimuli used in the studies reported in this thesis.

Chapter 2

The literature review up to the time of the Internet Delay Study (1999)

2.0 Introduction

There are two broad objectives for this chapter. The first objective is to provide an introduction to the literature on delays. The second objective is to provide a rationale and framework for analysing internet delays using theories developed in the literature on how people evaluate sequences, defined here as gestalt theory. The first two sections (including this section) cover the first objective, while the second objective is discussed in the rest of the chapter, from section 2.2 to 2.5. The outline of this chapter is as follows. The remainder of this section describes the context of delays from a marketing perspective, and provides a rationale for further investigations into the effects of internet-based delays. The next section (2.1) reviews the extant literature on delays to provide a foundation on which to base a theoretical construction of how people experience and evaluate internet delays. The second objective is achieved by reviewing the gestalt theory literature. The review of this literature is further divided into three sections, the first of which presents the key findings of the literature (section 2.2), the second discusses its limitations, and the third presents a more in-depth analyses of key methods and parameters used in the literature reviewed (section 2.4). Finally, section 2.5 summarises and concludes this chapter.

Internet delays emerged as a dominant factor in the internet marketing literature towards the end of the 1990s. Anyone who uses the internet would agree that delays rank amongst the highest irritants of computer usage. Many would joke that the world wide web (www) should be renamed the world wide wait, due to slow downloading speeds frequently experienced by the majority of users who connect using dial-up modems. Such frustrations are not merely

restricted to slow technologies on the part of the user, but may also be caused by cumbersome websites themselves.

In addition, delays are costly from an economic perspective. Larson (1987) estimated that Americans spent 37 billion hours per year queuing, with a significant proportion of the lost time related to the purchase of services. He also quoted another study showing that citizens of the former Soviet Union also wasted a similar number of hours in queuing to buy provisions. A more recent study on the effects of time delay by the Boston Consulting Group (as reported in ZDNet, 2000) found that in a survey of 12,000 online customers, 48% of them gave up trying to buy some products because the web pages took too long to load up.

While rapid advances in technology such as broadband have done much to reduce waiting times, the problem of internet delays, however, cannot be completely eliminated in the near future for a variety of reasons. For example, the benefits of greater bandwidth (i.e. faster internet downloading speeds) are often cancelled out by the increased breadth and complexity of services offered by websites, and an increase in the traffic volume of unwanted (i.e. spam) information, which slows internet downloading speeds. In addition, consumers' expectations of a better and more rapid service from websites could also increase and amplify the potential dissatisfaction when the technology fails to meet such increased expectations. Therefore, there is a need to understand how internet delays affect consumers' evaluation of websites.

In 1999, research on the effects of internet delays was just emerging. Researchers had started to use a number of different theoretical frameworks to analyse the effects of such delays. A commonly adopted starting point is the broader marketing literature on delays and waiting in everyday consumer experiences, because there is already a large, well-established literature on such effects on consumer evaluations of their experiences, which is discussed in the next section below. The topic of internet delays constituted a natural extension to such a literature. In addition, the emerging literature on sequence

effects (the sequence literature) was proposed as an alternative perspective in which to view such delays, given that the delays are usually not evaluated in isolation, but often in the context of other events that occurred before and after the delay. These two strands of literature are now explored in greater detail.

2.1 The literature on delays

The objective of this section is to introduce in greater detail the general literature on consumer evaluations of time delays in both an internet and non-internet context. In addition, it will also be argued that the literature on sequence evaluations (i.e. the sequence literature) can be used as a framework in which to analyse internet delays that occur in sequences, which is the main topic of investigation in this thesis.

Consumers consider their time to be a scarce resource, therefore any waiting or delay would mean that precious time is lost, and consequently, they would evaluate this loss as an economic cost to them (Osuna, 1985; Taylor, 1994; Dube, Leclerc and Schmitt, 1991; Larson, 1987). In addition to the economic cost, delays also invoke strong negative emotional reactions such as anger and its related feelings of annoyance, irritation and frustration in consumers' experiences (Sawrey and Telford, 1971). These negative emotions are invoked because consumers perceived the delays as an obstacle that impedes their consumption of the service (Lawson, 1965). Thus, the implications from an internet-based retailer's point of view are that delays have the potential to reduce consumers' evaluations of their retailing experience if nothing is done to mitigate their effects.

However, because the evaluations of such delays are based on human perceptions, this means that there are a number of ways in which retailers can seek to exploit differences between perception and reality. For example, researchers have found that there are discrepancies between the actual duration of the delay experienced (physical time) and perceived duration (subjective time). Hornik (1984) and Maister (1985) argued that perceived duration was one construct that explained consumers' emotional reactions to delay. The longer the consumer's perception of the duration of the delay, the

more dissatisfied they were with the experience (Hornik, 1984). Hornik also argued that while perceived duration was largely a function of physical time, various situational and personal variables could also influence perceived duration. For example, Hornik showed that consumers who enjoyed a shopping experience perceived delays at the checkout as being shorter than those who did not enjoy shopping.

In addition to perceived duration of delays, other environmental factors can also influence consumer's emotional responses towards delays, and mitigate some or all of the negative effects of the delays. For example, a consumer who faces a longer delay for public transport to arrive in a nicer environment (e.g. sheltered and heated waiting room) may evaluate that experience less negatively compared to facing a shorter delay in a more uncomfortable environment (e.g. draughty waiting room). In general, there are many studies which show that the relationship between the waiting duration and the evaluation of an experience is not always inversely proportional, that is, longer waiting times do not always result in more negative evaluations of the service, or more negative emotional reactions (e.g. see Maister, 1985; Larson, 1987; Taylor, 1994; Hui and Tse, 1996; Hui and Zhou, 1996; Hui, Dube and Chebat, 1997).

For example, Maister (1985) and Osuna (1985) showed that retailers could reduce consumer stress and anxiety caused by uncertainty over the delay by providing information on how long the delays would last. In the context of waiting in a queue, Larson (1987) had also argued that consumers put a high value on qualitative aspects of their waiting experience, such as adherence to the 'first come first serve' principle. Consumers may get upset if they experience 'social injustice', where the first to the queue is not the first to get served. This can result in the provider of a service with a shorter queue receiving a poorer evaluation compared to a service with a longer queue but one that strictly adhered to a 'first come first serve' principle. Larson also argued that factors such as the queuing environment and feedback on the expected length of the delays might matter more in consumers' evaluations than the total duration of the delays alone. Larson's research showed that a

more pleasant environment in which to wait, and/or the presence of a mechanism in which to inform consumers on how long they are expected to wait, was able to mitigate some of the adverse effects of delays.

In agreement with the results of Maister (1985) and Osuna (1985), Taylor (1994), along with Hui and his colleagues (Hui and Tse, 1996, Hui and Zhou, 1996 and Hui et al., 1997) also found that information provided on the expected duration of delays contributed significantly to reducing the negative experiences of consumers. Taylor (1994) found that the degree to which people perceived events that caused the delays were controllable was another factor that could mitigate the negative impact of delays. For example, in her research on the effects of delays on consumer evaluations of their experiences with services provided by an airline, Taylor found that consumers got more frustrated when the delays were caused by human incompetence, which was perceived to be more controllable (e.g. poor co-ordination between the airline and airport staff; managerial or personnel incompetence etc.), rather than by factors such as bad weather and force majeure, which were uncontrollable.

Hui and Tse (1996) suggested that there are two strands of arguments that explain the relationship between consumers' perceived controllability over events that causes delays and the evaluation of those delays. The first was through 'information gain', where the increased knowledge on the causes of the delay is similar to that of uncertainty reduction, where the additional information increases the perception of predictability and hence controllability of the experience. Controllability in this respect comes from consumers' ability to anticipate and plan for the delays. This led to the second argument, where the negative impact of delays could be mitigated by increased perceptions of controllability, because people believe that they are able to reappraise the situation they are in, and change their expectations or reference point as to what they consider to be an acceptable service (Hui and Tse, 1996). This was illustrated in an example used by Maister (1985), who wrote, "if a patient in a waiting room is told that the doctor will be delayed thirty minutes, he experiences an initial annoyance but then relaxes into an acceptance of the inevitability of the wait" (p.118). Folkman (1984) argued that reappraisal is an

effective coping mechanism employed by consumers in situations where they have no choice but to wait for the service to be completed.

In addition to notions of perceived controllability in the evaluation of delayed experiences discussed above, Dellaert and Kahn (1999) have also found that the evaluation of a delayed internet surfing experience is dependent on the positioning of the delay within the experience. For example, a delay experienced earlier in the experience was evaluated more aversively to delays experienced in the middle of the experience. This study will be discussed in further detail in chapter 3.

The descriptions of the negative impact that delays have on the evaluation of a service, along with a discussion of factors that mitigate its impact so far has only focussed on delays that occur once during the experience of a service. There are, however, many situations in everyday consumer experiences where people face not just a single delay, but also a sequence of delays. Therefore, an important conceptual distinction has to be made between studies that focussed on single-period delays, where the delays would occur only once within a particular consumer experience, with those that involved multiple delays that occur sequentially. Most if not all of the studies discussed so far had focussed on delays that occur in a single-period, rather than multiple delays occurring in sequence. For example, Taylor's (1994) experiments on consumer experiences with services provided by an airline can be classed as a single-period delay study, because the entire period of delay occurred before the consumer boarded the aircraft. Multi-period sequential delays, however, are delays that a consumer experiences more than once within the same retailing experience. For example, a consumer surfing the internet may face delays every time he/she moves from one webpage to the next. If the entire internet experience is perceived and evaluated as a whole, then the delays at every change of page occur sequentially.

In comparison with single-period delays, multiple delays that occur in sequences have an additional characteristic of sequence 'pattern', which distinguishes them from single-period delays. A sequence pattern is a generic

term used to represent the various sequential combinations for which delays could be ordered, or distributed, across the entire consumer experience. For example, a sequence of delays could be increasing in duration, or constant in duration throughout, or alternating between increasing and decreasing durations (oscillation). Pattern describes the shape that results when such delays are plotted on a two-dimensional graph, where duration of delay (vertical axis) is plotted against time (horizontal axis). For illustration, a man's shopping experience consisted of three queues at the till to pay for the goods. If the first queue was 10 minutes long, the next 15 minutes, and the final queue 20 minutes, the pattern of delays in the sequence of queues is said to be increasing over time. Having to take into consideration the sequence pattern, however, increases the complexity of the analyses, and may reduce the transferability of theories developed in the single-period delay literature (e.g. uncertainty reduction as a strategy to mitigate the negative effects of delay etc.) to a consumer experience with multiple sequential delays. The presence of pattern also introduces a number of other observations that are unique to sequences of multiple delays.

For example, Carmon and Kahneman (1993), in a study that manipulated the various durations of delays that were distributed over a queuing experience, showed that the pattern of delays did influence evaluation. They found that their participants were not impartial to the various sequences of delays with different patterns between them, despite these sequences all having delays of equal total duration. Instead, their participants preferred sequences that ended on an improving note (i.e. one with significantly less delay towards the end) compared to a sequence of delays that did not. Carmon, Shanthikumar and Carmon (1995) argued, in the context of queuing theory, that the conventional wisdom of the theory is to prescribe that delays of services should be spread evenly across the sequence (i.e. one that relied on distributing the delays to make it more convenient for the retailer). However, the findings suggest that this must be balanced against the psychological and behavioural implications of consumer's lack of impartiality to delay patterns. To put the implication of Carmon et al.'s (1995) findings in another way, while the even distribution of delays across a sequence may be operationally and

logistically expedient for the retailer, it may, from a psychological and marketing (i.e. consumers') viewpoint, however, be more beneficial for the marketer to exploit consumers' preferences for some patterns over others, since the ultimate goal of service delivery is to improve consumer satisfaction.

To summarise, the objectives of this section were to introduce the reader to the literature on consumer evaluations of delays. It was conceptualised that delays can be experienced as a single-period, or in multiple-periods in a sequence. The literature on single period delays has been well-researched, where the general findings suggests that it is possible to mitigate some of the negative impact of such delays by providing additional information on the duration of the delays, or by improving the environment in which the consumer experiences the delays. It is unclear, however, whether such methods will apply for an experience that consists of multiple sequential delays. It was argued that it is difficult to extrapolate the findings from single-period delays to multi-period delays experienced in a sequence. This is because sequences of delays have an additional characteristic of sequence pattern that is not present in single-period delays. This means that evaluations of delay sequence may also be influenced by its pattern, in addition to the other factors described earlier in the literature review.

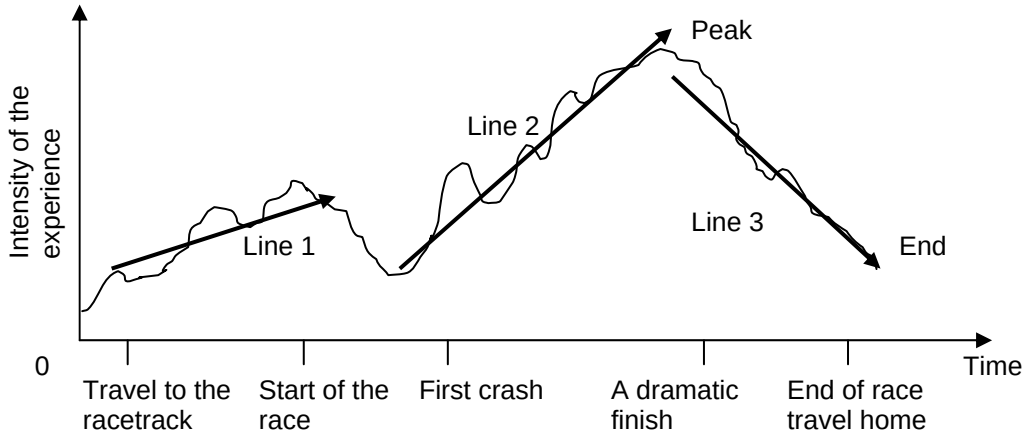
The remainder of this chapter reviews the literature on sequence patterns, also known as the 'sequence literature', in greater detail. The structure of the literature review is as follows. In section 2.2, the discussion focuses on describing the key themes of the sequence literature up to 1999. For orientation, the sub-sections of 2.2 describe the main pattern effects that are found to have significantly influenced people's evaluation of their sequential experiences, which is the trend effect (section 2.2.1), the velocity effect (2.2.2), and the peak-end effect (2.2.3). Section 2.3 discusses the boundary conditions on the applicability of these pattern effects. Section 2.4 summarises and discusses the methodologies used in the literature, while the final section of this chapter (section 2.5) concludes the chapter by summarising the key theoretical and methodological implications in order to develop a model of consumer evaluations of internet delay sequences.

2.2 Review of sequence literature: presentation of key findings

The objective of this section is to review the literature on sequence evaluations, otherwise defined in this thesis as the gestalt theory, as a basis in which to conceptualise how consumers evaluate sequences of internet delays. The issues raised earlier in the chapter pointed to a need for a better understanding of how people evaluate a sequence of multiple delays, as the literature on this topic is substantially less researched compared to the effects of single-period delays. In the previous section, the sequence pattern was briefly described, but is examined in greater detail in this section.

Consider the following hypothetical scenario and the conceptualisation that follows it. Imagine a young boy who has an interest in cars attending, with his father, a Formula One race for the first time. During the car journey to the racetrack, both father and son eagerly discuss their expectations of how the events will unfold. They also make predictions of the race outcome. On arrival, they make their way to their allotted seats, and wait excitedly for the race to start. A few moments later, the race marshal signals to the cars that they have to take their respective starting positions. The lights signal red. The count down begins, and on the lights turning green, the cars let off a mighty roar and accelerate down the podium straight. At the first bend, a crash occurs, which eliminates two cars from the race. The race continues, and lap after lap, a whole series of events that include overtaking, more dramatic crashes, and bungles during the pit stops ensures that the race turns out to be exciting for both father and son. They both leave the race tired but exhilarated. The father then asks his son to (retrospectively) summarise his experience of the race, which the son does by saying that he was thrilled overall. The son also adds that the highlights for him were the start of the race, the crashes, the overtaking manoeuvres and the climax at the end.

Figure 2.1: An example of an experience profile



The Y-axis represents the momentary intensity of the experience, which is the intensity at any one given moment in time. On the Y-axis, higher numbers represent more intense moments, and vice versa. The X-axis represents the progression of time.

This Formula One race scenario is used to illustrate the various issues that relate to the evaluation of experiences that extend over time. Before the race started, the discussion between father and son revolved around how the race would unfold, indicating that they were anticipating or predicting what their future experience might be. After the race concluded, the son then provided a summary of his overall opinion of the race. This evaluation is based on remembered experiences. Throughout the experience, the boy's excitement ebbed and flowed, tracing the highs and the lows of the race. If the boy's emotions (e.g. excitement) are recorded at every moment throughout the race, it can then be represented by an experience profile, where time is represented on the horizontal axis and the moment-to-moment intensity of the emotions on the vertical axis (see Ariely, 1998). Figure 2.1 illustrated the profile of the boy's experience that was just described.

An emerging body of research had explored the relationship between how people summarised their sequential experiences (e.g. when the boy said that he was thrilled with the experience of watching the race with his father) and the experience profile described in Figure 2.1. Normative reasoning suggested that the value or utility communicated from the summary evaluation should be based on a value that could represent the changes in intensity, as well as the

duration, of the experience. Kahneman, Wakker and Sarin (1997) proposed one such principle, where a sequence of stimuli that invoked emotions, say x1, x2, x3,…, xn (where 1, 2, …, n denote the position of the individual stimuli within the sequence), should be evaluated according to the normative utility integration principle, where the utility for the whole sequence is calculated as:

$$U(x1, x2, x3,\ldots,xn)=U(x1)+U(x2)+U(x3)+\ldots+U(xn)$$

For example, if the sequence had the following distribution of moment-to-moment intensity {2,5,8}, where the time intervals between each moment are equal, and where the range of intensity is from 1 to 10, with 10 being the highest and 1 being the lowest, evaluation based on utility integration would give a total of (2+5+8)=15. Therefore, if {2,5,8} was compared to a sequence with the distribution, or pattern {8,5,2}, the utility integration principle predicted that people would be indifferent between the two, as the integrated utility for both is 15. However, research findings on the summary evaluations of experiences often did not support such a prediction. Instead, this body of research reported that the value of the integral of utility mattered less in evaluation compared to the sequence pattern (for a review of this literature, see Kahneman et al., 1997). One reason was that people were often not able, or willing, to use such normative rules in evaluation. Instead, they chose to summarise their experiences using only the sequence patterns, such as the peak and end moments, or its trend, or velocity. These three patterns were illustrated in Figure 2.1, and are the patterns that would be discussed in this literature review. Line 1 and line 2 represented examples of an increasing trend, and line 3 a decreasing trend. The slopes of line 1 and line 2 represented velocity, with the velocity of line 2 higher compared to line 1, due to the steeper slope of the former (i.e. greater magnitude in the rate of change). The moment with the highest intensity in the experience is defined as the peak of the profile, while the final moment or conclusion of the experience represents the end of the profile. The following subsections reviews the literature that investigates the effects of these three patterns (trend, velocity and peak-end) on the evaluation of sequences.

2.2.1 The trend effect

In the sequence literature, the trend effect refers to people's preference for a sequence of events that are ordered in a particular fashion. The literature's interest in the trend effect came from studies that investigated people's preferences for sequences of events or sequences of outcomes where the events had different orders. For example, Ross and Simonson (1991) found that in a choice between two pairs of outcomes, one negative and the other positive, people preferred the pair of outcomes where the negative outcome was experienced first before the positive outcome. In one their studies, Ross and Simonson found that people preferred the pair of events where the loss of money was experienced before the gain of money, rather than vice versa. The finding that people preferred an improving, rather than deteriorating, trend was also replicated in another of Ross and Simonson's study in the same paper, but this time the domain used was that of a computer game, rather than monetary sequences, and where the outcome was success/failure at the game. Their finding in this study suggested that the preference for improvement also extended to choices between sequences of non-monetary events. Ross and Simonson's (1991) paper, however, did not provide any substantial theoretical explanations as to why people have such preferences for particular sequence patterns such as an improving trend.

Loewenstein and Sicherman's (1991) approach in developing theory to explain the preference for sequence patterns was focussed on departures from the predictions of rational choice. Their model proposed is quite similar to the one prescribed in Kahneman et al. (1997), but with the crucial difference that Loewenstein and Sicherman (1991) had only discussed this model in the context of sequences of monetary stimuli, while the Kahneman et al. (1997) model extends the rationalising motives to other non-monetary stimuli as well. Rational choice theory suggested that people should make decisions that maximise their utility. For choices of monetary sequences over time, people should choose sequence patterns that provided the highest returns. The tool used to calculate a sequence with the highest return is Present Value (PV), which is a technique that converted the value of future income streams into today's money using the discount rate for future income. If the discount rate is

positive (in the developed world, the discount rate for money is usually positive, and is often proxied by the interest rates set by a central bank), then each £1 of money in the future is worth less than it is now. In addition, the further into the future the money is received, the smaller its present value will be. Negative discount rates work in the opposite way, where money in the future is worth more (in real terms) than at present. Since interest rates for money are usually positive, it is also assumed that people's discount rate is positive, which means that people will prefer an income sequence (£500 now, £400 in a years' time, £300 in two years, £200 in three years and £100 in four years) should be preferred to the sequence (£100 now, £200 in a years' time, £300 in two years, £400 in three years and £500 in four years), because the former sequence has a higher present value.

However, Loewenstein and Sicherman (1991) challenged the rational presumption that people should always choose the monetary sequence that maximises their present value, or always have positive discount rates. They found that people in the real world did not exhibit such preferences. In a survey-based study that asked people to rank a series of monetary sequences that varied in present value and in sequence pattern, Loewenstein and Sicherman reported that people preferred some types of monetary sequences to others. For example, an increasing income pattern is favoured over a decreasing one, even though the decreasing pattern may have a higher present value compared to the increasing pattern. Rationally, this represents a sub-optimal choice on the part of the decision-maker, because the sequence that gave the highest present value, which maximised the utility from the income stream, was not chosen or preferred.

Loewenstein and Sicherman argued that while this observed behaviour might not necessarily be optimal from a PV maximising viewpoint, what it suggested was that people's preference for monetary sequences was also influenced by a number of complex psychological factors that may not be adequately captured in the rational PV model. Instead, they found that in addition to rational principles, behavioural factors such as signalling, self-control, adaptation, and anticipation were also important and this could explain people's preference for

monetary sequences deviated from PV predictions. Further explanation of these four concepts is presented next.

In signalling, people associated their income with productivity, and therefore derived utility from a feeling that their job performance was improving when their income increased. In this case, the pattern of the sequence (i.e. the sequence of increasing income over time) represented the signal that the worker was increasingly improving in his/her job, which led to the increases in income. The second factor, self-control, denoted a desire for people to match the pattern of consumption with the pattern in which the income is received. For example, a young worker planning to have a larger family in the future would therefore need more income in the future rather than now. If, however, people receive more money now, and less later (i.e. the PV optimal choice of sequence patterns), they are afraid that they would not be able to save the extra income received now for the future, due to poor self-control. Therefore, they may prefer to receive more of the money at a later period rather than earlier. This suggests that people obtain positive utility from the sequence pattern itself, because it helps them to control their spending in such a way that they were able to match the pattern in which income is received with the pattern in which the income is consumed.

The third factor, adaptation, is based on the assumption that workers grow increasingly accustomed to a particular level of consumption, and therefore would find it more difficult to adjust to a pattern of decreasing income over time. Adaptation is similar to self-control, in the sense that people would find it behaviourally difficult to match the pattern of income with the pattern of consumption. From a rational viewpoint, people could easily save the money earned in the earlier periods for spending in the later periods, but from a behavioural viewpoint, adaptation indicates that people would find it difficult to manage the matching of savings patterns and with consumption patterns.

The fourth and final factor, anticipation, is based on the assumption that monetary sequences with increasing patterns are preferred because people can anticipate and ‘savour’ the increase in future income at the current point in

time (e.g. positive utility is generated from people thinking about how they would spend their future increases in income). Decreasing patterns, however, are less preferable, because people 'dread' the future decrease even though it has not happened yet (see Loewenstein, 1987).

The four explanations proposed by Loewenstein and Sicherman (1991) were also supported in Schmitt and Kemper's (1996) study, which found that people preferred sequences with increasing monetary rewards. Schmitt and Kemper's main contribution to the literature was that their experiments actually paid out monetary rewards to participants, which Schmitt and Kemper claimed to make the choices between sequences more realistic, compared to the survey-based method adopted in Loewenstein and Sicherman (1991). Subsequent studies that explored the preference for an improving trend found that the phenomena could also be extended to non-monetary experiences, such as restaurant meals, and social activities such as visiting friends and family (Loewenstein and Prelec, 1993). For example, in one of Loewenstein and Prelec's studies, they asked people to choose between two sequences, where the first option is to eat out at a fancy French restaurant this weekend, but eat at home for the next two weekends, versus the second option to eat at home this weekend, at the French restaurant the next weekend, and at home the weekend after. The large majority of participants preferred the second choice (84%) as opposed to the first (16%), indicating that they prefer to delay their rewards and to control the pattern of their dining experiences to create an improving trend. In another illustrative study that demonstrated the preference for an improving trend, Loewenstein and Prelec (1993) found that 90% of their participants preferred the sequence where they got to spend the first of two weekends with an abrasive aunt, followed by a pleasant weekend with their friends, rather than the other way around.

Chapman (1996a) also found that people generally prefer a sequence of increasing income over their life time, which further supported Loewenstein and Prelec's (1993) findings. However, she found that people did not always prefer an improving trend for health sequences. For short-term preferences, she found that people preferred the increasing trend for both health and

money. Over the lifetime of a person, however, she found that people still preferred a sequence of increasing income, but a sequence of decreasing (i.e. deteriorating) health. She attributed such a result to the change in expectations of how sequences should progress in the real world, where people expect their health to improve over the short term (via their expectations that they will improve their diet, exercise, and other improving environmental factors), but realise that over their life-time, their health would deteriorate, and that they would eventually die. For income, people expected that in their life-time, their incomes would increase to reflect the progress and experience that they expected to accumulate in their jobs, and to a lesser extent, to reflect the fact that their income would keep pace with inflation. Therefore, expectations have an important role to play in shaping people's preferences for sequences, and in certain circumstances, might even moderate the preference for an improving trend. Expectations in this respect are a driver of preferences, and Chapman has provided a further explanation, on top of the four factors presented by Loewenstein and Sicherman (1991) earlier, for why people prefer some sequence patterns to others.

In summary, the trend effect is a robust pattern effect that is found in many experiences that extended over time, ranging from monetary sequences, to health, educational attainment, and social events. For monetary sequences, the preference for improvement can lead to sub-optimal decisions being made, especially when the sequence with the improving trend is not the utility maximising sequence in the narrowly defined rational choice using PV techniques. The broad reason attributed to preferences for sub-optimal sequences is traced to behavioural issues such as signalling, self-control, adaptation, anticipation (of how sequence outcomes are consumed) and expectations (of how sequence outcomes should turn out). For non-monetary sequences, the reasons for the preference for an improving trend were more complex, which ranged from savouring and dread (e.g. in social situations such as visits to friends and family, and restaurant meals) to socially based expectations of how sequences should progress, such as long-term health.

2.2.2 The velocity effect

In a related development to studies that investigated the trend effect, Hsee and Abelson (1991) proposed that another sequence pattern, which they called velocity, also affected the way in which people evaluate sequences. Their theory postulated that people's evaluation of a sequence is dependent on the rate of change of the outcome. The more rapid the rate of change for a sequence of events that is positive in nature, the more satisfied people are, and vice versa for events that are negative. This is different from the trend effect, because it measured the rate of change, not just the direction of change (i.e. the trend) or the magnitude of the change (Hsee and Abelson called this position). Hsee and Abelson's theory enabled trends in sequences to be differentiated by their rate of change, in addition to their position. Hsee and Abelson (1991) presented the evidence to support their hypothesis in two studies. In their first survey-based study, they asked people to evaluate changes in their income and academic position in a university. Their results showed that people preferred sequences where their salaries and academic positions improved at a faster rate compared to a slower one. In their second, laboratory-based study, they replicated their earlier findings under two additional domains, where participants were required to evaluate sequences of gambles, academic position, and share prices. These findings led Hsee and his colleagues to develop a more formal theory to explain the relationship between position and velocity. Hsee, Abelson and Salovey's (1991) model stated that the evaluation with a sequence of events (instead of evaluation, Hsee et al. called it satisfaction, or S) is a weighted average of its position and velocity. Notationally, Hsee et al.'s (1991) model is written as:

S = w.Position + (1-w).Velocity, where w is the relative weight.

There are a number of implications of this model for sequence evaluation, 'S'. The first was what Hsee et al. (1991) called the dynamic framing effect, where the emphasis of the frame in which the sequence was experienced would also influence evaluation. Framing in this regard is similar to the framing effect reported in Tversky and Kahneman's (1981) study, where people's preferences vary according to the context in which the decision is framed, even

though the outcomes for all frames are equivalent (i.e. lives lost in a health epidemic versus lives saved). In a study that asked people to evaluate sequences of income and class position, Hsee et al. found that people preferred the sequence with the more favourable position outcome (position in this case was defined as the average yearly income, while class position was defined as the average of all grades for pupils in a particular class) when the income (or class position) was framed as culmulative averages over the years. However, when the sequence was framed in terms of the rate in which the outcomes were changing over time, people then preferred the sequence with the more favourable velocity outcome. An example of the stimuli that Hsee et al. used is provided to illustrate the dynamic framing manipulation (all paraphrased from Hsee et al., 1991, p.264 para.2). The dependent variable satisfaction (S) was measured on a 9-point scale, with 1 indicating greatest satisfaction with the 'greater positional outcome' and 9 indicating greatest satisfaction with the 'greater velocity outcome'.

Illustration of the dynamic framing manipulation used in Hsee et al. (1991):

Emphasis on the average:
1a) Greater Positional Outcome: Your salary was $18,000 for the first year; your average salary over the first two years was $17,500, over the first three years $17,000, and over the entire four years was $16,500.

1b) Greater Velocity Outcome: Your salary was $12,000 for the first year; your average salary over the first two years was $12,500, over the first three years $13,000, and over the entire four years was $13,500.

Emphasis on the rate of change:
2a) Greater Positional Outcome: Your salary decreased $1,000 for each of the four years you worked, and your salary for the last year was $15,000.

2b) Greater Velocity Outcome: Your salary increased $1,000 for each of the four years you worked, and your salary for the last year was $15,000.

In Hsee et al.'s (1991) example above, participants were asked to rate their satisfaction between 'a' versus 'b'. The choice between 'a' and 'b' was such that either the average or the rate of change was emphasised. Their results showed that 'b' had higher satisfaction ratings when the rate of change was emphasised, but were more satisfied with 'a' when the average was emphasised. This indicated that the dynamic frame was influential in evaluation, despite the fact that the outcomes under both framing conditions (emphasis on average versus emphasis on satisfaction) were identical (i.e. both frames were tested using either a decreasing sequence of $18000, $17000, $16000 and $15000 from years one to four or an increasing sequence of $15000, $16000, $17000 and $18000).

In addition to the dynamic framing effect, Hsee et al. (1991) postulated that another two factors, motive and control, affected the relative weighting between position and velocity. Motive here is defined as the reasons that people have when they experience a sequence of events. For the purposes of this argument it is separated into two, instrumental and consummatory. An instrumental motive is one where people undertake a sequence of tasks with the sole intention of achieving a final outcome, such as house chores. This sequence is experienced purely as a means to an end, to derive utility from the final outcome of the sequence (e.g. enjoy a clean and tidy house). A consummatory motive, on the other hand, also provides utility throughout the progression of the sequence of tasks. In a mountaineering example taken from Loewenstein (1999), climbers also enjoy the sequence of challenges that they have to undertake in order to climb a mountain, in addition to the sense of achievement once they finally reach the top of the mountain. Hsee et al. (1991) found that position weighed more in sequence evaluation when the motive of the experience was instrumental, but velocity weighed more when the motive was consummatory. For control, they found that its external versus internal nature determined the relative weighting of position with velocity. Internal control is defined as the condition where a person has some influence over the outcome of the sequence, whereas external control is defined as the condition where the outcome is beyond his or her control. Hsee et al. found

that if the outcome of the sequence was perceived as internally controlled, the velocity effect would be stronger than position, and vice versa for perceptions of external control.

To recap, the contribution of Hsee and Abelson (1991) was to show that people do take velocity into consideration when evaluating a sequence, but there were a number of conditions that moderated its application, such as the way in which the decision is framed, the motive for pursuing the experience, and the perceived control over the outcomes (Hsee et al., 1991). In addition to velocity effects, Hsee et al. (1994) also found that 'quasi-acceleration effects' had an influence on sequence evaluation. They found that, in a number of computer-based experiments that were very similar to those employed by Hsee and Abelson (1991), people's satisfaction was greater the more positive the changes in velocity, but was less when the changes in velocity were smaller. In summary, velocity effects were very similar to that of the trend effect, except they measure the rate of change, as well as the changes in levels. However, such pattern effects were moderated by a number of factors, such as the decision frame, motive and control. This later finding strongly suggests that future research needs to be more specific about the conditions under which their results would apply.

2.2.3 The peak-end effect

Kahneman and his colleagues (Varey and Kahneman, 1992; Fredrickson and Kahneman, 1993; Carmon and Kahneman, 1993; Kahneman, Fredrickson, Schreiber and Redelmeier, 1993; Redelmeier and Kahneman, 1996; Kahneman et al., 1997) proposed that some of the pattern effects described in the studies reported above could be interpreted in the context of their 'peak-end' model. This model postulated that the peak intensity of an experience, and its end intensity, could account for a large proportion of people's evaluation of a sequence.

Varey and Kahneman (1992) carried out a survey-based study where their participants had to provide overall (or global) evaluations of a number of sequences that recorded their moment-by-moment experience of an

uncomfortable experience (such as sitting in a vibrating room, exposure to loud drilling or hissing noises, or standing in an uncomfortable position). Their participants were told that these moment-by-moment evaluations were recorded at constant 5 minute intervals, and was made up of numbers that ranged from 0 (no discomfort at all) to 10 (almost unbearable). The dependent variable that was used to capture the overall evaluations of the moment-by-moment ratings was a question that asked participants 'how bad the overall experience is' (p.178, para.1), on a 0 to 100 point scale, where 0 is not bad, to 100, which is extremely bad. The results of Varey and Kahneman's study suggested that the overall evaluations of the sequence suggested that decreasing sequences of aversive stimuli were preferred to increasing sequences of aversive stimuli, even if the utility integral for both sequences were equal. For example, for the sequence {8,5,2} (in Varey and Kahneman's study, this sequence had a 15 minute total duration, because it comprised of three moment-by-moment intensity evaluations at a five minute intervals for each evaluation) was rated more favourably (at 44.9 on a 0-100 scale) compared to one with {2,5,8} (at a higher and more unfavourable 64.0), despite both patterns having identical utility integrals (i.e. (8+5+2)=(2+5+8)=15). In addition, Varey and Kahneman also pointed out that the longer 20 minute pattern {2,5,8,4} had a more favourable overall evaluation (53.0) than the shorter 15 minute {2,5,8] pattern (64.0). They proposed that this evaluation violated the normative principle of utility integration, where people should prefer to experience less overall discomfort to more overall discomfort. Varey and Kahneman explained that people base their overall evaluations of sequences on only key features of a sequence, such as its trend (was it increasing or decreasing), peaks, and the way in which it ended. For example, they found that people prefer sequences that ended on a better note, even if that sequence of discomfort had an ending that was prolonged.

Fredrickson and Kahneman (1993) further tested the finding that people rely on the key features of a sequence. In their study, they employed a laboratory-based study that required participants to provide moment-by-moment and global retrospective evaluations of both pleasant and aversive film clips. These pleasant film clips included a short (range from 29 to 41 seconds) and

long (range from 75 to 125 seconds) clip showing a puppy playing with a flower, waves breaking on a beach, ski jumping, flying over African landscapes etc., while unpleasant clips consisted of the aftermath of Hiroshima, dying people in Singapore, and a medical film of an amputation. Fredrickson and Kahneman's general conclusion was that people's retrospective evaluations of the film clips could largely be determined by an un-weighted average of the peak and end moments of the clips, which did not seem to have taken into account of the duration of the sequence. For example, in a comparison between unpleasant films, a film that ends in an unpleasant manner could be evaluated more negatively than a similar film that ended somewhat better, despite the latter film having a longer duration (i.e. more overall exposure to unpleasantness).

The importance of the peak and end of a sequence in influencing evaluation was further tested in a cold water experiment by Kahneman, Fredrickson, Schreiber and Redelmeier (1993), where their participants were required to immerse one of their hands in a cold water basin at 14 degrees C for 60 seconds, and the other at 14 degrees C for 60 seconds, followed by 15 degrees C for an additional 30 seconds. Kahneman et al. (1993) found that when participants were asked which of the two cold-water experiences they preferred to repeat, the significant majority preferred the longer one. Kahneman et al. (1993)'s main conclusion was that this violation of temporal monotonicity (i.e. when an experience with a longer duration of pain is preferred to one with a shorter duration) was a mistake, and that participants preferred the longer experience only because they remembered that it ended on a better note than the shorter experience, not for any masochistic reasons. Kahneman et al. (1993) added that it was reasonable to assume that people would prefer to repeat experiences that they remembered to be more pleasant, but that their study has showed that such a decision could prove fallible because people's memories for hedonic and affective experiences, such as those from the pain induced by the cold water trials, were often unreliable (e.g. see Kent, 1985; Rachman and Eyrl, 1989; Thomas and Diener, 1990).

These memory failings and the dominance of the peak and end moments of a sequence in influencing evaluation were also reported in other real-world contexts, such as in painful medical treatments (Redelmeier and Kahneman, 1996) and TV advertisements (Baumgartner, Sujan and Padgett, 1997). For example, Redelmeier and Kahneman (1996) found that patients at a Canadian hospital who had just undergone painful medical procedures such as colonoscopy or lithotripsy had evaluations of total pain that were based largely on the peak (i.e. most painful moment in the experience) or end (i.e. final moment in the experience) of the procedure, rather than on total the amount of pain, as defined by the utility integral of the momentary pain with duration. This result was analogous to Kahneman et al.'s (1993) experiments, in the sense that individuals demonstrated substantial variations in their memory of total pain, and that the sequence patterns of their painful experiences was used, to a significant extent, to determine their retrospective evaluations of their experiences. In a study on consumer evaluations of TV advertisement, Baumgartner et al. (1997) found that the duration of an advertisement was only a small predictor of consumers' evaluations of advertisements that they had watched, compared to the peak and end moments of an advertisement, which was shown to significantly influence evaluation. In Baumgartner et al.'s study, they showed that consumers preferred delayed peaks, high ends and an improving trend, which corresponded to the predictions that people generally prefer sequences that were improving over time and/or one that ended on a better note. Baumgartner et al. argued that the adaptation explanation developed by Loewenstein and Sicherman (1991), was a mechanism that explained consumers' need for improvement in their experiences. The principle of adaptation to advertising stimuli is one where people get used to the positive emotions generated by the advertisement, and therefore required increasing amounts of stimulation to be captivated by the advertisement. This explained why they preferred an increasingly stimulating (i.e. one with characteristics such as delayed peaks and improving ends) sequence over one that has the opposite patterns (i.e. early peaks and deteriorating ends).

The substantial evidence of the effects of peaks and ends on sequence evaluation had been summarised and formalised into a theory by Kahneman,

Wakker and Sarin (1997). In the peak-end theory, it is proposed that the evaluation of a sequence is only dependent on the unweighted average of its peak and its end. Returning to the illustrative example provided earlier by Varey and Kahneman (1992), application of the peak-end theory to the two sequences {2,5,8} and {2,5,8,4} will give a peak-end intensity of ((8+8)/2)=8 for the first profile (8 represents both the peak and the end intensity of the sequence), and ((8+4)/2)=6 for the second. Since higher numbers in this example represent more intense aversion, the first sequence has a less favourable evaluation compared to the second, because of the higher peak-end average of the first sequence (8 versus 6). This means that sequences of negative stimuli with an extension added on at the end are not necessarily evaluated as being more aversive compared with a sequence without one, even if it has a higher amount of aversiveness, measured by utility integration. For example, the sequence {2,5,8} has a utility integral of 15, which is smaller (and therefore has less total or integrated aversiveness) compared to the sequence {2,5,8,4}, with a utility integral of (2+5+8+4)=19. Therefore, this phenomenon was termed duration neglect because evaluation by peak-end favoured the longer sequence over the shorter (6 versus 8), even though the longer sequence had a higher total aversiveness, measured by utility integration (19 versus 15). Evidence of such duration neglect had been demonstrated in a variety of different sequence domains, including real-life medical treatments, such as colonoscopy and lithotripsy (Varey and Kahneman (1992); Fredrickson and Kahneman (1993); Kahneman et al. (1993); Redelmeier and Kahneman (1996); Baumgartner et al. (1997) and Kahneman et al. (1997)..

In summary, the peak-end theory prescribed that the evaluation of sequences can be represented by the un-weighted average of its most intense moment of a sequence (its peak) and the way it concluded (its end). The sequence stimuli in which the peak-end effect could be found were varied, but it usually involved stimuli of an aversive nature. Reliance on the peak and end moments of a sequence for evaluation, however, may lead to duration neglect, where people were willing to experience more aversive stimuli to obtain a more favourable overall experience of the sequence based on peak-end evaluation.

To conclude, this section (2.2) reviewed the gestalt theory literature to provide a theoretical foundation for the forthcoming conceptualisation (in the next chapter) on how consumers evaluate sequences of internet delays. Three key themes in gestalt theory were discussed, trend, velocity and peak-end, which was based on the type of sequence patterns observed. The main finding was that people do not simply evaluate a sequence based on an integral of the event intensity over time (i.e. the Utility Integration Principle or UIP), but also consider factors such as patterns, in their evaluations. The next section discusses some of the factors that limit the influence of patterns on sequence evaluation.

2.3 Review of sequence literature: limitations of the findings

So far, the review of the literature has largely focussed on extending the domains in which such pattern effects might apply to sequence evaluation. There are, however, a handful of studies that raise questions over the transferability of such findings across different domains, or at least qualify the conditions and domains in which such effects might apply. In Hsee et al. (1991), the influence of velocity over sequence evaluation was determined by three factors, namely framing (i.e. was velocity or position emphasised more in the framing of the decision?), motive (i.e. was the sequence experienced for instrumental or consummatory purposes?), and control (i.e. do people perceive that the outcomes of the sequence is within their control?). Chapman (1996a) found that people's expectations of how sequences should progress, which was often based on observations from their real-world experiences, might moderate the preference for an improving trend/end. This was true for the case of long-term health sequences.

In a study on painful sequences, that was partly inspired by the results and theories on duration neglect proposed in Kahneman et al. (1993) and Redelmeier and Kahneman (1996), Ariely (1998) found some support for the proposition made by Kahneman and his colleagues that the effects of patterns (such as peak-end, trend and velocity) had a significant influence over the evaluation of painful stimuli. Ariely (1998), however, reported that there were a

number of important theoretical and methodological qualifications to these effects that warranted further attention. In his study, two different methods of inducing pain were used. Since some of Ariely's (1998) concerns were related to the way in which the dependent variable was measured in gestalt research, it is necessary to provide a brief description of the experiments that he performed.

Ariely's (1998) first experiment required participants to make contact with a heated thermode, where the experimenter controlled exposure to the device, in terms of its duration and temperature. Duration (two levels, at 10 and 14 seconds) and temperature (ranging from 45 to 48 degrees C) was systematically varied by creating a number of 'intensity-patterns', such as 'low&up', where the first half of the sequence (i.e. for a duration of 5 or 7 seconds) started off with a temperature of 45 C, and gradually increased over the second half of the sequence to finish at 48 C. Other examples of intensity patterns include 'high&down' (first half at 48 C and finished at 45 C), 'up&down' (start at 45 C, increasing to 48C halfway through, and finished at 45 C) and 'constant' (which was held at the same temperature throughout the sequence for four different temperature gradients, 45 to 48 C). The main result of his first experiment was that the end intensity as well as the final trend of the different patterns played an important role in influencing global retrospective (i.e. post-experience) evaluations of painful sequences. This result supported the conclusions reached in the literature discussed in the previous section (e.g. Kahneman et al., 1993 and 1997; Hsee and Abelson, 1991; Hsee et al., 1991), which had demonstrated that sequence patterns (e.g. trend, velocity, peak and end) had a significant effect on people's evaluations of aversive sequences. In contrast with the duration neglect concept proposed by Kahneman and his colleagues (Kahneman et al., 1993 and 1997; Redelmeier and Kahneman, 1996), however, Ariely (1998) argued that there is a flaw in the method used to support the duration neglect concept. Ariely showed, from the results of his first experiment, that duration neglect was only present in the constant case (i.e. where the temperature was held constant over time), but not when temperatures were varied across the sequence (i.e. the other intensity-patterns). He explained that people relied on changes in the stimuli intensity to

help them perceive the amount of time that had passed, a concept illustrated by Poynter and Holma (1985), which meant that duration neglect was more likely to occur when participants experienced a constant intensity of pain over the sequence, because they were unable to obtain any information on the passing of time by inferring it from the changes in stimuli intensity. Ariely argued that some of the methods used by Kahneman and his colleagues, for example, Kahneman et al. (1993), may need further testing and replication. This is to ensure that the observed neglect of duration was not confounded by participants' inability to distinguish the amount of time that had passed.

The main purpose of Ariely's (1998) second experiment was to investigate the effects of having to provide moment-by-moment responses, in addition to that of an overall evaluation of the sequence. A moment-by-moment response was a key dependent variable used by Kahneman and his colleagues in their research on the peak-end effects in sequences. In his second experiment, Ariely utilised an alternative method to induce pain, by locking participants' fingers in a vice, which was then used to vary the physical pressure exerted on the finger. Participants were required to provide two dependent variable responses, one of which was the moment-by-moment response of how painful they feel the physical pressure generated by the vice was, and the second was the global retrospective evaluation of their experience. Ariely's results indicated that the measurement of moment-to-moment responses had confounded the sequence evaluations, because it influenced retrospective evaluations of sequences when it should not have. Ariely explained that when participants were asked to provide such momentary responses, the saliency of such responses increased, thus affecting their retrospective evaluations of the sequence. This is because when participants are providing moment-by-moment responses, they are not attending to the sequential stimuli. Ariely claimed that this questioned the validity of the experimental methods used by Kahneman and his colleagues in their research on duration neglect (Kahneman et al., 1993; Redelmeier and Kahneman, 1996; Baumgartner et al., 1997; Kahneman et al., 1997), because Kahneman and his colleagues' studies did not control for the possibility that the feedback from moment-by-moment evaluations might contaminate the retrospective evaluations.

A further implication of Ariely's (1998) study for the gestalt research literature was its demonstration that many of the theories (e.g. velocity, trend and peak-end) developed in studies that employed prospective methods for measuring sequence evaluation were also supported in his experiments using retrospective methods. The methodological distinction between prospective versus retrospective studies is important, because in prospective evaluations, participants do not actually experience the sequential stimuli and report back on their experience, but instead have to imagine themselves experiencing the sequence and provide anticipated responses on how they think they would feel. In the Formula One example illustrated earlier in this chapter, the boy and his father's predictions of how the race would unfold could be classed as a prospective evaluation, and the boy's opinions at the end of the race as a retrospective evaluation. Although both prospective and retrospective evaluations involved elements of memory and judgement, it is important to consider the distinction between the methods used to measure people's evaluation of sequences, because of the different psychological processes involved. Elster and Loewenstein (1992) illustrated one of these distinctions:

> "Anticipation is similar to memory in that the consumption and contrast effects are operative. But there is a crucial difference between backward and forward effects. Memory has as its object events that have already occurred, which are thus *certainties*. Anticipation focuses on events in the future that are by this fact inherently *uncertain*. The *likelihood* that the anticipated event will actually occur is an important determinant of the intensity of savouring and dread." (para.2, p.227, emphasis in italics added).

So far, the evidence has suggested that the pattern effects found using prospective methods, such as trend (Ross and Simonson, 1991; Loewenstein and Sicherman, 1991; Loewenstein and Prelec, 1993; Chapman, 1996a), velocity (Hsee and Abelson, 1991; Hsee et al., 1991; Hsee et al., 1994) and peak-end (Varey and Kahneman, 1992) have also been successfully replicated under a retrospective methodology (trend – Ross and Simonson, 1991; Schmitt and Kemper, 1996; Ariely, 1998; velocity – Hsee and Abelson, 1991; Ariely, 1998; peak-end – Fredrickson and Kahneman, 1993; Kahneman et al.,

1993; Redelmeier and Kahneman, 1996; Baumgartner et al., 1997; Kahneman et al., 1997, Ariely, 1998). In addition, the literature have not reported any difficulties or divergences in the results obtained from the two different experimental paradigms (i.e. prospective v retrospective methods) used.

To conclude this section, while a couple of earlier studies (Hsee et al., 1991; and Chapman, 1996a) have attempted to specify some of the conditions under which people were more likely to rely on sequence patterns for evaluation, the most significant criticism so far of gestalt theories (especially duration neglect) had come from Ariely (1998). Ariely's findings were mostly supportive of the broad findings in the literature, where people preferred improving endings and improving trends. However, he claimed that there is a need for further evidence to support the duration neglect concept proposed, given the criticisms of the methods used. In addition, Ariely also showed that the use of moment-by-moment evaluations during the sequential experience diverted participants' attention away from the sequential stimuli itself, which confounded the retrospective global evaluation of the sequence.

As a summary of the first half of this chapter, section 2.1 reviewed the literature on delays, and proposed that sequences of internet delays can be analysed using the theoretical framework developed in the gestalt sequence literature. Section 2.2 reviewed the gestalt sequence literature, which had showed that sequence patterns had an important and significant influence over evaluation and/or preferences. There were, however, some methodological concerns over the theories developed in the literature that were highlighted in the present section (2.3), but these criticisms were specific to the choice of methodologies used, rather than questioning the overall soundness of the theories developed. The next section examines the methodological aspects of the studies reviewed up to 1999 in greater detail.

2.4 Review of sequence literature: A systematic review of the methods and results

The two previous sections have covered key theoretical themes of the sequence literature, and discussed some of its limitations. The primary purpose of this section, however, is to present a more systematic review of the methods and results from the experimental studies, given that methodological issues such as type of stimuli used, choice of dependent and independent variables, and other concerns over the experimental design have not yet been debated in a systematic manner. There are a number of uses for the results of these analyses. First, they may provide additional insights on the relationship between the various methods and parameters used in the experiments that had not been previously discussed. Second, these analyses can also be used as a guide to inform the choice of methods and parameters for use in future experimentation. Third, the separation of the methods and results review from the general literature review in section 2.3 and 2.4 will enable the construction of a table to summarise key theoretical and methodological aspects of the literature. A tabular representation will make it easier to update the findings of newer literature without having to reconstruct the entire review of theory and methodology every time a new study is published, especially if the findings of the study complements or extends current theory. The ability to represent the key methodological constructs and summary findings of the literature in tabular form is an important point, because it can help the reader follow more easily the evolution in the literature, which is separated into two in this thesis. As a reminder, the literature up to 1999 is reported in this chapter, while the literature post-1999 is reported in chapter 4, after the completion of the first study (chapter 3). The outline for this section is as follows. The next sub-section (2.4.1) describes the methods used to analyse the key methodological constructs in the literature. Section 2.4.2 presents the results of these analyses in tabular form, and section 2.4.3 discusses the implications of the results.

2.4.1 Method

This sub-section describes the methods used in the analyses of key methodological constructs used in the literature. First, the key methods and parameters used are summarised in a table with separate categories (see Table 2.1), which is based on one used in Guyse et al. (2002, p. 268, Table 3). These categories, however, are not specific to the analyses presented by Guyse et al., but are generic to the literature. The categories represent a summary of the key independent and dependent variable manipulations adopted, number and type of participants used, time horizon or duration of the sequence, and key findings of a study, which are standard issues that any researcher in the field of gestalt theory has to consider. In addition, all papers containing experimental studies do report them in one form or another. These categories serve a number of purposes, such as the provision of information on parameters (e.g. statistical power, the boundaries of sequence durations, the type of stimuli etc.), and on documenting the experimental paradigms used (e.g. prospective and retrospective methods of capturing evaluation, hypotheses made and results obtained).

Readers may note that the timing of the methodology review up to the time of the first study in 1999, and the subsequent publication date in Guyse et al.'s research (year 2002) is not in chronological order. This is because the current methodological review in this chapter, which was originally written in 1999, has been revised and re-written so that it can be incorporated into the tabular form adopted by Guyse et al. (2002). The use of Guyse et al.'s table as a structure in which to summarise the literature does not significantly alter the pre-1999 analyses which led to the first study. The reason for the incorporation of the pre-1999 analyses into the format of Table 2.1 is merely to standardise the presentation of material for the reader, so that the pre and post-1999 literature (which will be discussed later in chapter 4) can be compared and contrasted without the need for further reconciliation on part of the reader.

The categorisation of key constructs from the studies up to 1999 is presented in Table 2.1 according to the following criteria below. A summary of the results from the categorisation exercise is presented in Table 2.2. For the

convenience of referencing Table 2.1, the studies cited in this section will follow the numbering order used in Table 2.1, instead of the conventional Harvard style of referencing used in the rest of this thesis. For example, the citation of Ariely's (1998) paper is referenced as [15]. The reason for the abbreviation of citations is because of the expected high frequency of references that are made to these citations in the ensuing discussions. This necessitated the need for a simpler citation mechanism, which lead to the development of the current citation system, to ensure brevity and to conserve page lengths. Literature reviews on the research topic that do not contain any original empirical study or studies, and unpublished but often-cited working papers (e.g. Carmon and Kahneman, 1993) were excluded from Table 2.1. On this basis, there were 15 studies in the literature that contained at least one original empirical study.

Table 2.2 presents a summary of the methods and results used in every study within the 15 cited research papers from Table 2.1. This is useful because some research papers have more than one study, with different hypotheses, methods, parameters and results. For example, research paper [1] consisted of two studies that deployed different methods to elicit the dependent variable, one based on participants' choices between different options, and the other based on participants' willingness-to-pay (WTP) for a certain option.

2.4.1.1 Participants

Participants were classified on the basis of their employment (students versus non-students) and numbers used per study within each research paper.

2.4.1.2 Stimuli and experimental design

The stimuli and experimental design column in Table 2.1 briefly described the set-up of the studies and stimuli used (e.g. colonoscopic procedures and income distributions), and methods used to elicit the dependent variable (i.e. evaluation of the sequence). There were three reported methods (see Table 2.2), two of which were commonly used, and the other was highly specific to the requirement of the individual experiment conducted. The two most commonly used methods were 'choice' (i.e. forced choice) and 'scale'

(numerical and graphical). Under choice, participants have to evaluate their experiences or reveal their preferences for sequences by choosing between options presented by the experimenter. Under scale, participants usually had to evaluate the sequences based on Likert scales or thermometer scales (e.g. 0 to 100 scales, where 0 is not painful at all, to 100, which is as painful as it gets). The least common method was a 'willingness-to-pay procedure', where participants had to provide information on their assessment of their choices in monetary terms. Another important category here was the evaluation paradigm (prospective versus retrospective). As argued in section 2.3, it was an important issue that needed to be taken into consideration, because there might be a domain where theories developed under a prospective paradigm could not be extended to retrospective paradigm, and vice versa.

2.4.1.3 Duration

The reason for the categorisation of duration was to present material that could be used to provide support for the selection of appropriate sequence duration to use in future experiments. It was expected that this analysis would provide estimates of the upper and lower boundaries of parameters used in the literature, which is useful in avoiding potential ceiling and floor effects when designing experiments. Duration here was defined as the total time of the sequence under evaluation. It ranged from 14 seconds (as in study [15]) to 60 years (as in study [12]), and it was further sub-divided according to whether the evaluation paradigm was prospective or retrospective.

2.4.1.4 Breakdown of the results

The results were divided into five different categories. The 'trend effect' referred to comparisons between sequences with different patterns or trends, such as increasing and decreasing intensity. It also included studies that investigated less commonly observed trends such as increasing then decreasing, or constant followed by increasing (e.g. those used in [15]), or simple contrasts between pairs of events, one with higher intensity than the other (e.g. those used in [1]). 'Velocity' referred to the effects that arise from variations in the rate of progression of an experience. This includes the (quasi) acceleration effects of [9]. The 'peak-end' effect referred to studies that

investigated the effects of the peak (highest intensity) of an experience, and/or its end (the final intensity of the experience). 'Duration neglect' referred to studies that focussed on a special case of the peak-end effect, where for some aversive sequences such as painful medical procedures, people prefer the dominated, over the dominant alternative (i.e. the sequence with a longer duration of pain over one with a shorter duration), as long as the longer sequence ends on an improving note. 'Qualification / Moderation of gestalt effect' referred to studies where the objective was to investigate the impact of various factors on results obtained. Studies that researched on the effects of moderation came under this category.

2.4.2 Results

The results of the categorisation are presented in Table 2.1, and further summarised in Table 2.2. The first column in Table 2.1 presented references to the empirical studies reviewed in chronological order. The second column reported on the number and type of participants used in the study. The third column described the stimuli and experimental design used. Included in this category was the type of dependent variable used, which ranged from choice or rating or other, and whether it the evaluation was prospective (indicated by P) or retrospective (indicated by R). The fourth column indicated the sequence duration, which can range from a few seconds to more than a few years. The fifth and final column summarised the key findings of the studies analysed.

Table 2.1: Synopsis of methods employed and results obtained in gestalt studies up to 1999

Study	Participants	Stimuli and experimental design	Duration	Results
[1] Ross and Simonson (1991)	202 under-graduates	Monetary gains and losses; Questionnaire involves choice, P	Not mentioned	Preference for an improving end
	135 under-graduates	Monetary gains and losses; Questionnaire involves choice, P	One day	Preference for an improving end
	40 under-graduates	Attractiveness of a computer game; Willingness-to-pay (WTP) experimental design, R	Not mentioned	Preference for an improving end
[2] Loewenstein and Sicherman (1991)	80 adults	Wage and rental income streams; Questionnaire involves choice, P	6 years	Increasing trend preferred; Moderated by factors such as education, age and current income
[3] Hsee and Abelson (1991)	48 under-graduates	Academic achievement and income stream; Questionnaire involves choice , P	5 weeks	Satisfaction positively related to velocity (for positive stimuli)
	36 under-graduates	Academic achievement, gambles and shares; Computer generated stimuli; 9 point scale (satisfied to unsatisfied), R	33 to 100 sec	Same as above

Study	Participants	Stimuli and experimental design	Duration	Results
[4] Hsee, Abelson and Salovey (1991)	96 under-graduates	Income streams and academic achievement; Questionnaire design; 9 point scale, P	4 years	Framing, motive and perception of control moderates the effects of velocity
[5] Varey and Kahneman (1992)	48 students	Uncomfortable stimuli (weights, posture, encumbrance, noise); Questionnaire involves choice, P	5 to 50 mins	No effect of trend if it is not made salient
	46 students	Uncomfortable stimuli (vibration, loud noises, standing uncomfortably); Questionnaire with 0-10 scale (no discomfort to almost unbearable), P	15 to 35 mins	Decreasing trend of aversive stimuli preferred; Contrast between highest and lowest intensity has a larger effect on evaluation than duration; Adding a better end improves the overall experience
[6] Loewenstein and Prelec (1993)	Example 1: 95 under-graduates	Restaurant meal; Questionnaire involves choice, P	1 to 2 months	Preference for an improvement
	Example 2: 48 museum visitors	Social activity; Questionnaire involves choice, P	1 to 27 weeks	Same as above

Study	Participants	Stimuli and experimental design	Duration	Results
[6] Loewenstein and Prelec (1993) continued	Example 3 & 4: 27 under-graduates; 51 museum visitors	Same as above	5 weeks	Preference for spreading
	Example 5: 101 museum visitors	Same as above	4 months to unlimited time	Inconsistent time preferences
	Study 1 & 2: 52 & 57 museum visitors	Social activity; Questionnaire involves rating; 1 (worst) - 10 (best) scale used, P	5 weeks	Preference for an improving profile of outcomes
[7] Fredrickson and Kahneman (1993)	32 students & 96 students	Pleasant and aversive film clips; Evaluation is on a moment-to-moment basis; Sliding scale (-7 to +7) used, R	55 to 138 sec	The effects of duration on global evaluations were small, explained by real time affect and reduced when made from memory; Retrospective evaluations governed by peak-end rule
[8] Kahneman, Fredrickson, Schreiber and Redelmeier (1993)	32 students (all male)	Cold water experiment; 11 point Likert scale (-5 to +5), R	60 to 90 sec	Peak-end rule is a good predictor of retrospective evaluations; Adding a better end can improve the overall experience

Study	Participants	Stimuli and experimental design	Duration	Results
[9] Hsee, Salovey and Abelson (1994)	62 students	Academic achievement and US$ bonds; Computer generated stimuli; 7 point rating scale, P	5 months	Satisfaction is positively related to the rate of change in velocity
	163 under-graduates	Share investment; Computer generated stimuli; Evaluation on a moment-to-moment basis using a 7 point scale, P	3 months	Same as above
[10] Schmitt and Kemper (1996)	20 under-graduates	Income streams; Experimental apparatus generated four different trends; Monetary incentives offered, R	50 mins per session over 5 days	In general, improving trend preferred; Strength of preference dependent on the magnitude, velocity and acceleration of improvement
	87 under-graduates	Same as above; Questionnaire used 9 point rating scale, P	No mention of duration	Same as above
[11] Redelmeier and Kahneman (1996)	53 hospital patients under colonoscopy & lithotripsy	Real pain; Moment-to-moment evaluation; 0-10 scale, R	Mean= 23 mins SD= 13 mins	Duration neglect and peak-end effect
[12] Chapman (1996a)	40 under-graduates	Health quality; Questionnaire, P	60 years	Decreasing trend preferred
		Income streams; Questionnaire, P	60 years	Decreasing trend preferred (small)

Study	Participants	Stimuli and experimental design	Duration	Results
[12] Chapman (1996a) continued	50 under-graduates	Health quality; Questionnaire, P	60 years	Decreasing trend preferred
		Health quality; Questionnaire, P	1 year treatment	Increasing trend preferred
	79 under-graduates	Income streams; Questionnaire, P	60 years	Indifferent between trends
		Health quality; Questionnaire, P	60 years	Decreasing trend preferred
			12 day treatment	Increasing trend preferred
		Income streams; Questionnaire, P	60 years	Increasing trend preferred
			12 day income	Increasing trend preferred
[13] Baumgartner, Sujan and Padgett (1997)	28 under-graduates	Advertisements; Computer-generated stimuli; moment-to-moment evaluation; 9 point scale (1= strong negative feelings to 9= strong positive feelings), R	30 to 90 sec	Peak-end effect and preference for improving profile of advertisements
	34 participants (type not mentioned)	Advertisements; Computer-generated stimuli; 9 point scale, R	60 to 90 sec	Results above cannot be attributed to recency effects

Study	Participants	Stimuli and experimental design	Duration	Results
[13] Baumgartner, Sujan and Padgett (1997) continued	94 participants (type not mentioned)	Same as above	90 sec	Adaptation effects result in preference for delayed peaks and high ends
[14] Kahneman, Wakker and Sarin (1997)	682 colonoscopy patients in hospital	Real pain; Moment-to-moment evaluation; 0-10 scale, R	1 min addition of pain	Duration neglect and peak-end effect
[15] Ariely (1998)	20 graduate students & university faculty	Pain from heat; evaluation when trial concludes; 0-100 scale (no pain to pain as bad as it can be), R	14 sec	End effect and trend effect for the latter part of the experience profile
	40 graduate students	Pain from a finger locked in a vice; 0-100 scale, R	10 to 40 sec	Moment-to-moment reporting of affect moderates retrospective evaluations

Notes:

- In the stimuli and experimental design column, ‘P’ denotes a prospective and ‘R’ a retrospective evaluation.

- Some studies (e.g. experiment 2 of [9]) that purport to use retrospective methods (e.g. participants asked to evaluate the stimuli generated by an apparatus) are classified as prospective (i.e. predictive), because the stimuli used in these studies often require that participants imagine their feelings from hypothetical (and not actual) outcomes (e.g. money in [9]).

- Participants who have volunteered for the studies cited in this table either did so on a free-of-charge basis, or otherwise received compensation in the form of course credit or an attendance fee. When participants were offered monetary or other incentives to perform better, this will be explicitly mentioned in the experimental design column.

Table 2.2: Summary and breakdown of the methods and results in Table 2.1 by the number of research papers and citations

Categories		Number	Papers cited (see Table 2.1 for reference number)
1. Method			
Study type	(a) Experimental		
	- Laboratory	8 papers	1, 3, 7, 8, 9, 10, 13, 15
	- Questionnaire or survey-based methods	7 papers	1, 2, 3, 4, 5, 6, 12
	(b) Real-life experiences	2 papers	11, 14
Elicitation of dependent variable	Choice (choices between options presented)	6 papers	1, 2, 3, 5, 6, 12
	Scale (Likert scales and thermometer scales)	12 papers	3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15
	Willingness to pay / accept (WTP / WTA)	1 paper	1
2. Results	a. Trend effect	7 papers	2, 5, 6, 10, 12, 13, 15
	b. Peak and / or end effect	8 papers	1, 5, 7, 8, 11, 13, 14, 15
	c. Duration neglect	5 papers	5, 7, 8, 11, 14
	d. Velocity	4 papers	3, 4, 9, 10
	e. Qualification / moderation of gestalt effect	3 papers	4, 12, 15

Note: Some papers contain more than one study, where each study may be classified under a different category (e.g. [1])

2.4.3 Discussion

The results from Table 2.1 showed that students make up the largest type of participants used in gestalt research. Only studies [2], [6], [11] and [14] employed non-students as participants. In studies [11] and [14], participants were patients undergoing medical treatment at a hospital. One of the reasons for the significant use of students as participants was because the testing of gestalt theories often requires a laboratory-based approach, or a survey-based approach with a sufficiently large sample size to ensure adequate statistical power. From a logistical and cost point of view, the availability of students on a university campus represented a pragmatic, if not perfect, solution to canvassing adequate numbers of participants required for the experiments.

In terms of experimental design, the two most commonly used dependent variable designs were choice and rating. The use of ratings as a method to elicit preferences for sequences was more popular than the use of choice. Choice was more common in sequences with monetary stimuli, whereas rating was more frequently used when the stimuli involved was aversive. The division of choice for monetary sequences and ratings for aversive sequences could be because people often have to make choices when faced with sequences of monetary stimuli (e.g. which pension plan or mortgage to choose). For aversive stimuli, however, it may be more important to elicit the degree to which the aversiveness affects people. This is because people who face aversive stimuli do not usually choose to experience it voluntarily, which means that it is important to elicit the degree to which the stimuli effects people's evaluations of their experiences. As a simple illustration, people do not prefer to experience a sequence of pain, and will try to avoid it if they can. However, if it is unavoidable, people may prefer a sequence of pain that makes them feel the least discomfort, which, according to gestalt theory, is the improving sequence pattern. It is therefore more relevant in this situation to elicit ratings from different sequence patterns than it is to measure choices, because choices may not be available when faced with aversive sequences.

In terms of independent variables, duration and pattern are the most important manipulations for gestalt research. For retrospective studies, durations used tend to be significantly shorter than those used in prospective studies. For example, no retrospective study duration was longer than one hour, presumably because many participants would not be willing or able to commit themselves to unreasonably long laboratory studies. Besides, there was also the issue of fatigue, which puts an upper limit on the duration of exposure to sequential stimuli (unless the study was deliberately designed to test fatigue, which was not the case for any of the studies in the literature). The range of duration used was from 10 seconds (study [15]) to 50 minutes (study [10]), but a large proportion of retrospective studies used a duration not longer than 180 seconds (e.g. [3], [7], [8], [13], [15]). Prospective studies were used when the duration under investigation had to be longer than one hour. Typically, durations range from weeks (e.g. [6], to years (e.g. [2]), to even decades (e.g. [12]).

The discussion in this section can be used to provide a guide to the most appropriate methods to use in gestalt research. The recommendations for forthcoming experiments in this thesis are as follows. It is acceptable for studies in this thesis to use students as participants, because this issue did not raise any major concerns amongst researchers, and the benefits of recruiting students as participants (e.g. easier logistics and lower cost) probably outweigh the costs (e.g. narrow field of participants). Studies that require the use of longer durations should employ prospective methods (e.g. surveys), but those that could afford to use shorter durations should do so, because the dependent variable measured would be derived from actual, not anticipatory, experiences. From Table 2.1, the most common pattern was the improving ending effect, which was found in all studies that examined the impact of trend, velocity and peak-end effects on evaluations. What is common to all pattern effects reported is a sequence classified as improving over time (whether it is defined in terms of its increasing trend, a favourable peak-end ratio, or increasing velocity) is often preferred (with a few exceptions) to one that deteriorated over time. In terms of delay sequences, this finding suggests an improving sequence of delays will be more favourable than a deteriorating

sequence, regardless of whether improvement is defined in terms of its trend, velocity, or whether it has favourable peak-end characteristics.

In summary, this section (2.4) has reviewed the methods and parameters used in the literature. In Table 2.1 and for the discussions that followed it, the studies were categorised based on the type of participants employed, the stimuli and design used, its duration, and results obtained. A guide that was used in the selection of methods and parameters to use in forthcoming experiments in this thesis was also provided.

2.5 Summary and conclusions

There were two broad objectives set out in this chapter. The first objective was to introduce the reader to the topic of internet delays. These delays were analysed by reviewing the extant literature on delays to understand how people evaluate delays and what can be done to mitigate its effects. Given that it was suggested that internet delays were often experienced as multiple delays in a sequence, rather than as a single-period of delay, this meant that it was also necessary to review the literature on gestalt theory. The second objective of this chapter was to review the literature on sequences (i.e. the sequence literature) to provide a theoretical framework for analysing sequences of internet-based delays. The review of the sequence literature was separated into three parts, the first of which presented key findings of the literature. The second part discussed the limitations of the literature, while the final part analysed the methodology of the literature in further detail.

In the review of the delay literature in section 2.1, the broad consensus of the literature was that while delays did impact on perceptions of service quality in a negative way, there were, however, a number of things that service providers could do to mitigate the aversiveness of the experience, such as by providing information on the expected duration of the delay, or by providing appropriate distractions such that participants were less attentive to the passing of time. In extending some of these ideas to the investigation of consumers' experiences with the internet, it was suggested that one of the more unique characteristics

of internet delays was that it was often experienced as a sequence of delays, rather than delays concentrated at one point in a service. This was because internet consumers would move from one webpage to another, which was often the way they would use the internet. It was highly likely that consumers would, in their internet surfing experience, encounter multiple and consecutive delays in between the web pages.

However, because most of the literature on delays reviewed only focussed on situations where the delays were located at one position, it was also necessary to extend the review to include the literature that investigated situations where the delays were spread across the entire consumer experience. In this literature, it was suggested that ideas from the sequence literature could also be applied to modelling consumers' experiences with sequences of delays, given that current theories on how delays should be distributed in a queue did not take into account of consumer's preferences for sequence patterns (Carmon and Kahneman, 1993; Carmon et al., 1995). This then led to the next section (2.2), which was a review of the sequence literature up to the time of the first study (reported in chapter 3), in 1999. This review reported that people's sequential experiences were influenced by a number of pattern effects, such as the trend, velocity, and peak-end moments of the sequence. Section 2.3 discussed the limits to which such pattern effects would apply. The conclusion that can be drawn from this section was that domain related factors such as expectations, and issues relating to the methods used in eliciting the dependent variable can have a significant effect on pattern effects.

The purpose of section 2.4 was to provide a more systematic representation of key methods and parameters used in the literature. A number of suggestions were provided on the choice of methods and parameters to use when designing a study to test hypotheses in gestalt research. The summary table provided in section 2.4 will also provide a common framework to analyse the combined literature when the pre-1999 literature in this chapter is linked up with the post-1999 literature in chapter 4.

Figure 2.2: A historical timeline of the literature up to 1999

Year	Trend effect	Velocity	Peak-end and duration neglect	Qualification / moderation of pattern effect
1991	Ross & Simonson: Money & computer games	Hsee & Abelson: Income & academic grades		
	Loewenstein & Sicherman: Income	Hsee et al.: Income & academic grades		Hsee et al.: Income & academic grades
1992			Varey & Kahneman: Discomfort	
1993	Loewenstein & Prelec: Social events		Fredrickson & Kahneman: Film clips	
			Kahneman et al.: Cold water trial	
1994		Hsee et al.: Money & academic grades		
1996	Schmitt & Kemper: Income		Redelmeier & Kahneman: Pain	Chapman: Income & health
	Chapman: Income & health			
1997			Baumgartner et al.: TV Advertisement	
			Kahneman et al: Pain	
1998	Ariely: Pain	Ariely: Pain		Ariely: Pain

Some studies (e.g. Hsee et al., 1991 and Chapman, 1996a) straddle more than one category.

Figure 2.2 provides an illustrative summary of the historical development of the literature, categorised by its main findings (e.g. type of pattern effect or qualification of such an effect), and by the stimuli used. Figure 2.2 shows that some of the early gestalt research (in 1991) was primarily focussed on studying people's preferences for monetary sequences. The findings for these

studies have over time, however, been extended to include non-monetary stimuli such as social events, health, medical treatments, TV advertisement and painful stimuli. Research was largely limited to defining more precisely the conditions under which the effects occur (Hsee et al., 1991; Chapman, 1996a). Even studies that were specifically set up to test the robustness of such pattern effects, such as Ariely (1998), found that the various forms of such effects were generally replicable (at least for the domain of painful stimuli that he had used in his studies). It is concluded that because pattern effects have a significant influence over the way in which people evaluate sequences of aversive stimuli, therefore it is conjectured that such effects will also have an impact on consumers' preferences for sequences of internet delays. The next chapter (chapter 3) described a study that further developed and tested this proposition.

Chapter 3

The Internet Delay Study

3.0 Introduction

The previous chapter (chapter 2) reviewed the literature on how delays affect consumers' evaluation of services, as well as the sequence literature. It was proposed that the sequence literature on the effects of sequence patterns could be used to model consumers' experiences with internet delays, given that these were often experienced as a sequence of multiple delays, rather than as a single period. The purpose of this chapter is to further develop and test a model of internet delays that incorporated these broad suggestions proposed in the previous chapter. This chapter is outlined as follows. The remainder of this introductory section revisits and expands on some of the key concepts previously reviewed in chapter 2, as these concepts will be used in the development of a model that evaluates consumers' experiences with sequences of internet delays (section 3.1). Section 3.2 reports the methods used in this study, while section 3.3 presents the results. Finally, section 3.4 discusses the implications of the results, and concludes the chapter.

In a study on queuing for a service, described in the previous chapter, Carmon and Kahneman (1993) found that peoples' retrospective evaluations of their queuing experience were almost entirely based on the way it ended, with little consideration for the events leading up to the end. They gave an example of a queuing experience that received a positive evaluation, despite having dissatisfactory evaluations up to just seconds before it ended on a positive note. This meant that it was possible for people to evaluate the longer sequence of queues as one that is more favourable compared to a shorter sequence, as long as the longer sequence has an improving ending. The finding that people cared about the pattern in which delays were distributed in a sequence was then applied by Dellaert and Kahn (1999) to model consumer experiences with internet delays.

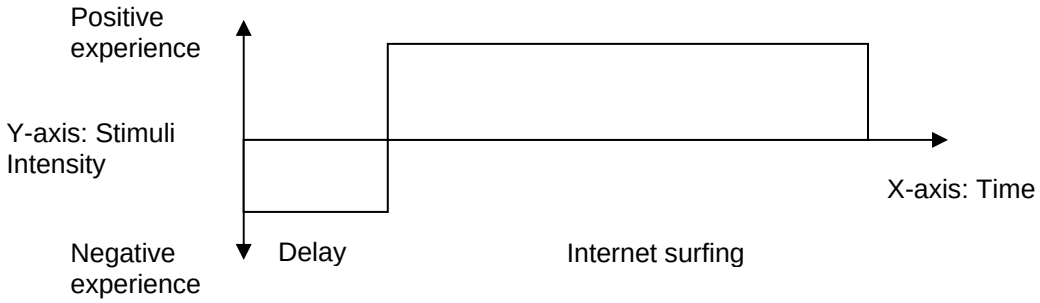
As previously reported, Dellaert and Kahn (1999) had investigated the effects of internet delays on website evaluation in four experiments. The purpose of revisiting Dellaert and Kahn's work here is to provide further details on one of their experiments, which will be used as a basis of the internet delay evaluation model that is being developed in this chapter. The other hypotheses tested by Dellaert and Kahn were not considered, because they were not directly relevant for the theories that would be discussed in this thesis, as they focussed on other aspects of the consumer experience with internet delays.

Dellaert and Kahn's third experiment on the positioning of delays was set up such that participants had to surf a number of web pages on a site that incorporated a delay. They manipulated the position of the delay within the internet surfing experience. In the first of two conditions of a between-participants design experiment, the position of the delay was placed at the start of the surfing experience. For the other condition, the delay was placed in the middle of the surfing experience, thus bisecting it. Dellaert and Kahn also manipulated parameters such as the duration of the delay. Participants were then asked to evaluate their entire web experience using a 10-point response scale, with the categories ranging from very poor to very good.

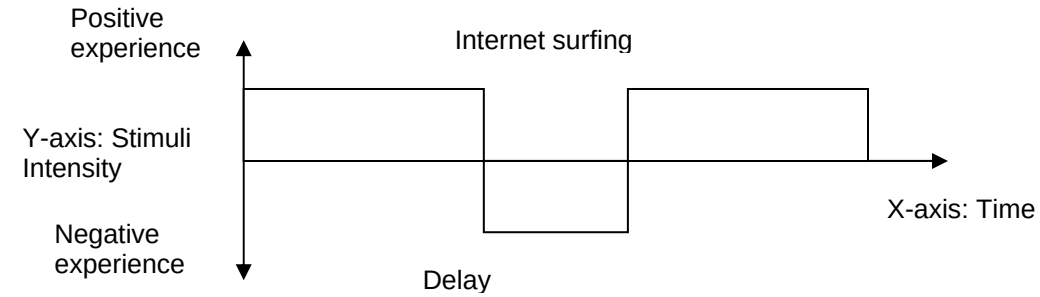
Dellaert and Kahn found that participants preferred the situation where the delay was placed at the start of the surfing experience, compared to the position in the middle of the experience. There are a number of ways in which this finding could be interpreted. Dellaert and Kahn's interpretation was that participants could separate their negative experience of the delay from the positive experience of the website when the position of the delay was at the start, but not so when the position of the delay was in the middle.

Figure 3.1: Reinterpretation of Dellaert and Kahn's (1999) third experiment

Delay at the start



Delay in the middle



Based on prior (non-delay) research, which showed that people relied on patterns to evaluate sequences, it is argued that the way events such as delays are positioned within a sequence, is important, because people attach meanings to the different patterns generated. In prior research, positioning of an event or events within a sequence was shown to be important, because it could invoke gestalt interpretations and evaluation based on patterns. For example, Loewenstein (1987) and Loewenstein and Prelec (1993) demonstrated that people preferred sequences of outcomes where the rewards were postponed (Loewenstein and Prelec used the term 'delayed consumption', instead of 'postponed', but their terminology was renamed to avoid confusion with the context of the word 'delays' used in this study) so that consumption was enjoyed towards the end of the sequence, compared to one where they were enjoyed early on. If Loewenstein and Prelec's theory was applied to re-interpret Dellaert and Kahn's results, the two 'sequences' described in Figure 3.1 (inverted commas are used because Dellaert and Kahn

did not explicitly defined their stimuli as a sequence) would be made up of one with an experience profile negative-to-positive (delay at the start), and the other, positive-to-negative-to-positive (delay in the middle). The preference for postponement would mean that the sequence with the delay at the start would be preferred to the sequence with the delay in the middle, because people would prefer that more of the rewards (i.e. the internet surfing experience) were postponed (or located) towards the end of the sequence.

A point to note is that the duration for which the rewards were postponed had to be within reason, since Loewenstein and Prelec also found that if the postponement was too far into the future (e.g. more than 12 months), people might not consider the events as being part of a sequence. For the purposes of the current discussion, however, this was not an issue, because Dellaert and Kahn's experiment had clearly defined the start and the end of the sequence, i.e. when the website was launched, and when it was shut down. The magnitude of the duration their experiment was in the order of seconds, rather than months.

It is argued in this chapter that Dellaert and Kahn's (1999) work, which had highlighted the importance of the positioning of delays within an internet surfing experience, however, was limited in terms of the scope of its findings. This is because their manipulation was restricted to positioning the delays at a single point within the experience, that is, either at its start, or in the middle. Other patterns of delays, such as one where multiple delays periodically interpose or punctuate the internet surfing experience, were not taken into account in their study. An extension to Dellaert and Kahn's (1999) work, which would investigate the effects of the positioning of delay or delays within an internet experience on its evaluation, is much needed. At the time of the study in this chapter, there was no research using gestalt theories to model consumers' experiences with internet delay(s). As previously argued in chapter 2, the need to research the impact of delays on evaluation has become more important, due to the increasing usage of the internet. In particular, it is suggested that in the arena of internet shopping, delays are expected to occur intermittently throughout the shopping experience, because shopping on a

website often consists of a number of ‘steps’, such as searching for a good or service, followed by refining the search results, selecting the appropriate good or service, entering payment and delivery details, and confirming the transaction. Delays are expected because each of these steps requires information to be transmitted and processed, which is time-consuming.

To summarise, it is proposed that the positioning of delays within an internet experience is important. It is argued that this is especially so for experiences with multiple delays that occurs in sequence, because prior research (in a non-delay arena) has demonstrated that people do take into consideration the sequence patterns that are generated. The next section develops a model of consumer evaluations of internet delay sequences.

3.1 A model of internet delay sequences

The objective of the proposed model of internet delays introduced in this study is to demonstrate that the findings from gestalt theory can be used to predict consumers’ evaluations of internet delay sequences. Because delays were shown to obstruct the consumer experience (Lawson, 1965), it was assumed that a longer delay would be more negative than a shorter delay (Hornik, 1984), because longer delays invoked more negative emotions than shorter ones (Sawrey and Telford, 1971). The sequence literature had shown that for sequences of aversive stimuli, people preferred patterns that improved over time (i.e. where the aversiveness of the stimuli decreased over time). Therefore, it is predicted that a sequence of increasing delays is more aversive than a decreasing delay sequence, because the pattern of aversiveness is increasing in the former sequence, while the pattern of aversiveness is decreasing in the latter. This model will now be developed using notational form.

If we let t= the time period in between two events, and when such time passing is experienced as delays (i.e. time=delays or t=d), and for delay 1 to be longer than delay 2 (d1>d2), then Sawrey and Telford’s (1971) result would give

V(d1)>V(d2), where V(.) is the negative utility generated by d. Larger values of V mean that a larger magnitude of negative utility is generated, and vice versa.

This utility function for the evaluation of delays assumes that there are no other factors that can influence evaluation, aside from the duration of delay. Hornik (1984), however, showed that the provision of information to reduce the uncertainty of the wait could mitigate the negative evaluations of the delayed experience. The literature has also shown that there are other situational factors that can mitigate the effects of delays, for example, Hui and Tse (1996) and Hui and Zhou (1996). In expanding Sawrey and Telford's (1971) model, where the evaluation of delays is a function of the delay duration, to incorporate non-delay factors that can mitigate its effects, let sn be a vector of situational factors, i.e., {s1,s2,…,sn), and let dn be a vector of delays {d1,d2,…,dn}. Therefore, the evaluation or utility from a delayed experience is written as:

V=f(dn,sn)

So, in the case of two different delay scenarios where d1>d2, then V(d1, s1)>V(d2, s2) would hold if the situational factors for both scenarios are identical, i.e., s1=s2, or where s1>s2. If s1<s2, the utility is then determined by the combined effect of d and s, which might moderate the negative effects invoked by the delay. In extension, if d1>d2>…dn, then V(d1, s1)>V(d2, s2)>…>V(dn,sn) would hold if s1=s2=…sn or where s1>s2>…sn.

From Dellaert and Kahn's (1999) findings, one of the situational factors that can affect the evaluation of a delayed experience was the positioning of the delay within a sequence. Let s1= delay at the start and s2=delay in the middle of the experience, and that d1=d2. In re-interpreting Dellaert and Kahn's findings using the current utility model, this gives V(d2, s2)>V(d1, s1). As previously argued (e.g. see Loewenstein, 1987; Loewenstein and Prelec, 1993), positioning is important because people prefer sequence patterns where the more aversive events were placed earlier in the sequence and more favourable events later on in the sequence (i.e. the pattern of improvement).

The model of internet delays described here, however, expands this single period model of delays to include multiple delays. It is argued that it is also important to analyse consumer evaluations of multiple delay experiences. This is because internet surfing experiences will often comprise of multiple sequential delays, in addition to single period delays, because consumers often visit more than one webpage at each experience, thus increasing the chances of experiencing sequences of multiple internet delays. Given that there is no research on sequences of multiple internet delays, it is proposed that theories from the sequence literature can be used to predict consumers' emotional responses towards such experiences.

In a simplification of Kahneman et al.'s (1997) model that was described earlier in section 2.2 of chapter 2, they showed that a sequence of aversive stimuli x1, x2, x3,…, xn (where 1, 2, …, n denote the chronological position of the individual stimuli within the sequence), when evaluated according to the utility integration principle or UIP, would give the following utility function:

$$U(x1, x2, x3,\ldots,xn)=U(x1)+U(x2)+U(x3)+\ldots+U(xn) \quad \ldots \text{UIP}$$

However, the general conclusion of gestalt research demonstrated that the pattern of the sequence does matter in evaluation. For example, Kahneman et al.'s (1997) model suggests that people attach different weights to the individual delays within the sequence, whereas under the UIP, the weights are identical for all individual delays. This is because under Kahneman et al.'s model, certain individual utilities, such as those that form part of a pattern (e.g. peak, end, and trend) are weighed more heavily than others. Therefore, when the individual utilities (which are uneven weighted) are added up together and compared with the UIP utility, Kahneman et al. (1997) has shown that their weighted model predicts people's evaluations more accurately compared to the UIP model.

Incorporating weights into the UIP model above, let:

$$U(x1, x2, x3,\ldots, xn)=w1.U(x1)+w2.U(x2)+w3.U(x3)+\ldots+wn.U(xn)$$

Where w1, w2, …, wn are relative weights. Note that UIP implies that $w1=w2=\ldots=wn$.

If a person has a utility function that weighs the peak-end or trend of a sequence more heavily than other parts, the model above will therefore exhibit a end-weighted bias. An example of a set of end-loaded weights is one with a monotonically increasing pattern or ($w1<w2<\ldots<wn$). In this weight distribution, a sequence with more aversive stimuli located at the end is predicted to have a more negative evaluation compared with a sequence with more aversive stimuli located at the start, assuming that the integral of utility for both sequences are equal. In other words, this set of weights reflects the preference for an improving trend, whereby a sequence with increasingly aversive stimuli will be evaluated more negatively than one with a decreasing pattern. This is because the end-loaded weights disproportionately inflates the influence of the most aversive moment of the increasing sequence (i.e. its end) compared to the most aversive moment of the decreasing sequence (its start).

In addition, there are also other possible patterns of end-loaded weighting, such as a constant pattern followed by an extreme end-weighting ($w1=w2=\ldots=w(n-1)<wn$) or one that combines both a monotonically increasing pattern followed by an extreme end ($w1=w2=w3<\ldots<wn$). Also, it is noted that an end-loaded weight distribution is only one possible distribution of weights. Examples of other distributions include front-loaded weights ($w1>w2>\ldots>wn$) or constant weights.

For the purposes of the theories proposed in this thesis, however, the distribution of weights considered will always refer to a monotonically increasing pattern of weights ($w1<w2<\ldots<wn$), unless the discussion explicitly points out otherwise. The reason for selecting this pattern is because it can represent both the trend and the end effects in sequences, which are among the most common patterns researched in the literature.

Therefore, when a sequence of increasingly aversive experiences with end-loaded weights is compared with a decreasingly aversive sequence with a similar distribution of weights, it is predicted that the increasing delay sequence will be evaluated more negatively (i.e. higher negative utility) than the decreasing sequence. This prediction is represented by the inequality below.

$$w1.U(x1)+w2.U(x2)+ \ldots+wn.U(xn) > w1.U(xn)+\ldots+w(n-1).U(x2) +wn.U(x1)$$

$$\text{or } U(x1, x2, \ldots, xn)>U(xn, \ldots, x2, x1) \ldots 3.1$$

If model 3.1 is extended to sequences of delays, when delays are considered to be aversive (i.e. xn=dn), therefore extrapolating from inequality 3.1 with an increasing ($d1<d2<\ldots<dn$) and a decreasing (dn, …, d2, d1) sequence will give the following inequality, which is also the first model proposed:

$$V(d1, d2, \ldots, dn)>V(dn, \ldots, d2, d1) \ldots \text{[Model 1]}$$

This inequality states that a sequence of increasing delays is more aversive than a sequence of decreasing delays, for delays that are identical in magnitude, but opposite in direction. For example, the sequence (d1, d2, …, dn) is {2,5,8,10} minutes, and (dn, …, d2, d1), {10,8,5,2} minutes. Note that Model 1 is a subset of the more general V=f(d,s) model described earlier. In the case of model 1, the situational factor(s) is the pattern of delays, that is, whether it increases (d1 to dn) or decreases (dn to d1) as the sequence progresses. Based on model 1, the hypothesis for this study is written as follows:

H1: In a comparison between two sequences with identical total delays, the sequence with the decreasing delay pattern would be less aversive than the increasing delay sequence.

In the next section, a study is presented to test the hypothesis made.

3.2 Method

This study utilised these two sequence patterns (increasing versus decreasing delays) as the independent variable manipulation. The experimental domain in which participants experienced these delays was an internet retail site, where they were required to make a number of web-based purchases that were interrupted by delays.

The parameters used in this study were selected to reflect the parameters employed in the literature. Prior to this study, pilot tests were conducted to simulate delays that were in common with real world internet experiences. These tests demonstrated that participants were annoyed by the delays that they had experienced. However, the use of excessive delays in this study was ruled out to prevent participants from getting bored, which may result in poor levels of concentration. The choice of parameters was restricted to durations used in the literature, where the upper limit was approximately 180 seconds, based on the discussion in section 2.4.3 of chapter 2. The dependent variable used (see Figure 3.4 for an illustration) was as follows. At the conclusion of each sequence (the 'evaluate sequence' box in Figure 3.2), participants were asked the question "How annoying were the delays?" They were required to provide their response to this question by circling a number on a ten point scale (with 1 = not annoying at all to 10 = extremely annoying) which best reflected their evaluation of their internet shopping experience. This scale and choice of phrase was developed based on the dependent variables used in a number of studies, including Dellaert and Kahn (1999), Ariely (1998) and Varey and Kahneman (1992), to measure the overall aversiveness of the sequential experience.

3.2.1 Participants

Forty-five undergraduates at a large British university volunteered to participate in a simulated internet shopping study. Their ages ranged from 18 to 25 years old. Each was paid £3.00 each as a participation fee. Twenty of the participants were female.

3.2.2 Stimuli, design and procedure

The experiment required participants to make hypothetical purchases of groceries in a simulated online supermarket, with one purchase per web page, for five purchases that made up a sequence. The five delays were positioned in between each of the five purchases (i.e. pages 1 to 5 in Figure 3.2), and its final page (i.e. the end page). See Figure 3.2 for the layout and flow of the experiment.

The experiment was operationalised as follows. Participants were sat at a computer terminal, and on the 'instructions' page in Figure 3.2, they were told that they had to make a number of purchases on a website. Participants were then taken to the 'familiarisation' page, where they were given the opportunity to undergo a trial run before starting the experiment for real. Once they were satisfied that they had understood the requirements of the experiment, they were told to begin.

The first purchase was made on 'page 1' of Figure 3.2. For an illustration of what the pages looked like, see Figure 3.3. Once the participant had completed the purchase, they were told to click to the next page (which was 'page 2') to make the second purchase. When participants clicked to move to the second page, however, it did not load up immediately, but was interrupted by a delay. Participants were not told how long the delay would last. Eventually, page 2 did load up and participants made a further purchase. This continued until they reached the last page of the sequence (the page titled 'end'). On this page, participants were told that they had successfully completed their shopping task, and were asked to evaluate their entire experience using the rating instrument (i.e. dependent variable) in Figure 3.3.

In particular, participants were asked to circle a number, on a scale of 1 to 10 (1 being not at all, and 10 being extremely) to indicate how annoying the sequence of delays was. Once they had done so, they were then asked to click the mouse to proceed. On the next page, participants were told that they had to make another set of five purchases in sequence. The entire sequential process was thus repeated again (i.e. moving from page 1 to 5 to the end

page), but this time with a different set of parameters for the items to be purchased and for the time delays. Participants, however, were not told what these parameters were, or that these parameters were different from the sequence that they had previously experienced.

Figure 3.2: Schematic layout of the time delay experiment

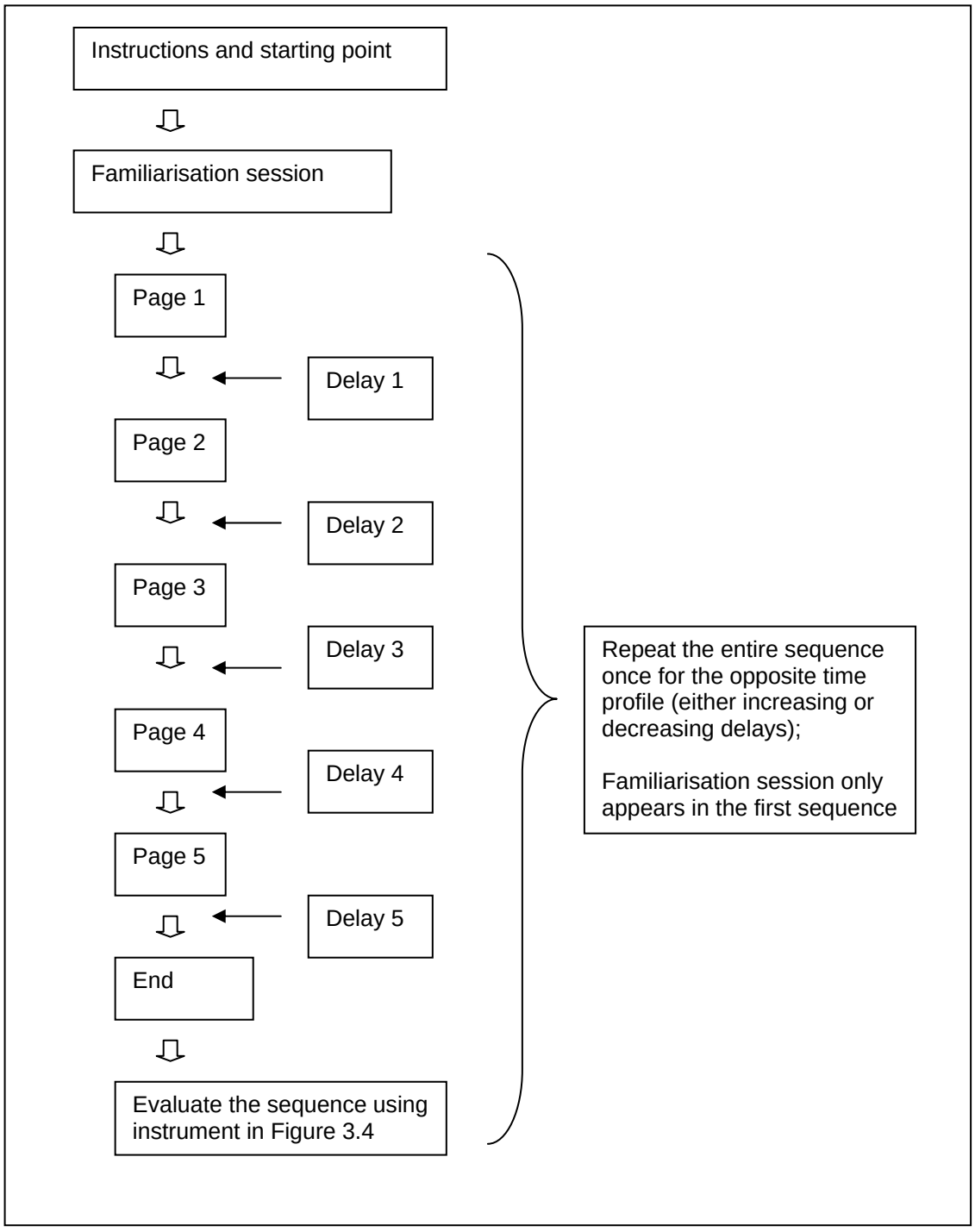


Figure 3.3: Snapshot of one of the web pages used in this study

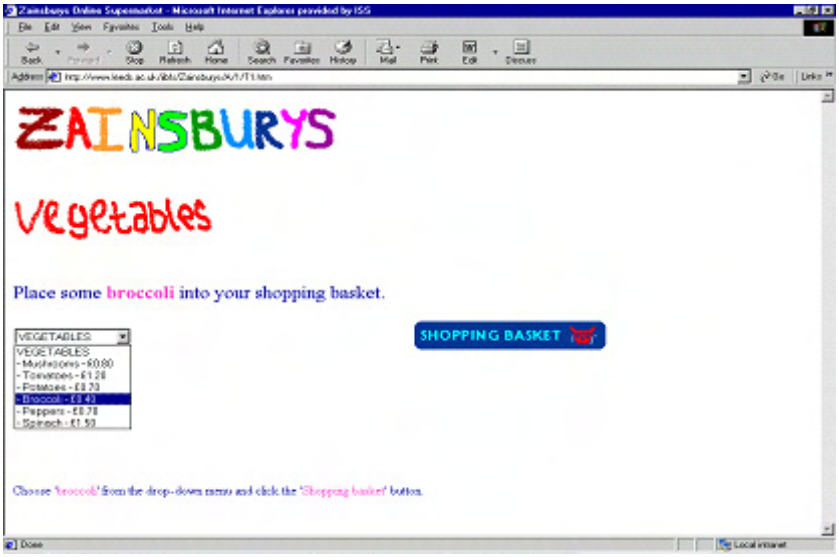


Figure 3.4: Rating instrument used to measure the dependent variable

How ANNOYING were the delays?

Circle ONE number only.

1	2	3	4	5	6	7	8	9	10

Not at all Somewhat Extremely

The parameters of the delays (in seconds) used in this study were as follows. Increasing delays – {2, 8, 18, 32, 50} seconds, and decreasing delays – {50, 32, 18, 8, 2} seconds. The order in which the delays appeared (e.g. increasing delay sequence before decreasing, or decreasing before increasing) was randomly distributed across participants, with twenty-three selected for the former, and the remainder for the latter, order. The delays were generated and controlled by a computer programme written in CGI (Common Gateway Interface) language. These delays were activated when participants clicked the 'next' button to move from one page to the next. If participants made a mistake at any stage during their task, a web-based prompt (written in JavaScript computer language) appeared on the screen to inform them to correct their mistake. None of the participants made any mistakes during the

actual study (all such problems were ironed out during the familiarisation session).

3.3 Results

Table 3.1 presented the results of the study. The means of the dependent variable 'rating of annoyance' were calculated for both sequence patterns, along with standard deviations in parenthesis. In addition, the means are also presented based on the order in which the sequences appeared, that is, whether the increasing delay sequence was experienced before the decreasing, or vice versa. To test for significance differences between the ratings of annoyance for increasing delays versus decreasing delays, a 2 (pattern) x 2 (order) mixed design model was used. Pattern was the within-participants factor that represented the increasing versus decreasing delays. Order represented the counter-balancing factor where the order of appearance of the sequences (i.e. increasing before decreasing, versus decreasing before increasing) was also taken into account. Order here was a between-participants factor.

Table 3.1: Descriptive statistics of the Internet Delay Study

Independent variable	Pattern of delays	
	Increasing sequence	Decreasing sequence
Order:		
Increasing before decreasing	M=5.65 (SD=2.10)	M=5.48 (SD=2.02)
Decreasing before increasing	M=6.41 (SD=2.09)	M=6.09 (SD=2.02)
Combined results	M=6.02 (SD=2.11)	M=5.78 (SD=2.02)

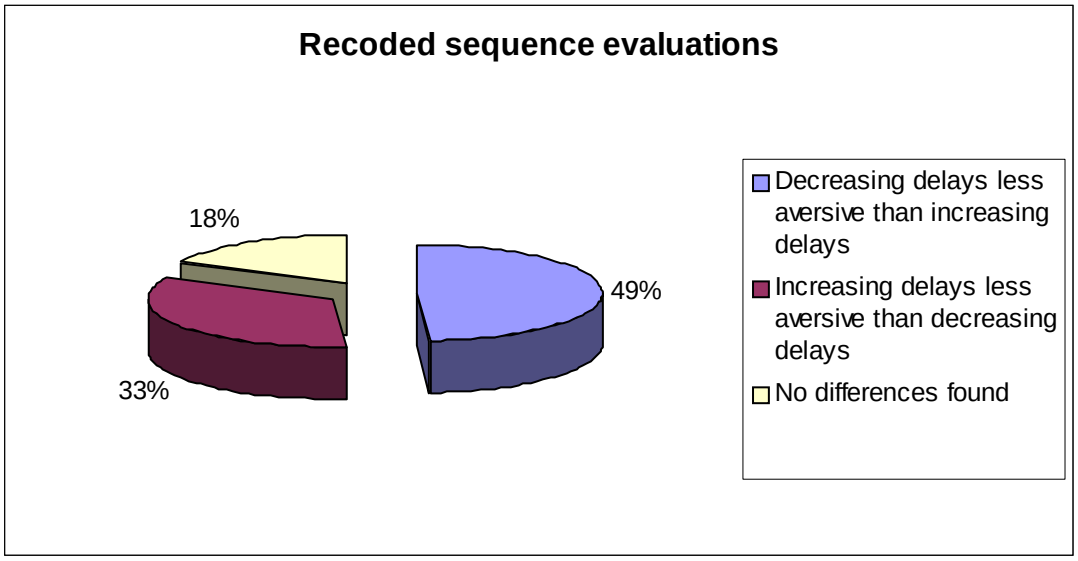
M indicates the mean for ratings of annoyance, while SD indicates the standard deviations for the ratings of annoyance. The scale used ranged from 1 (not annoying) to 10 (very annoying).

An ANOVA calculated for the 2x2 mixed model did not reveal any significant main effect of pattern ($F(1,43)=1.15$, $p>0.05$) nor of order ($F(1,43)=1.45$, $p>0.05$). There was also no interaction effect between pattern and order

(F(1,43)=0.99, p>0.05). The lack of a significant main effect of pattern meant that the hypothesis H1 is not supported.

Given the unexpected nature of these findings, the data reported in Table 3.1 were recoded into different categories. The purpose of this transformation was to provide an additional insight into the data. For example, all participants in this study provided two evaluations, one for the increasing delay sequence, and the other, for the decreasing. The recoding exercise provided the basis for distinguishing those participants who rated the increasing less aversive than the decreasing sequence from those who preferred the opposite, and from those that rated both sequences equally aversive. The data transformation was achieved by converting the scale variables into one of three categories: [category1] for when the decreasing delay sequence is less aversive; [category2] for when the increasing delay sequence is less aversive, the code; and [category3] for equal levels of aversion. The results of the recoding exercise are illustrated in Figure 3.5.

Figure 3.5: Chart illustrating the recoded results of the study



The data in Figure 3.5 showed that 49% (N=22) of participants rated the decreasing delay sequence less aversive than the increasing delay sequence, and 33% (N=15) rated the opposite. This suggested that the distribution of the data was bimodal, because (49%+33%)=82% of participants rated one

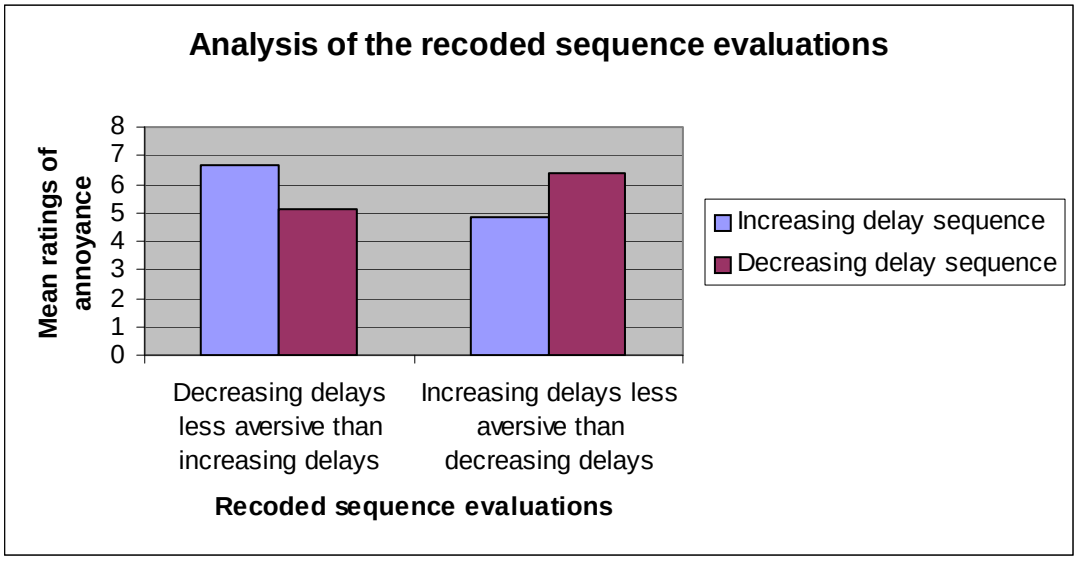
sequence higher than the other, with only 18% rating both equal. Table 3.2 displayed the results of the recoded data by separating the ratings of annoyance for both increasing and decreasing delay sequence according to the direction of preference (i.e. category).

Table 3.2: Selective results of the recoding exercise

	Comparison between sequence evaluations	
	Decreasing delay sequence less aversive [category1]	Increasing delay sequence less aversive [category2]
Ratings of annoyance for:		
Increasing delay sequence	M=6.68, SD=2.03, N=22	M=4.87, SD=1.85, N=15
Decreasing delay sequence	M=5.14, SD=1.96, N=22	M=6.40, SD=1.92, N=15

Note: Mean = M, Standard Deviations = SD, and frequency = N. The scale used ranged from 1 (not annoying) to 10 (very annoying)

Figure 3.6: Chart illustrating the mean rating of annoyance from Table 3.2



There are two possible explanations for the results observed in Table 3.2, which is also illustrated in Figure 3.6. The first argues that the instrument used to measure the annoyance rating is not reliable enough to produce consistent measures. For example, participants might have felt that both sequences were similarly annoying, but may have made a mistake when evaluating their experience using the instrument. This means that the differences between the increasing versus decreasing sequence, when separated by the direction of

preference or category, could be representative of such errors in evaluation, rather than a true bimodal preference. The other explanation argues that there are indeed two groups of participants with different preferences; one evaluating the decreasing less aversive than the increasing delay sequence, while the other, the opposite. From the results of the present study, it is not possible to deduce which of these two arguments is correct. Therefore, the possibility that the manipulations and/or the model developed in the present study may be is dealt with in subsequent studies in this thesis (i.e. chapter 5).

3.4 Discussion

This study presented a model of consumer behaviour with sequences of internet delays in the form of Model 1, which predicted that consumers preferred a decreasing sequence of delays over an increasing one in an internet experience. The study in this chapter extended the single period of internet delay used in Dellaert and Kahn (1999), to a situation where the internet experience consisted of multi-period delays experienced in a sequence. It was argued that the study of multi-period delays in a sequence is important, because this was held to be a more realistic example of real-life internet experiences. The model proposed in this study, Model 1, predicted that the preference for an improving sequence pattern could also be used to model consumers' evaluations of internet sequences of delays. An important assumption made in this model was that gestalt theories on peak-end and trend effects in sequences of aversive stimuli also extended to the domain of internet-based delays. The results of this study, however, did not support the hypothesis made for it. Further analyses of the data revealed evaluations that were bi-directional in distribution, instead of the predicted direction in favour of the decreasing sequence pattern. While some participants evaluated the decreasing delay sequence as being the less aversive of the two (i.e. the hypothesised result), there was also a number of them who had the opposite evaluation, where the increasing sequence was less aversive. The direction of evaluation for this second group of participants suggested that the weights at the start of the sequence might have been heavier than the weights at the end (e.g. w1>w2> …>wn instead of w1<w2<…<wn). However, because concerns over the reliability of measures cannot be ignored, the manipulation and/or the

model used in the present study have to be re-examined in subsequent studies. The current discussion proceeds on the assumption that the observed bimodal evaluations were a true representation of the evaluation of the stimuli used in the present study.

The intended interpretation when Model 1 was developed was for participants to perceive the sequence of delays as being annoying. The literature (e.g. Kahneman et al., 1997; Ariely, 1998) had documented that a sequence of escalating annoyance (i.e. which was hypothesised to be the increasing delay sequence in the context of this study) was more aversive than a sequence of diminishing annoyance (i.e. the decreasing delay sequence). The hypothesised direction was based on gestalt theories (e.g. peak-end and trend) that emphasised the end-loaded nature of these sequences, which was the reason for the asymmetric evaluations.

However, the possibility that preferences were indifferent or perhaps bi-directional was unexpected. While Dellaert and Kahn (1999) had demonstrated that the positioning of a single period of delays in an internet sequence has an effect on how people evaluate the sequence, the results of the present study suggested that the evaluation of multiple delays might be more complex than realised. This is because unanticipated factors might have influenced evaluation. It is suggested that the intensity of the aversion caused by the delays may be a factor that determines the distribution of weights in evaluation. While all participants perceived the two sequences of delays as being aversive overall, it is possible that the aversion caused by the delays is insufficient to induce a more systematic bias, where the sequence evaluation is always based on end-loaded weighting. For example, the stimuli used in the sequence literature involved examples of aversive stimuli where pain (e.g. cold water, colonoscopy, loud sounds, fingers in a vice etc.) is being inflicted on participants, and it is assumed that the attention of participants are highly focussed on dealing with and encoding such pain.

In contrast, it is argued that the study in this chapter had utilised a less severe form of aversive stimuli (i.e., internet-based delays), where perhaps attention

on the aversion caused by the delays is less focussed. This may allow for the possibility that other (unanticipated) factors, such as the domain in which the stimuli are experienced, to unduly influence the distribution of weights of the delays. If the argument that preferences were bimodal is correct, this could explain why a significant minority of participants preferred the front-loaded distribution of weights to the predicted end-loaded distribution, which was reflected in the less aversive evaluations for the increasing delay sequence over the decreasing. Domain-specific effects are discussed in further detail in the sub-sections below. The important issue of how people interpreted the distribution of weights in delay sequences is a core research question that the remainder of this thesis attempts to address. This issue is discussed in greater detail in studies reported in chapters 5 and 6.

3.4.1 Domain-specific effects – The carry-on effect

It is suggested that factors such as the carry-on effect and expectations are some of the ways in which domain might influence the distribution of weights in sequences of delays in an internet retailing environment. This sub-section describes the first of these, while the next sub-section addressed the second.

In a study that examined consumer internet surfing behaviour (which was published after the completion of this study), Menon and Kahn (2002) showed that consumers preferred an internet experience to start off well, because this would encourage them to further engage with the online retailer. Menon and Kahn explained that the preference for a good start was because of the presence of carry-on effects in the domain of internet shopping, where the positive utility generated from the start was retained throughout the internet experience. Such carry-on effects are also common for consumer experiences with bricks-and-mortar shopping, whereby a pleasant start to the shopping experience induced people to continue with their shopping activity (Donovan and Rossiter, 1982; Holbrook and Gardner, 1993; Mehrabian and Russell, 1974).

It is argued that these ideas could be extended to explain the phenomenon observed in the current study on sequences of delays in an internet shopping

experience. It was suggested that the group of participants who had evaluated the increasing delay sequence less aversely than the decreasing sequence did so because a less aversive start to the internet experience was more important for them than a less aversive end (i.e. preference for front over end-loaded distribution of weights). The increased importance of the way a sequence started could be the result of carry-on effects, where the more positive or negative emotions generated at the start would be retained by the consumer throughout the entire experience.

In addition to the (positive/negative) emotions retained or carried-on, another consequence of the carry-on effect, when the experience started positively, is the formation of trust between the consumer and internet retailer. Menon and Kahn (2002) suggested that, as was commonly observed in the bricks-and-mortar world of retailing, trust has to be form early on in the experience in order to prevent negative carry-on effects later. Similarly, they also argued that a good start to the internet experience would allow consumers to rapidly build trust with an internet-based retailer. Applying Menon and Kahn's reasoning to describe the findings in this study, it is argued that a website with minimal delays at the start is less annoying overall compared to one where most of the delays are at the start, because the pattern of delays on the former website would reassure consumers that the website was functioning properly and therefore could be trusted to deliver the goods purchased, thus gaining consumers' trust early on. Analogously, these experiences demonstrate that 'first impressions count', in terms of trust-building between the consumer and the online retailer, at least for some individuals.

3.4.2 Domain-specific effects – Expectations

Another explanation for why some participants evaluated a sequence of delays with a better start less aversively compared to a sequence with a better ending is attributed to consumer's expectations of how sequences should progress. For example, Chapman (1996a) showed that people preferred a deteriorating sequence when they expected such sequences to be more realistic, such as for sequences of long-term health. In applying Chapman's reasoning to interpret the results of this study, it is possible that participants who had

evaluated the increasing delay sequence less aversively than the decreasing sequence expected that the increasing delay pattern to be more in-line with their real-life experiences with internet retailers, compared to the latter sequence. For example, it is possible that some participants perceive the increasing delay sequence to be the more realistic pattern of the two, because they experienced an increasing pattern of processing times (i.e. increasing delays) more often than a decreasing pattern, when purchasing on retail websites. The preference for an increasing pattern of delays could be attributed to the fact that it is expected for retail websites to have such a pattern of delays, which is reflective of the website experiencing greater computational demands (e.g. consumers continually add extra items on the shopping basket etc.) as the consumer progresses to check-out (where the greatest computational demand would be).

Therefore, the evaluation of the delay sequence would be more reflective of consumers' expectations of how a retail website should progress (i.e. they are less annoyed when their experience of delay patterns matched their expectations), instead of their generic preference for a sequence with an improving pattern (i.e. a sequence with decreasing aversion). Some participants may have expected, and therefore preferred, an increasing to decreasing delay sequence, because the most influential factor for sequence preference was driven by expectations that the increasing pattern was the more realistic of the two patterns. The preferences of other participants, however, may not be driven by expectations, but more by their preference for an improving pattern (i.e. the decreasing delay pattern). This suggested that the consideration of the internet shopping domain widens the range of possible factors that could influence preference, from the predicted desire to have an improving pattern in H1, to expectations, which may have accounted for the lack of a significant result supporting Model 1's prediction.

3.4.3 Summary and conclusions

In summary, the prediction that a decreasing sequence of delays would be less annoying than the increasing pattern was not supported by the results of the study in this chapter. In the ensuing analysis to uncover possible reasons for

the results, it was suggested that perhaps people's evaluations of delay sequences is more complex than at first realised. The subsequent analysis suggested that the evaluations were split into two, where one set of evaluations supported the hypothesised direction (increasing evaluated worse than decreasing), and the other in the opposing direction. This observation cannot, however, be concluded from the results of the study, because the possibility of measurement error in the ratings could not be eliminated from the analyses. Nevertheless, assuming that preferences were bimodal, it was suggested that the mechanisms in which domain may have (inadvertently) influenced evaluation was to introduce carry-on effects of a more positive start to the sequence, or introduced expectations that guided people's evaluation of the delay sequences. These suggestions were based on the assumption that the data underpinning the observed bimodal evaluations were reliable.

These two theories proposed (carry-on effects and expectations) suggested that studies on consumer evaluations of delay sequences should take account of the domain in which delays are experienced. Therefore, from the model V=f(dn,sn), where the negative utility from a sequence of delays is a function of both the delay and situational factors, it is proposed that, in addition to pattern, domain is also situational factor that influenced utility from a sequence. From the perspective of the model V=f(dn,sn), the effect of introducing another situational factor (e.g. domain) has a significant affect on how other situational factors, such as patterns, would influence sequence evaluation. For example, the new factor could amplify, or moderate the influence of pattern on sequence preference. The preference for an improving pattern could be very strong, or absent. From an online retailer's viewpoint, the challenge is to identify the factors and conditions that cause the amplification or moderation, because retailers would be able to use the increased predictability in behaviour to enhance consumer experience in general. For example, if the total internet delays could not be avoided, retailers should re-organise the sequencing of the delays within the experience such that the pattern effects, when amplified by the new factor, could be used to minimise the negative emotions invoked by the unavoidable delays.

The importance of the topic of moderation had gained even greater momentum from year 2000 onwards, just after the completion of this study. For example, in 2000 a number of papers were published in the literature which showed that while the many of the pattern effects described in the pre-1999 literature were still relatively robust and replicable, there was, however, some important qualifications to the effects of gestalt patterns over sequence evaluation. In this post-1999 literature, it will be shown that the influence of patterns on sequence evaluation are not necessarily applicable for all types of stimuli and domains under all conditions, and that there are some remaining questions over the appropriateness of methods used to measure the dependent variables. This post-1999 literature is described in greater detail in the next chapter (chapter 4).

Future studies in this thesis need to address the issue of domain in sequences of delays. Given that the results in this Internet Delay Study did not support the proposed Model 1 (where people would prefer a decreasing delay sequence over increasing), the immediate challenge is to design a new and improved study that would constitute a more appropriate test (or rebuttal) of model 1. The first set of studies, reported in chapter 5, investigates the issue of how people interpret the distribution of weights in delay sequences, because it was a core assumption made in Model 1. The second set of studies, reported in chapter 6, explored the effects of domain on people's preference for delay sequences. However, given that many of the ideas underpinning these studies is based on research published post-1999 (in contrast to the original study reported in this chapter), the next chapter outlines the post-1999 research.

Chapter 4

The post-1999 literature review

4.0 Introduction

The objective of this chapter is to review the sequence literature published after the completion of the first study (the Internet Delay Study or IDS reported in the previous chapter) in 2000. The emergence of a number of papers published in 2000 had raise further theoretical and methodological issues concerning the assumed universal applicability of theories developed in the sequence literature. These studies found that there was a variety of factors that moderated or qualified the influence of pattern on evaluation. The topics of moderation and qualification of pattern effects are important for this thesis, because in the discussion of the IDS, it was suggested that the internet retailing domain was a factor that could have influenced the evaluation of internet delay sequences. The rapid evolution of the literature from 2000 onwards means that there is a need for a second literature review to update and extend the findings of the first literature review in chapter 2. The structure of this chapter is as follows. The first part of this chapter, in section 4.1, reviews the post-1999 literature, while the second part (section 4.2) extends the analysis of methods used to take into account of the new literature. This analysis includes the pre-1999 literature reported in section 2.4 of chapter 2 to the most recent literature outlined in the present chapter.

Figure 4.1: A historical timeline of the development of the literature from 1999

Year	Trend effect	Velocity	Peak-end and duration neglect	Qualification / moderation of pattern effect
2000	Matsumoto et al.: Income			Matsumoto et al.: Income
	Ariely & Carmon: Pain		Ariely & Carmon: Pain	Ariely & Zauberman: Annoying sounds & Stock market
	Ariely & Loewenstein: Annoying sounds		Schreiber & Kahneman: Annoying sounds	Ariely & Loewenstein: Annoying sounds
	Chapman: Health			Chapman: Health
2001	Raghunathan & Irwin: Consumer experience		Diener et al.: Life experiences	
2002	Guyse et al.: Health, environment & income			Read & Powell: Income & health
	Read & Powell: Income & health			
2003	Ariely & Zauberman: Consumer experience			Ariely & Zauberman: Consumer experience
				Novemsky & Ratner: Taste
2004	Zauberman & Ariely: Academic grades & factory output			Zauberman & Ariely: Academic grades & factory output
2005			Langer et al.: Income	Langer et al.: Income

Some studies (e.g. Chapman, 2000 and Ariely and Zauberman, 2000c.) straddle more than one category.

4.1 The post-1999 literature review

Figure 4.1 illustrated the historical time-line of the post-1999 literature. When Figure 4.1 is compared to Figure 2.2 in chapter 2, which summarised the themes of the pre-1999 literature, there is a noticeable shift in the focus of the studies from discovering new patterns and domains in which sequence pattern would influence evaluation, towards investigating the factors that moderated or qualified the influence of pattern. This shift is characterised by an increasing number of the post-1999 studies where their objectives straddle the columns on new discoveries and applications of pattern effects with qualifications and moderations of pattern effects (see the first three columns in Figure 2.1 and compare the changes in objectives with those in Figure 4.1). The tables that summarised the key methods and parameters of the pre-1999 studies reviewed (Table 2.1 and Table 2.2 in chapter 2) are extended in the present chapter to include studies from 2000 onwards, in the form of Table 4.1 and Table 4.2, which are presented in section 4.2.

One of the objectives of this extended literature review is to analyse in more detail the implications of the shift on the development of studies in this thesis. The review of this new literature was sub-divided into a number of new themes, to reflect the shift in emphasis of the new literature. The first theme, reported in section 4.1.1, was on methodological issues that lead to pattern effects being moderated. The second theme (section 4.1.2) was on the domain-specificity of pattern effects. The third theme (section 4.1.3) was on the role of cognitive ability in moderating pattern effects, and the fourth and final sub-section (section 4.1.4) provides an overall summary of the post-1999 literature review.

4.1.1 Issues on the boundaries of pattern effects

The post-1999 debates on the boundaries of pattern effects focussed on two theoretical issues. The first was on the definition of what is a sequence pattern (section 4.1.1.1), while the second was on the appropriateness of the dependent variables used to demonstrate pattern effects (section 4.1.1.2).

4.1.1.1 Definitions of what is a sequence pattern, and its associated effect on evaluation

Some authors claimed that the pattern effect, such as the preference for an improving sequence, was defined by the dynamic properties of a sequence, such as its trend and velocity, while others suggested that it was the static properties, such as the start, peak, and end moments of a sequence, that defined the sequence pattern (see Ariely and Carmon, 2003). Under the dynamic definition, it was assumed that there is a relationship between the various moments or static properties of a sequence. Under the static definition, however, only key moments of a sequence were relevant for the prediction of evaluation. In the sequence {8,5,2} for example, static properties were the start {8}, peak {8} and end {2}. Theories such as the peak-end rule (Kahneman et al., 1997) predicted that such static properties would be able to account for most of people's evaluation of sequences containing such properties. Dynamic properties, on the other hand, defined the sequence pattern based on the changes in the static properties over time (Ariely, 1998; Ariely and Carmon, 2000). Therefore, the sequence {8,5,2} had a decreasing trend, because the magnitude of the static properties diminished as the sequence progressed. If each of the moments in the sequence were sufficiently 'spread out' in time, the dynamic or trend effect would therefore disappear, because people would consider the moments to be separated, and therefore 'unrelated' to each other. For example, Loewenstein and Prelec (1993) showed that when people were asked to choose between two sequences of social events, an unpleasant one with an abrasive aunt, and a pleasant one with friends, people preferred the sequence of events aunt before friends (i.e. the improving sequence), if the duration between the two events was only a week apart, but were indifferent between the sequences when the duration was half a year apart. Loewenstein and Prelec suggested that the events in the second example was too far apart to have any meaningful dynamic relationship, which was why people were indifferent to the sequence pattern of events.

The debate between whether patterns should be defined based on their static or dynamic properties remained unresolved in the literature, however. The

issue in contention relates to how well the data fits with the hypothesised pattern. For example, in a study of patients undergoing bone marrow transplant treatment at a US hospital, Ariely and Carmon (2000) reported that both the end moment of the sequence (the sequence being the daily pain ratings generated by a patient undergoing treatment for bone marrow transplant) and the trend were significant predictors of the summary evaluations of the treatment. Ariely and Carmon found that patients generally preferred sequences of pain that improved rather than deteriorate over time. In contrast to Kahneman et al.'s (1997) findings, however, Ariely and Carmon (2000) did not find that the peak moment of the sequence was a significant predictor of summary evaluations. Ariely and Carmon argued that this was because the effect of the peak was moderated by repeated exposure to the treatment, thus making it difficult for patients to distinguish clearly the peak moments of pain in the treatment.

In a study that manipulated the intensity of aversive sounds presented to participants, Schreiber and Kahneman (2000) modified Kahneman et al.'s (1997) original theory of duration neglect to one of additive extension effect (AEE). The reason given by Schreiber and Kahneman (2000) was that some of the earlier findings on duration neglect were highly specific to the domains in which they were examined and may not be applicable to all sequence domains. In addition, they also had to respond to the methodological criticisms put forward by Ariely (1998; see also discussions in the next two sub-sections of this chapter). Schreiber and Kahneman (2000) postulated that AEE is a milder and more general version of the duration neglect effect. In AEE, unlike theory of duration neglect (Kahneman et al., 1997), people do care about the duration of a sequence (i.e. a longer sequence of painful stimuli is worse than a shorter one, e.g. {2,5,8,4} is worse than {2,5,8}), but they do not fully integrate the extension (i.e.{4}) to the sequence, as prescribed by the utility integration principle or UIP proposed in Kahneman et al. (1997). Therefore, this implied that the intensity of the extension is unimportant: In the context of the example above, this means that while people can distinguish between {2,5,8} and {2,5,8,4}, they do not distinguish between the intensities of the extension to the sequence. This means that the summarised evaluation of a

sequence of pain with the pattern {2,5,8,4}, where the last moment {4} is an extension of the sequence, is about the same as the summarised evaluation for the extended sequence {2,5,8,1}. In addition to the AEE theory, Schreiber and Kahneman (2000) also argued that their findings contradicted some of the claims made by Ariely and Carmon (2000) and Ariely (1998), where Ariely and his colleagues found that the dynamic properties such as the sequence trend accounted for more of the summary evaluation of a sequence, compared to static properties, such as the peak and end. Instead, Schreiber and Kahneman (2000; pp.36-7, main text and footnotes 6 and 9) argued that in a re-analyses of Ariely's (1998) data, the peak-end pattern had a higher explanatory power in predicting summary evaluations of sequences compared to the trend pattern proposed by Ariely.

It can be concluded from the discussions in this sub-section that the definitions of what is a sequence pattern remains unresolved, due to ongoing debates on the methodology used to define patterns. There is, however, agreement in the literature that people generally prefer an improving sequence, regardless of how improvement is defined (i.e. in peak-end terms, according to Kahneman and his colleagues, or in sequence trend terms, according to Ariely and his colleagues). The implication for the construction of theory in this thesis is that there has to be careful consideration into how a sequence pattern is used as experimental stimuli. If a sequence of improvement can be defined under both peak-end or trend criteria, this will reduce potential methodological criticisms involving definitions of improvement.

4.1.1.2 Methodological concerns over the way in which dependent variables are measured

In the previous sub-section, the dynamic properties of a pattern were defined as the relationship or changes over time in the static properties of a sequence. This definition, however, assumes that the individual events or moments that make up a sequence are sufficiently proximate to each other for it to be perceived as belonging to a sequence. Loewenstein and Prelec (1993) had demonstrated that moments that are too far apart in time and distance would not be perceived and evaluated as a sequence, but interpreted as individual

moments that are unrelated to each other. For example, having a meal in a Greek restaurant in a year's time maybe too distant into the future to be compared with having a meal in a French restaurant today. Building on this finding that moments have to be sufficiently proximate for a series of moments to be interpreted as a sequence, Ariely and Zauberman (2000 and 2003) argued that the moment-by-moment method of evaluating people's painful experiences, which was used in research by Kahneman and his colleagues (Varey and Kahneman, 1992; Fredrickson and Kahneman, 1993; Redelmeier and Kahneman, 1996; Kahneman et al., 1997) would disrupt the smooth changes between the moments that make up a sequence. This disruption breaks up the sequence to the extent that individual static moments are no longer interpreted as being part of a sequence. This has the effect of moderating the influence of patterns on sequence evaluation.

For illustration, consider the following hypothetical example proposed by Ariely and Zauberman (2000). In the example that they used, they assumed that people would evaluate a sequence of aversive stimuli based on the rule that evaluation = un-weighted average of the mean intensity and the end intensity. The intensity scale used was similar to those previously described, ranging from 1 (not aversive at all) to 10 (as aversive as it gets). They also assumed that each unit of time was equally spaced, i.e. the sequence {2,2,2,2} is twice as long as the sequence {2,2}. Applying the evaluation rule above to the sequence {4,5,6,7,8,9} and {9,8,7,6,5,4} would produce an evaluation of 7.75 (mean=6.5, end=9, giving an un-weighted average of the mean and end intensity =7.75) for the increasing sequence, and an evaluation of 5.25 (mean=6.5, end=4, un-weighted average=5.25) for the decreasing sequence. The difference in evaluation of the increasing and decreasing sequences based on the rule above is 7.75-5.25=2.5 (any positive difference indicated a preference for the decreasing sequence). Ariely and Zauberman (2000) further postulated that when a sequence is disrupted, the disruption would segment the sequence to the extent that evaluation would only be based on the average evaluations of each segment, and not on the overall sequence pattern. This is because each segment is conceptualised as a stand-alone experience, and not as part of the sequence. For example, if the sequence

has two points of disruption, it would be segmented into three sub-sections {[4,5],[6,7],[8,9]} and {[9,8],[7,6],[5,4]}. Each sub-section would have its own mini or nested sequences, with evaluations based on the un-weighted average of the mean and end intensities, giving rise to the nested evaluations of {[4.75],[6.75],[8.75]} and {[8.25],[6.25],[4.25]}.

The overall evaluation of the sequence would then be based on an average of the three segments above, rather than on the un-weighted average of the mean and end. This is because there is no perceived relationship between the three segments (i.e. Loewenstein and Prelec's (1993) concept of spread in sequences, discussed above). The average of [4.75][6.75][8.75] is 6.75 and for [8.25][6.25][4.25] is 6.25, giving a difference between the increasing and decreasing sequence of 6.75-6.25=0.5. The difference between the evaluations in the segmented sequence (0.5) is significantly smaller than the 2.5 difference in the non-segmented sequence. Segmentation here has moderated the effect of pattern on evaluation by reducing the differences between sequence patterns. If the extreme case of segmentation was considered {[4],[5],[6],[7],[8],[9]} and {[9],[8],[7],[6],[5],[4]}, the evaluation, based on the average of the sequences (both=6.5), would show that the more favourable evaluation of the decreasing sequence pattern has disappeared altogether. The implication of Ariely and Zauberman's (2000 and 2003) results demonstrated that the segmentation of a sequence moderated the influence of pattern on retrospective evaluations of aversive stimuli. The segmentation of a sequence has the effect of reducing the more global effect of the sequence trend, which then increases the effects of averaging or moderation (Ariely and Zauberman, 2000). In addition, the segmentation of a sequence at 'strategic' points has the effect of creating a number of smaller sequences. 'Strategic' points are defined as points of local minima and maxima, which create mini or nested sequences with increasing and decreasing trends, which are then evaluated separately from the overall experience. Ariely and Zauberman (2003) found that these individual and separate evaluations have a stronger influence over the evaluation of the entire sequence compared to the global pattern of the larger sequence. For example, a larger decreasing sequence that is segmented in a way that it comprised of three smaller increasing

sequences would be evaluated based on the combined effects of the three increasing sequences, instead of the overall decreasing pattern. The 'strategic' segmentation has the effect of reducing the preference for the decreasing sequence.

In addition to demonstrating the potential disruption and segmentation that could be caused by measuring the dependent variable on a moment-by-moment basis, Ariely and Loewenstein (2000) outlined further circumstances that can lead to a duration neglect effect. They pointed out that there are two mechanisms that influence the relative importance of sequence duration on evaluation. The first mechanism concerns the purpose of the sequence evaluation, distinguishing evaluations made for the purpose of encoding the overall goodness or badness of a sequence, with those made for the purpose of assisting future choices among future sequences. The second hinges on whether the sequence evaluation is comparative or not. Ariely and Loewenstein (2000) argued that people seek to maximise their utility if the purpose of evaluating a sequence is to assist the decision on future choices. They use the example of a person who has to decide whether to repeat a particular holiday experience. However, if the purpose is to encode the sequence (i.e. provide an overall representation of the sequence), they argue that it may be sensible for people to ignore duration in their evaluation. The example Ariely and Loewenstein (2000) used was a person who has to convey the overall impression of their holiday to friends. When people summarise their holiday experience for their friends, they usually highlight the key moments of their holiday, but do not give much details relating to the duration. They are more likely to say that the holiday was great and the highlights for them were the sightseeing, but they are less likely to provide detailed descriptions of hour-to-hour movements.

For the second mechanism, Ariely and Loewenstein (2000) claim that it is difficult for people to judge the duration of a sequence in isolation without direct comparison to other events, much in the same way that people have difficulty trying to evaluate the attributes of new stimuli without the ability to make comparisons (Hsee, Loewenstein, Blount and Bazerman, 1999). To test

their propositions, Ariely and Loewenstein (2000) conducted four studies that orthogonally varied the factors ratings versus choices, and comparative versus separate evaluations. Ratings required participants to evaluate sequences based on a scale that ranged from 0-100, which participants have to use to represent variations in the intensity of a stimuli, from least to most intense (e.g. for pain, not painful at all to as painful as it gets). For choice, participants' evaluations of sequences were elicited by getting them to choose between different options. They showed that the importance of duration in evaluation increased when the evaluations were comparative rather than separate, and when evaluation was based on choices rather than on ratings. The implications of Ariely and Loewenstein's (2000) finding for the general applicability of duration neglect is to demonstrate that some of the effects reported in the pre-1999 literature may be suspect, due to the lack of a more careful consideration of the measurement methods used to demonstrate duration neglect.

To summarise, section 4.1.1 discussed methodological issues on the boundaries of pattern effects, which was divided into two dimensions. The first dimension described the literature that debated the definitions of what is a sequence pattern, while the second dimension debated the appropriateness of the dependent variables used to demonstrate the pattern effects. On the first dimension, the definition of what is a pattern can broadly be divided into two, based on whether one thinks that the effects are based on static features (e.g. peak-end theories of Kahneman and his colleagues), or the dynamic characteristics (trend and velocity theories of Ariely, Loewenstein, Hsee and their colleagues) of a sequence. On the second dimension, Ariely and his colleagues questioned the method in which peak-end theories were developed. For example, Ariely and Zauberman (2000 and 2003) argued that the moment-by-moment measurement used (e.g. by Fredrickson and Kahneman, 1993; Kahneman et al., 1993; and Redelmeier and Kahneman, 1996) interfered with the sequential experience, to the extent that it segmented the sequence into smaller sub-sequences, resulting in an evaluation that may not reflect a summary of the overall sequence. In addition, Ariely and Loewenstein (2000) argued that the methods used to demonstrate duration

neglect was also inappropriate, because it did not take into account of the variations in people's interpretation of the dependent variable. People are more sensitive to duration (i.e. less chance of observing duration neglect) when asked to make future decisions based on utility maximisation, and when there is a basis on which to make comparisons. They are, however, less sensitive to duration when asked to encode the overall goodness/badness of a sequence, and where no comparison is possible (which was the method Kahneman and his colleagues used).

4.1.2 Domain-specificity of pattern effects

In addition to the growing literature on methodological debates over the boundaries that moderate the influence of pattern on sequence evaluation, there is also an emerging literature that explores the role in which different domains affect the influence of patterns, either through moderating or amplifying their influence. The sub-sections below describe the various mechanisms that explain why and how different domains affect the influence of sequence pattern on evaluation. Section 4.1.2.1 describes the literature that attributed domain-specificity of pattern effects to people's expectations of how sequences should progress. Section 4.1.2.2 expands the discussion on expectations to include arguments on ideal consumption and appropriateness. Finally, section 4.1.2.3 argues that the effects of domain on the influence of pattern on sequence evaluation vary according to people's motives for experiencing a sequence.

4.1.2.1 The role of expectations

In the earlier literature review in chapter 2, Chapman (1996a) found that while people preferred an improving sequence of monetary outcomes, this preference for improvement was moderated for long-term health sequences. She argued that the reason for the moderation was because people's preferences for health are driven by expectations of how such sequences should be evaluated. Expectations serve as a reference point and are derived from personal observations of real-life sequence progression. This means that any sequence not having the same pattern as the expected sequence would then consist of both positive and negative deviations from the expected pattern

(i.e. gains and losses). Chapman (1996a, para.1, p.204) argued that when comparing an expected with an unexpected sequence pattern that is mean preserving (i.e. the mean intensity for the unexpected sequence is the same as the expected), people would prefer the expected, because the choice of the unexpected sequence pattern would constitute a net loss, as negative deviations from the expected pattern (i.e. losses) loom larger than positive deviations (gains), assuming a prospect theory value function (Kahneman and Tversky, 1979).

In a more recent paper on the same topic, Chapman (2000) explored in further detail the reasons behind people's preferences for health. In her study, Chapman (2000) manipulated two independent variables, the health domain (which was further divided into sub-domains such as headaches, athletic ability, facial acne and facial wrinkles), and the duration (from one hour to 20 years) of the sequence. She reported a significant effect of domain, sub-domains, and duration on expectations and preferences for health sequences, and that preferences for these domains of health sequences generally tracked expectations. For headache sequences, people expected and preferred improving patterns regardless of duration, but for athletic ability, the preference and expectation of an improving sequence pattern declined (i.e. it was moderated) when the sequence duration increased. There was also a stronger preference of improvement for acne sequences compared to wrinkles, and that the difference in expectations between acne and wrinkle sequences was largest for the longest (i.e. 20 year) duration. The general implications of Chapman's (2000) findings demonstrated the highly conditional nature of sequence preference on the stimuli domain (e.g. health versus money), and that even within a domain (e.g. health), significant differences in preferences and expectations exist.

4.1.2.2 Ideal consumption and appropriateness

To determine the reasons why people have a preference for sequence patterns in the domain of money and health, Read and Powell (2002) conducted a study that used a verbal protocol technique to get people to express their thoughts aloud. This technique was used because they wanted

to obtain the widest possible range of responses. For money sequences, Read and Powell found that the most frequent reason for people's preference for sequences of money and health was that it should reflect their ideal pattern of consumption. This is in addition to rational economic reasoning, where people preferred a pattern because it provided the maximum returns, subject to any behavioural constraints (e.g., a less optimal pattern may be chosen if it aided self-control). In other words, people preferred a sequence pattern of health and money that was 'appropriate' for the personal and societal values that they hold. Read and Powell gave the following example 'if people wanted a lavish retirement, they would prefer an income sequence that was end-weighted; if they needed the money to pay off debts quickly, they would prefer a front-loaded sequence pattern of income'. These preferences for sequence patterns, which match the consumption of money with the receipt of money, were demonstrated to be more important to people, as compared to a pattern that only maximises present value, but not does not match income with consumption.

Read and Powell added that the relationship between expectations and appropriateness is close, where people think what they expect is appropriate, but will prefer another pattern if they think what they expect is inappropriate. For example, people may generally expect that income would increase over their lifetime. If they were to contemplate purchasing a house, they would prize the stability of income over their lifetime, in order to know that they can meet their long-term mortgage repayments. Therefore, an example of an appropriate pattern of income for the potential house purchaser is an increasing sequence, or at least one with very stable characteristics. However, if people expect their income sequence to have another pattern that did not match with their ambitions (e.g. free-lance musicians may have sporadic and fluctuating income, depending on the number of, and timing for, the gigs that they perform, both of which can be highly unpredictable), they would not prefer this expected pattern, but instead prefer the appropriate pattern (i.e. the increasing or stable pattern). Read and Powell's study explained the mechanisms for why and how sequence domains influenced people's

preferences for sequence patterns, which is through their expectations and idealised consumption (i.e. appropriateness) of it.

In addition to health and money, expectations also feature in other domains (e.g. advertising, consumer products, and the environment) as a key explanation for people's domain-specific preferences for sequence patterns. The mechanisms in which expectations influence preferences are different from those proposed by Chapman (2000) though. In a study on the impact of sequence patterns on the domain of the marketing of consumer products such as holidays, Raghunathan and Irwin (2001) demonstrated that consumer satisfaction with products was dependent on the domain in which they experience the sequence of products. In their study, they asked consumers to evaluate a particular holiday (e.g. going to Thailand) after they were shown a holiday brochure containing other possible holiday destinations around the world that was increasing versus decreasing in pleasantness. The consumers they asked rated that they were less satisfied with choosing the Thai holiday when the brochures was an improving sequence, but more satisfied when the brochures was a deteriorating sequence. Raghunathan and Irwin attributed their findings to differences in expectations of the potential purchase of a holiday, as compared with the raised expectations caused by the increasing pattern, and by the lowered expectations caused by the decreasing pattern. They argued that consumers' raised expectations from the increasing pattern and the lowered expectations from the decreasing pattern is attributable to their reliance on sequence patterns to form their expectations and reference point in which to compare the product with. Raghunathan and Irwin's results suggest that consumer's expectations of the relative attractiveness of products are highly dependent on the reference frame, which in turn is dependent on the background sequence pattern.

People's reliance on their expectations to construct their preferences, however, can sometimes be faulty, and may lead them to erroneously prefer or choose sequence patterns that may not maximise their actual enjoyment. For example, in a study that examined how people order the sequence of their consumption of consumer food products (jelly beans), Novemsky and Ratner

(2003) showed that people's expectations can sometimes (mis) lead them into thinking that their preference for an improving sequence would actually provide more enjoyment than preferring other sequence patterns. In other words, people (wrongly) over-estimate the extra utility that they could gain from choosing an improving pattern, because actual evaluations of their consumption experience did not support such an expectation that the improving pattern would be indeed the best pattern for them.

Research in other unique sequence domains, such as the quality of life, and the environment, also supported the general finding that preferences for sequence patterns are conditional on the domain. This is because in such domains, people attribute meaning to particular sequence patterns, which may not apply to other domains. For example, Diener, Wirtz and Oishi (2001) found that, in the domain of the quality of a life, the way a life ends has the largest influence over its overall evaluation (i.e. people have a strong preference for a pattern of an improving end to their lives). The importance of the end moment is such that duration (in this example, duration is the number of years that a person lives for) may sometimes be less important in the evaluation of one's life, especially if it means that they get to enjoy a more favourable end (i.e. period leading up to death). Diener et al. compared this duration neglect effect to the 'James Dean' effect, because James Dean had a short but wonderful life up to the point of his untimely death. Therefore, for quality of life sequences, the end moment of a pattern is most important in evaluation because it encapsulates the meaning of the life that one has lived.

For environmental sequences, however, Guyse, Keller and Eppel (2002) found that people generally preferred an improving (increasing) sequences of air quality and near-shore ocean water quality, but did not offer any reasons as to why such patterns are preferred, except to say that such preferences were specific to the domains, when compared with their other findings where people preferred decreasing sequences of money and health. Guyse et al. added that for health, preferences tracked expectations (as per Chapman, 2000), but they did not find any relationship between preference and expectation for the domain of environment and money. Guyse et al.'s results were, however, in

contrast to the studies reported in the literature review above, which showed that people generally preferred deteriorating health sequences (Chapman, 1996a & b) and increasing income sequences (Loewenstein and Sicherman, 1991). Guyse et al. accounted for the differences in their results from that of the general literature by saying that they had recruited more mature business graduate students as participants, who had different perspectives in allocating risks and economic benefits (Keller and Sarin, 1995), when compared to the younger psychology students used by Chapman (1996a & b), and Loewenstein and Sicherman (1991). This suggested that the domain-specificity of preferences could also be attributed to the age and experience of the evaluator.

To summarise, the literature discussed in this sub-section had shown that people's evaluations of sequences could be driven by the values that they hold, such as their notions of what constitutes an ideal consumption, and what is an appropriate preference for a sequence pattern, after consideration of the prevailing circumstances. The evaluations of sequences based on these values were not infallible though.

4.1.2.3 Motivation

In addition to the effect that expectations and ideal consumption (which was previously discussed in sub-sections 4.1.2.1 and 4.1.2.2) have on sequence preference, another reason for why people's preferences are domain-dependent is related to their motive(s) for experiencing the sequence (Hsee et al., 1991). For example, when people pursue an experience such as a holiday, they are doing it primarily for consummatory reasons. When people have to perform a task such as assembling devices in a factory, their motives are primarily instrumental. Hsee et al. (1991) reported that pattern effects were stronger when people's motives for experiencing a sequence are consummatory, compared to when their motives are instrumental.

In a more recent study on the effects of motivation on the influence of pattern on sequence evaluation, Zauberman and Ariely (2004) provided further insights into the psychological mechanisms behind consummatory versus

instrumental motives on pattern effects. They found that consummatory sequences triggered an evaluation that is forward-looking, thus increasing the weight of the trend (i.e. preference for an improving pattern) on sequence evaluation. In contrast, instrumental sequences triggered an evaluation that is backward looking, which increases the weight of the initial part of the sequence. The implications of Hsee et al. (1991) and Zauberman and Ariely's (2004) findings suggest that people's motives for experiencing a sequence is dependent on the domain in which the sequence is experienced. The relationship between motivation and cognitive constraints would be discussed in the next section.

To summarise section 4.1.2, the domain-specificity of pattern effects was categorised into three different dimensions, which explained why and how different domains affect the influence of pattern on sequence evaluation. The first dimension (in section 4.1.2.1) was expectations, which showed that people prefer the expected pattern over the unexpected because of the net loss of the unexpected (Chapman, 1996a). Expectations are formed through personal and societal observations of sequences, which means that an improving sequence pattern will not always be preferred for all domains (Chapman, 2000; Diener et al., 2001; Guyse et al., 2002). The second dimension (in section 4.1.2.2), ideal consumption and appropriateness, argued that, in the context of monetary sequences, people's preferences are driven by their expectations. If, however, such expectations are inappropriate, people will then prefer the appropriate pattern (Read and Powell, 2002). The construction of preferences based on expectations and appropriateness, however, is not infallible. People sometimes do not sufficiently adjust for elevated or lowered reference points caused by the patterns (Raghunathan and Irwin, 2001). In addition, reliance on expectations as a basis on which to construct preferences may result in the over or under-estimate of the actual strength of the pattern effect (Novemsky and Ratner, 2003). The third dimension (in section 4.1.2.3), motivation, explained that people's preferences vary according to the sequence domain because their motives in each domain are different. Some domains are experienced primarily for instrumental reasons, which mean that the preference for improvement is less strong. Other domains are experienced for

consummatory reasons, where the preference for improvement is stronger (Hsee et al., 1991). Zauberman and Ariely (2004) attributed the difference in preferences to the evaluation process, where instrumental motives trigger backward evaluations, emphasising the start of a sequence (i.e. a primacy effect), while consummatory motives trigger forward evaluations, emphasising the trend/end of a sequence.

4.1.3 The role of cognitive ability in moderating pattern effects, and its relationship with motivation

Ariely and Carmon (2003) argued that one of the reasons why people rely on sequence patterns for evaluation is due to people not being able or want to process all information when evaluating experiences. The consideration of all information is a pre-requisite if people were to evaluate sequences based on the utility integration principle (Kahneman et al., 1997). The alternative was to rely on patterns as a heuristic to assist sequence evaluation (Kahneman and Frederick, 2002; Kahneman, 2003). Ariely and Carmon (2003) argued that the use of patterns as a form of heuristic was a reasonable and highly adaptive way for people to summarise large amounts of information relating to their experiences, without having to invest in mental effort where the costs may outweigh the benefits of doing so. Ariely and Carmon added that their argument was supported by evidence from research on how people summarise information based on key attributes, which was also more generally related to the adaptive/constructive perspective of decision-making (Payne, Bettman and Johnson, 1993). For example, in a study involving the recognition and memory of circle sizes, Ariely (2001) found that people only extracted and retained a few key statistics (such as the mean and variance of the circle sizes) that were subsequently used to represent the entire set in a memory-based task.

The patterns-as-heuristic perspective was supported by studies that examined the preferences of more experienced decision-makers versus those with less experience. In a study on preferences for monetary distributions over time, Matsumoto et al. (2000) found that participants who had more financial knowledge (senior year under-graduate accounting students familiar with time

discounting techniques) were more likely to use rational methods (such as Present Value calculations that involved concepts of time discounting) to choose between various monetary distributions, compared to less knowledgeable participants (sophomore or second year undergraduates with little or no exposure to time discounting techniques). Matsumoto et al.'s results showed that knowledgeable participants were able to distinguish between patterns of money that produced the highest present value, even if they had a preference for improving (i.e. increasing) sequences which do not necessarily maximise present value. In contrast, less knowledgeable participants based their decisions primarily on their preference for an improving (i.e. increasing) pattern. Matsumoto et al. reported that their less knowledgeable participants were unaware that the sequence with the decreasing pattern produced a greater present value, compared to the increasing sequence.

Matsumoto et al.'s results, which showed that not all knowledgeable participants prefer the sequence with the highest present value, suggested that people can sometimes explicitly prefer patterns of monetary sequences that may not be normatively optimal, but possess psychologically attractive features (Loewenstein and Sicherman, 1991; Read and Powell, 2002). For example, O'Donoghue and Rabin (2000) reported that people prefer certain sequence patterns over others because they use them as a tool to reflect their self-control – the preference for an increasing sequence of rewards over a decreasing one is a form of delayed gratification (Loewenstein, 1987). In another study that also investigated people's preferences for sequences of monetary outcomes, Langer, Sarin and Weber (2005) showed that people's reliance on sequence patterns for evaluation was dependent their ability to work out the normatively optimal choice between monetary sequences. Langer et al. found that their participants relied more on patterns such as the peak-end for evaluation when they were unable to work out the normative solution, because the tasks that they were asked to perform were beyond their abilities to cope with them. However, when the task difficulty was within their capabilities, they were able to make choices based on the normative solution that they had worked out.

Other (non-sequence) studies that investigated the relationship between the decisions made in situations where people's cognitive abilities were not exceeded also supported the findings of Matsumoto et al. (2000) and Langer et al. (2005). In a series of studies based on actual market trading environments (the sports card market in the USA), List (2003, 2004) found that behavioural models, such as prospect theory (Kahneman and Tversky, 1979) and the endowment effect (Kahneman, Knetsch and Thaler, 1990) had strong predictive power for novice or inexperienced decision-making, where the normatively optimal solution may be difficult to determine. However, List found that this behaviour did not extend to the decisions made by highly experienced traders, whose trading behaviour was more in line with the predictions of normative principles. List (2004) added that people have the ability to overcome their lack of experience to make more rational decisions, which is contrary to evidence in the psychological literature that claims the limited transfer of learning across tasks (Loewenstein, 1999b).

A distinction, however, has to be made between domains where people have to choose between sequences where a normatively better outcome does exist (e.g. monetary sequences; pain etc.) versus domains where the argument for preferring a normative outcome is less compelling (e.g. health, cinematic experiences, advertising). The results from Matsumoto et al. (2000) and Langer et al. (2005), which argued that people rely on patterns for evaluation because of cognitive constraints, would only apply to sequences where there is a compelling case for adopting normative principles to preferences in the first place, and where the benefits of choosing a less optimal but more psychologically attractive option (e.g. choosing distributions that most closely matches receipt of income with its consumption – see Loewenstein and Sicherman, 1991; Read and Powell, 2002) is not outweighed by the costs of doing so.

For example, the case for using normative principles (such as maximisation of present value) in instrumental sequences, such as monetary distributions over time, is a more compelling argument than for consummatory experiences (e.g. cinematic experiences, advertising etc.), because the objective of an

instrumental experience is to maximise the outcome of the experience, and where the journey or the path taken to reach that outcome is ancillary to the evaluation. Therefore, for instrumental experiences, it is argued that any preference for an outcome that is non-optimal from a normative perspective can be attributed to bounded rationality constraints. It is observed that almost all of the research on preferences for monetary distributions have not recruited real experts (see the analysis of types of participants used in Table 4.1 in section 4.2.2.1), such as finance professionals (e.g. accountants, investment bankers, money managers, stock traders etc.), who presumably are less hindered by the bounded rationality constraints (i.e. by choosing sub-optimal monetary distributions due to a lack of will power and/or self-control, inability or lack of incentives to calculate the optimal distribution profile, endowment and framing effects etc.), compared to less experienced decision-makers, such as sophomore undergraduate students.

For consummatory experiences, however, because utility is derived from the journey or the path of the experience itself, in addition to the final outcome, it is more difficult to argue that people should prefer experiences that are normatively superior from the economic viewpoint of present value maximisation (Ariely and Loewenstein, 2000). Ariely and Loewenstein (2000) argued that while normative rules of choice such as dominance (if X is better than Y on all dimensions, choose X) and transitivity (if X is preferred to Y, and Y is preferred to Z, then X is preferred to Z) may be a useful principle to follow in economic decision-making, it, however, makes less sense to use such principles for the evaluation of consummatory or emotional experiences. In such experiences, the literature (e.g. Hsee et al., 1991) had suggested that the pattern of an experience – in addition to the mean or integrated utility of a sequence – is also an important consideration in evaluation. Consider the examples of people pursuing experiences such as mountain climbing, planting and caring for rare flowers that bloom only once every 20 years. It could be argued that the average evaluations of such experiences is less meaningful compared to instrumental experiences, since the key utilities derived from mountain climbing is the pleasure from ‘re-living’ their tales of hardship (Loewenstein, 1999a), where key moments of the experience such as the

peak-end intensities feature more highly in memory (Fredrickson and Kahneman, 1993), and are intuitively more meaningful compared to the average evaluation when relating experiences to others. The same could also be said of the gardening example provided, where the evaluation of joy, happiness, and elation etc. of the experience hinges almost entirely on the final moment, when the flower blooms. The retrospective evaluation of the 20-year experience based on its average is obviously less meaningful here compared to its peak-end evaluation. All these examples suggest that the meanings conveyed by the gestalt characteristics of an experience are carriers of meaning, which are not captured by normative rules such as the utility integration principle and maximum present value (Fredrickson, 2000; Diener et al., 2001).

In summary of section 4.1.3, the differences between instrumental versus consummatory experiences implied that the psychological mechanisms underpinning the ability of patterns to influence the evaluation of sequences are different. Cognitive ability moderates the influence of patterns for evaluations of instrumental sequences, but for more consummatory sequences, factors such as expectations and appropriateness are the primary determinants of the influence of pattern on evaluation. Normative rules of evaluation are less compelling for consummatory compared to instrumental experiences, because there is no optimal choice of pattern for such consummatory motives, as it would all depend on people's individual differences and the societal values that they hold, which in turn shape their preferences and choices for sequences.

4.1.4 Summary of the post-1999 literature

The purpose of this literature review in section 4.1 was to extend the previous pre-1999 review reported in chapter 2 to incorporate new developments since 1999. These new developments reflected an increased emphasis on qualifying the extent to which patterns would influence sequence evaluation. The literature review was also motivated by the discussion of the findings of the Internet Delay Study (IDS) in chapter 2, which suggested a number of reasons to explain the moderation of the pattern effects in that study. The

post-1999 literature was organised into three inter-related themes (sections 4.1.1 to 4.1.3), which highlighted the factors that may qualify/moderate the pattern effects.

The first theme (section 4.1.1) highlighted issues concerning the boundaries of pattern effects, such as sequence pattern definitions and methods of eliciting preference. It was reported that some of the methodologies used in the pre-1999 literature were problematic, so the findings and interpretations need to be re-examined in more detail, because there were a number of weaknesses that may lead to inaccuracies in the theories developed. The literature consisted of two themes, the first of which was on the definition of what a sequence pattern is (static versus dynamic features), and the second on the appropriateness of methods used to measure the dependent variables. It was shown that for studies that used experimental methods such as moment-by-moment evaluation, methods interfered with the sequential experience. In addition, people's intention or purpose in evaluating a sequence also matters. Their sensitivity to the duration of the sequence increases if the purpose of their evaluation was to make future decisions. However, sensitivity to duration decreases when the purpose is to summarise the overall goodness/badness of the experience.

The second theme (4.1.2) described some of the mechanisms that demonstrated the domain-specificity of pattern effects. In many domains, preferences are driven by expectations, which are in turn derived from people's personal experiences with sequences. Sometimes people may not prefer their expected outcome, especially if it seems inappropriate for them to follow their expectation (e.g. see example given in section 4.1.2.2). The literature has also reported that preferences that are constructed from expectations may not always lead people to preferring the option that maximises their enjoyment. This is because people over or under-estimate the strength of the predicted versus the actual experience. In addition, motivation also explains why preferences are domain-specific. This is because some sequential domains are predominantly instrumental in nature, which means that pattern effects

such as the preference for improvement is less strong, while other domains are predominantly consummatory, which enhances pattern effects.

The third and final theme (4.1.3) proposed that cognitive ability plays a role in moderating pattern effects for some sequence domains. This perspective takes the view that sequence patterns are a type of heuristic that people use to summarise their experiences. Pattern effects become more influential on evaluation when people do not have sufficient ability and/or knowledge to evaluate a sequence based on normative principles. This argument would only be applicable for sequences that are primarily instrumental in nature, where a normative solution does exist. For sequences that are more consummatory, however, people's use of patterns for evaluation reflect psychological benefits that may be obtained from the patterns, such as self-control, which are not fully captured by normative models.

4.2 Discussion of methodology

The previous section (4.1) has covered key theoretical themes of the post-1999 sequence literature. In parallel with the division of the literature review into a section on theory, and a section on methodology adopted in chapter 2, the primary purpose of this section is to present a more systematic review of the methods and results from the experimental studies post-1999, given that methodological issues such as type of stimuli used, choice of dependent and independent variables, and other concerns over the experimental design has not yet been debated in a systematic manner. The results of the post-1999 will be combined with the literature up to 1999, which was reported in chapter 2, and presented jointly.

To reiterate the reasons for reviewing the post-1999 literature in this manner, there are a number of uses for the results of this analysis. First, it may provide additional insights on the relationship between the various methods and parameters used in the experiments that had not been previously discussed. Second, this analysis can also be used as a guide to inform the choice of methods and parameters for use in future experimentation. Third, the

separation of the methods and results review from the general literature review in the previous section (4.1) will enable the construction of a table to summarise key theoretical and methodological aspects of the literature. A tabular representation will make it easier to update the findings of newer literature without having to reconstruct the entire review of theory and methodology every time a new study is published, especially if the findings of the study complements or extends current theory. The ability to represent the key methodological constructs and summary findings of the literature in tabular form is an important point, because it can help the reader follow more easily the evolution in the literature, which is separated into two in this thesis.

As per section 2.4 in chapter 2, the key methodological constructs used in this section are divided into a number of common components. These consist of, but are not limited to, components such as the number and type of participants, stimuli and experimental design, duration of the experience, and results of the studies. The categorisation of such methodological constructs serve a number of purposes, such as the provision of information on parameters (e.g. statistical power, the boundaries of sequence durations, the type of stimuli), and on documenting the experimental paradigms used (e.g. prospective and retrospective methods of capturing evaluation, hypotheses made and results obtained).

The categorisations of key constructs from studies cited in section 4.1 of this chapter are presented in Table 4.1 according to the following criteria below. A summary of the results from the categorisation exercise from Table 4.1, which also includes the previous literature review from Table 2.1, is reported in Table 4.2. In similar fashion to Table 2.1 and Table 2.2 in chapter 2, citations of the studies listed in Table 4.2 will follow the numbering order used in Table 2.1 and 4.1, instead of the conventional Harvard style of referencing used in the rest of this thesis. For example, citations of Ariely's (1998) paper will be referenced as [15]. Research papers that do not contain any original empirical studies are excluded from Table 4.1. On this basis, there are 14 new papers from the post-1999 literature, which when combined with the 15 from pre-1999 literature (see Table 2.1) would give a total of 29 papers that contains at least one

original empirical study. Table 4.2 presents a statistical summary of the methods and results used in every study within the 29 cited research papers from Table 2.1 and 4.1.

4.2.1 Results

(see next page)

Table 4.1: Synopsis of methods employed and results obtained in gestalt studies from year 2000 onwards

Study	Participants	Stimuli and experimental design	Duration	Results
[16] Matsumoto, Peecher and Rich (2000)	54 under-graduates	Wage and rental income streams; Questionnaire involves choice, P	6 years	For all three studies: Decreasing trend preferred amongst participants with accounting/finance knowledge. Increasing trend preferred when the normative or expected value of series is constant
	376 under-graduates	Income streams and Restaurant meal; Questionnaire involves choice, P	6 years; 2 months	See above
	50 under-graduates	Income streams; Questionnaire involves choice, P	6 years	
[17] Ariely and Loewenstein (2000)	45 under-graduates per study x 4 studies	Annoying sounds; a) 0-100 scale (not annoying to very annoying), R b) Willingness to accept (WTA), R c) Rating relative to a standard, R d) Choice, R	10 to 22.5 sec	Response modes that reduce reliance on conversational norms or standard of comparison increased the attention paid to duration

Study	Participants	Stimuli and experimental design	Duration	Results
[18] Ariely and Carmon (2000)	37 bone marrow transplant patients	Pain from transplant; 0-100 scale (no pain to worst pain possible), R	11 hours	Significant trend effect and end effect
[19] Ariely and Zauberman (2000)	54 under-graduates	Annoying sounds; 0-100 scale (not annoying to very annoying), R	50 sec	Increased segmentation and on-line evaluation moderates the effect of trend
	120 under-graduates	Shares; Computer generated stimuli, P	Not mentioned	Same as above
[20] Schreiber and Kahneman (2000)	20 & 24 under-graduates	Annoying sounds; Computer generated stimuli; -250 to 250 scale (extremely pleasant to unpleasant), R	32 sec	Confirmation of peak-end and other gestalt effects
	37, 36 & 28 students	3 experiments; same stimuli as above; -10 to 10 scale (extremely unpleasant to pleasant), R	10 to 24 sec	
[21] Chapman (2000)	100 & 90 members of the public; 115 & 32 under-graduates	Health quality; Questionnaire involves choice; The final group made predictions of other participants' choices, P	1 hour to 20 years	Improving profile generally preferred but moderated by expectations of how health experiences usually progress

Study	Participants	Stimuli and experimental design	Duration	Results
[22] Diener, Wirtz and Oishi (2001)	115 under-graduates	Life experiences; Questionnaire; scale: 1 = unhappy, 5 = neutral, 9 = happy, P	30 or 60 years	Duration neglect and end effect
	55 older friends & family of students	Same as above	25 to 50 years	Same as above
	34 under-graduates	Same as above	30 to 60 years	End effect
[23] Raghunathan and Irwin (2001)	134, 255 & 220 under-graduates	Computer generated experience profiles of consumer products; scale: 1= no pleasure to 11= great pleasure, P	Not mentioned	Improving profile preferred (results from other experiments are not reported since they are unrelated to gestalt research)

Study	Participants	Stimuli and experimental design	Duration	Results
[24] Guyse, Keller and Eppel (2002)	48 graduate business students			Preference for:
		Health quality; Questionnaire involves choice, P	50 years	Constant trend
			5 year treatment	Increasing and constant trend
		Air quality, P	50 years	Same as above
		Air quality, P	5 years	Same as above
		Near-shore ocean, P	50 years	Same as above
		Water quality, P	5 years	Same as above
		Income streams: rental, P	50 years	Decreasing trend
		Income streams: restaurant profits, P	5 years	Decreasing trend
[25] Read and Powell (2002)	40 university members (students & staff)	Income stream and health quality; Think aloud protocol, P	1 year and lifetime (80 years)	Constant trend generally preferred. Preference for other trends dependent on context and duration.
	25 members of the public	Income stream and health quality; Choice-only questionnaire, P	Same as above	Verbal explanations of trend preferences concurs with previous research

Study	Participants	Stimuli and experimental design	Duration	Results
[26] Novemsky and Ratner (2003)	136 & 77 under-graduates	Taste comparison; scale: 0-100 (not enjoyable to very enjoyable), P & R for 1st study, R only for 2nd	Not mentioned	(For all three experiments): Participants expectation of a preference for improvement does not match the actual experience, which shows no such preference
	91 under-graduates	Taste comparison; 7 point Likert scale, P	Not mentioned	
	182 under-graduates	Restaurant meal; 7 point Likert scale, P	1 day to 1 week	Same as above
[27] Ariely and Zauberman (2003)	45, 66 & 60 students	Performance of communication providers (e.g. mobile phone service providers); Computer-based stimuli; 0-100 (not satisfied at all to extremely satisfied) scale, P	24 hours	Retrospective evaluations: Improving profile preferred; This preference decreases in magnitude as the level of segmentation within the profile increased (see [20])
	82 students	Same as above	Same as above	Prospective evaluations: Improving profile preferred, but the effect of trend is reduced by segmentation

Study	Participants	Stimuli and experimental design	Duration	Results
[28] Zauberman and Ariely (2004)	40 students	Test performance; Computer-based stimuli; 0-100 (not satisfied at all to extremely satisfied), P	Not mentioned	Improving preferred to declining profile in hedonic evaluations; Declining preferred to improving in informational evaluations
	100 students	Evaluation of rate of defects at various factories; Computer-based stimuli; 0-10 scale (not satisfied to extremely satisfied), P	12 months	Demonstrated that the preference for an improving sequence is driven by the 'start' and 'trend' rather than by the 'start' and 'end'
	45 members of the public	Performance of communication providers (e.g. mobile phone service providers); Computer-based stimuli; 0-100 (not satisfied to very satisfied), P	12 months	In prospective evaluations, the trend has a larger impact than the start. This is reversed in retrospective evaluations.
[29] Langer, Sarin and Weber (2005)	36 students	Income stream; Computer-based stimuli; Monetary incentives provided to influence performance; 0-100 scale, P	Rate of change =2 sec	No peak-end or duration neglect
	168 students	Same as above	0.1 to 5 sec	Duration neglect and peak-end effect present

Notes:

In the stimuli and experimental design column, ‘P’ denotes a prospective and ‘R’ a retrospective evaluation.

Some studies (e.g. experiment 2 of both **[9]** and **[19]**; all experiments in **[28]** and **[29]**) that purport to use retrospective methods (e.g. participants asked to evaluate the stimuli generated by an apparatus) are classified as prospective (i.e. predictive), because the stimuli used in these studies often require that participants imagine their feelings from hypothetical (and not actual) outcomes (e.g. money in **[9]** and **[19]**).

Participants who have volunteered for the studies cited in this table either did so on a free-of-charge basis, or otherwise received compensation in the form of course credit or an attendance fee. When participants were offered monetary or other incentives to perform better, this will be explicitly mentioned in the experimental design column.

Table 4.2: Summary and breakdown of the methods and results from both Table 2.1 in chapter 2, and Table 4.1, by the number of research papers and citations

Categories		Number	Papers cited (see Table 4.1 for reference number)
1. Method			
Study type	(a) Experimental		
	- Laboratory	16 papers	1, 3, 7, 8, 9, 10, 13, 15, 17, 19, 20, 23, 26, 27, 28, 29
	- Questionnaire or survey-based methods	12 papers	1, 2, 3, 4, 5, 6, 12, 16, 21, 22, 24, 25
	(b) Real-life experiences	3 papers	11, 14, 18
Elicitation of dependent variable	Choice (choices between options presented)	12 papers	1, 2, 3, 5, 6, 12, 16, 17, 21, 24, 25, 29
	Scale (Likert scales and thermometer scales)	21 papers	3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 17, 18, 19, 20, 22, 23, 26, 27, 28
	Willingness to pay / accept (WTP / WTA)	2 papers	1, 17
	Think aloud protocol	1 paper	25
	Prediction of other participants' evaluations	1 paper	21
2. Results	a. Trend effect	19 papers	2, 5, 6, 10, 12, 13, 15, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 28
	b. Peak and / or end effect	14 papers	1, 5, 7, 8, 11, 13, 14, 15, 18, 20, 22, 26, 28, 29
	c. Duration neglect	8 papers	5, 7, 8, 11, 14, 20, 22, 29
	d. Velocity	5 papers	3, 4, 9, 10, 20
	e. Qualification / moderation of gestalt effect	9 papers	4, 12, 15, 17, 19, 25, 27, 28, 29

Note: Some papers contain more than one study, where each study may be classified under a different category (e.g. **[1]**)

4.2.2 Analysis and discussion

The key issues from an analysis of Table 2.1 and Table 2.2 from chapter 2, and its extensions Table 4.1 and Table 4.2 respectively in this chapter are further described in this section. The purpose of this analysis is to provide a quantitative summary of the key parameters used in the literature, and incorporated both the pre and post-1999 literature, thus expanding the methodological discussion started in section 2.4 of chapter 2. This information was then used as a guide in the selection of appropriate parameters for use in further studies reported in this thesis.

4.2.2.1 Choice of participants and power analysis

The discussion on the type and number of participants used in the literature helps to inform an appropriate selection of participants for the experiments in this thesis. The number of participants used is important, because a sufficient sample size is needed in order to meet statistical power requirements to support any significant effects found.

Table 4.3 describes the statistics for the number of participants, which are sub-divided according to the type of experimental methods they have been employed in. The table is divided into two categories. The first category divided participants based on whether they were students or non-students. The second category meanwhile divided participants according to experimental methods, which contained two major experimental methods used, laboratory-based methods and questionnaire-based methods. Given the significant use of computer-generated stimuli in laboratory-based studies reported in the literature, the statistics also included information on the number and type of participants employed in these types of studies. Both the student and non-student participant columns include two figures, a ‘before adjustment’ and an ‘after adjustment’ figure. The adjustment here refers to the need to take into account the number of between-participants conditions for each study in order to make the figures more comparable. This is because some studies employ a large number of participants due to the complexity of their factorial design, while others need fewer numbers due to having a simpler design.

Table 4.3: Number of participants employed

Experimental methods	Student participants		Non-student participants	
	Before adj.	After adj.	Before adj.	After Adj.
Laboratory (all stimuli)				
Mean	77.1	31.7	41.7	14.5
Standard Deviation	58.2	26.0	6.7	2.8
N	37	37	3	3
Laboratory (computer-generated stimuli only)				
Mean	82.7	25.6	45.0	15.0
Standard Deviation	64.9	9.7	-	-
N	20	20	1	1
Questionnaire and survey				
Mean	86.1	56.6	65.9	52.3
Standard Deviation	81.1	42.5	25.3	24.9
N	20	20	10	10

Notes on the calculations of figures in the table above:
N = number of individual studies contained within the 29 research papers in Table 2.1 and Table 4.1.
The category on 'laboratory – computer-generated stimuli' is a subset of 'laboratory – all stimuli', i.e. there is overlap in the calculated figures.
The figures for all columns exclude participants from study [11] and [14]. This is because these two studies are classed as 'non-experimental' for the purposes of this table, because the methods used in these studies involve the measurement of pain from patients undergoing real-life medical procedures, which are not experimental in nature.
Numbers in the 'before adjustment' columns are calculated from Table 2.1, Table 4.1 and Table 4.2; Numbers in the 'after adjustment' column have been adjusted to take into account of the number of 'between-participants' factors in the experimental design.
The adjustment (Adj=adjustment) was based on the number of participants employed per between-participant cell. See Appendix 4.1 for a detailed list of the experimental designs for all studies in each research paper cited in Table 4.1, and the calculations that followed.

The analysis of the results from Table 4.3 is as follows. From the table, students make up the largest type of participants used in the research (25 out of 29 research papers cited). While some academics argued that the injudicious use of students as participants in psychological experiments could potentially be problematic due to the bias in demographic characteristics such as average IQ, family background, and motivation (Higbee and Wells, 1972; Jung, 1969), recent research presents a more sanguine view. In a study on the adequacy of students as surrogates for real practitioners, Liyanarachchi and Milne (2005) found that students' financial investment decisions fared well in comparison to those made by professionals. It is therefore argued that the use of students as the largest group of participants in the thesis (non-students

would also be used) would not present a problem in the context of research on the influence of pattern effects on sequence evaluation.

The results and general conclusions from the handful of experimental studies (laboratory and questionnaire-based) that employed non-student participants (e.g. [2], [6], [13], [21], [22], [25], [28]) are similar to those using student participants, in that the theories developed using student participants can also be applied to studies using non-student participants, and vice versa. Studies that employed non-student participants all involved the use of questionnaire or survey-based methods, rather than laboratory methods, apart from one study ([28]). For example, the theories developed using experimental methods involving students (e.g. [5], [7] and [8]) were deemed to be generalisable enough by their respective author(s) for use in modelling patients' real life evaluations of painful sequences (see [11] and [14]).

There were only two studies ([16] and [24]) that examined the relationship between the preference for sequence patterns and participant type. As reported earlier, the study [16] found that inexperienced students were more likely to rely on pattern for evaluation of monetary sequences compared to more experienced students. In [24], the authors attributed the outcomes of their results (i.e. preference for decreasing monetary sequence), which were different to other studies in the literature (which generally demonstrated the preference for an increasing monetary sequence), to the more experienced nature of their student participants. Therefore, these studies demonstrated that key difference between participants was primarily due to their level of experience in evaluating monetary sequences, rather than their student/non-student status.

In addition to the type of participants used, the number of participants or sample size used in the sequence literature was also analysed. The purpose of this analysis was to justify the numbers of participants that should be recruited for the empirical studies in this thesis. As previously argued, a sufficiently large number of participants is important, because the results of the studies will not be robust enough if the statistical power used (β) is insufficient.

The level of statistical power in each study is a function of the experimental design, that is, the number of participants used in each between-participants condition, and the expected size of the effect.

In the analysis of sample sizes (in Table 4.3), a number of statistics that are useful for the studies conducted in this thesis are identified. Given that the experiments in the next two chapters (5 and 6) will adopt a laboratory-based method using computer-generated stimuli (i.e. studies in chapter 5) and a questionnaire-based approach (chapter 6), the results in Table 4.3 suggest that the number of student participants that should be employed is 25.6 and 56.6 participants respectively. The adjustment procedures used to construct the figures reported in Table 4.3 are detailed in Appendix 4.1. These statistics (i.e. 25.6 and 56.6) are highlighted because it is most relevant to the purpose of the analysis here, which is to obtain an estimate of an appropriate sample size of participants to use.

Table 4.4: Selective reproduction of Cohen's (1969) power table

	f, the effect size for the F-test (F(1,n))		
Power level or (1-β)	Small effect size (f=0.1; η^2=0.001)	Medium effect size (f=0.25; η^2=0.0625)	Large effect size (f=0.4; η^2=0.138)
0.1	22	5	3
0.5	193	32	13
0.8	393	50	20
0.9	526	64	26

Notes:
Cohen's (1969) power table is from Table 8.4.4 (p.377).
The values in the table are the sample size, or n, needed to detect f at the significance level of α=0.05 (5%)
Both f and its equivalent η^2 are shown [$f = \eta^2/(1- \eta^2)$]
n = sample size
The small, medium, and large effect size categories are subjective boundaries suggested by Cohen (1969)

Cohen's power table (1969; p.377, Table 8.4.4) is partially reproduced in Table 4.4 here. The reason for using the values calculated by Cohen is because he had designed a statistical table to guide researchers on selecting the appropriate parameters to use, to ensure that studies have sufficient power. Cohen argued that 0.8 (80%, or β=0.2) is a reasonable power level to operate

with, because it takes into consideration the implicit convention of using a significance level of 5%, or α=0.05, which gives a ratio of the relative seriousness of type II to type I errors of 4 to 1 (from β/α=0.2/0.05=4). Any attempt to further increase the power of a study beyond the level of 0.8 will be prohibitive from a practical and logistical position (difficulty and cost in finding sufficient numbers of participants etc.), because the benefits of doing so do are not commensurate with the additional costs associated with using larger sample sizes (Cohen, 1969, pp.53-54).

The purpose of the analysis below is to determine, from the parameters obtained in Table 4.3 and Table 4.4, the effect sizes obtained in the literature. Once these effect sizes are determined, they can then be used to estimate the effect sizes, and correspondingly, the appropriate sample sizes needed, for studies in this thesis.

If the values reported in Table 4.3 (i.e. the previously highlighted figures of n=25.6 and 56.6 for laboratory-based and questionnaire and survey-based studies respectively), which represents the average sample sizes calculated from the literature are cross-referenced to the nearest statistics in Table 4.4, at the 0.8 power level, it suggests that the effect size of results from laboratory-based studies in the literature are large (f=0.4), because n=25.6 is closer to 20 than it is to 50. The effect size of results from questionnaire and survey studies are medium (f=0.25), because n=56.6 is closer to 50 than it is to 20 or 393. It is unlikely that the real effect size for laboratory-based studies will be medium, given that this implies the use of a power level (in the literature) that is less than 0.5 (50%), which is inadequate to prevent type II errors (i.e. the incorrect denial of the existence of an effect or a relationship that actually does exist) from seriously affecting the interpretation of results (Cohen, 1969). The same argument also applies for questionnaire and survey-based studies, where it is unlikely that the real effect size is small (f=0.1), given that it implies a power level usage of less than 0.5. In summary, assuming a power level of 0.8, effect sizes obtained from laboratory-based studies in the literature are on average large (f=0.4), while effect sizes of questionnaire and survey-based studies are medium (f=0.25).

It has to be emphasised that the f value calculated here are only estimates of the actual effect sizes, since it is assumed that all the empirical studies in the literature adopt the base (ANOVA) F-test design with one degree of freedom (df) for the numerator. While many other experimental and F-test designs exist, it is argued that this approximation to the F-test design with df=1 for the numerator is sufficient for the purposes of approximating power levels and sample sizes that would form the basis of guideline parameters to adopt in future experiments. This is because all studies in the empirical chapters in this thesis are based on variants of this design (i.e. ANOVA, with df numerator=1).

4.2.2.2 Stimuli and experimental design

The majority of the studies reported in Table 2.1 and Table 4.1 (26 out of 29) employed experimental methods for hypotheses testing. One of the main reasons why the use of experimental methods is more common compared to the analyses of people's real-life experiences with sequences is because experimental methods allow for a highly controllable environment, where noise and other extraneous variables can be kept to a minimum. This is an important point, because most of the real-life experiences do not allow for easy or direct comparison. For example, it is difficult to compare an increasing (say {1,2,3}) versus a decreasing sequence (say {9,8,7,6} of pain, if the total duration of the two sequences (or its average) are not equal. In other words, it is not possible to establish a basis for comparisons between sequences for real-life experiences.

Only a small number of research papers based on real-life experiences with sequences have managed to incorporate a substantial element of experimental control over the delivery of stimuli. All three of these kinds of studies ([11], [14], [18]) were based on patients' evaluations of clinical procedures, where the delivery of stimuli (both intensity and duration) to the patient was precisely controlled and accurately measured using medical instruments (colonoscopy and lithotripsy procedures in [11] and [14], and a bone marrow transplant operation in [18]). The methods used provided a basis for comparing the patients' evaluations in all three studies, because there were systematic and

objective methods available to measure evaluation, and the intensity and duration of the stimuli administered. This degree of measurement accuracy, however, is not possible for other real-life experiences of sequences, which is why most of the studies adopt experimental methods. Apart from medical procedures, it is difficult to find real-life experiences that would allow for such precision of control over the quantity and duration of stimuli delivered. For the majority of researchers, this has meant that experimentation in a laboratory-based setting is the next best and often the only feasible alternative for testing hypotheses in gestalt research. Of the 16 research papers that adopted laboratory-based methods for the experiments, 9 of them chose to use computers to deliver the stimuli. The advantages of doing so are that the stimuli can be delivered with a high degree of precision (e.g. very short durations such as those involving one or two seconds can be accurately delivered) and control (e.g. the entire process is encapsulated on-screen, and all parameters are inputted by the experimenter, thus reducing the amount of experimenter 'presence' to a minimum).

Table 4.5: Analysis of the relationship between evaluation timing and results

	Number of research papers		Number of individual studies within the research papers	
Evaluation timing	prospective	retrospective	prospective	retrospective
Results obtained				
Trend and velocity	19	9	39	18
Peak-end and duration neglect	6	9	11	16
Qualification / moderation of pattern effect	4	5	7	11

The purpose of the analysis in Table 4.5 is to determine whether the results obtained from studies can be attributable towards a particular paradigm, given that these evaluation paradigms involve different psychological processes (Elster and Loewenstein, 1992, p.227, para.2; see also earlier discussions in section 2.3 of chapter 2). Table 4.5 presents data based on dividing up the literature into two categories, based on the evaluation paradigm (prospective versus retrospective) of the study. The table also reports on the number of

individual studies within a research paper, given that many of the papers contain more than one study. As a brief reminder of the paradigms, prospective studies measure people's anticipatory/predicted responses to sequential stimuli, whereas retrospective studies measure responses after the stimuli are experienced. It is observed from the first two columns of Table 4.5 that the largest category of studies comes under the prospective paradigm (19 research papers), which investigated trend and velocity effects in sequences. There is, however, no discernable bias when the results obtained are compared with the evaluation paradigm. The only obvious difference is from the large number of individual studies (39) using the prospective paradigm to investigate the effects of trend and velocity. It is argued that this number reflects studies that extend the earlier findings of trend to new domains (e.g. from monetary and aversive stimuli, to advertising stimuli, health etc.). Studies extending the peak-end and duration neglect theories are more limited in frequency in comparison to studies investigating trend and velocity. This is perhaps because peak-end theories are largely derived from studies that research sequence evaluations for stimuli of an aversive nature, apart from study [29] from Table 4.1, which may impose limits on its applicability.

4.2.2.3 Duration

The analysis of the sequence duration (i.e. the length of time from start to end of sequence) and its relationship with the type of experiment and the evaluation paradigm is reported in Table 4.6. The results of the categorisation show that studies involving questionnaire and survey-based methods exclusively use the prospective evaluation paradigm, while laboratory-based studies consist of roughly equal numbers of both retrospective and prospective evaluation paradigms. Note that laboratory-based studies that involved longer durations (more than 1 hour) are exclusively based on the prospective evaluation paradigm. Studies that used longer sequence durations only utilise the prospective, but not retrospective, evaluation paradigm. There are practical reasons for this: it is unrealistic to expect that volunteers would willingly participate in retrospective, laboratory-based studies that last for more than an hour. So, the major reason for the choice of paradigm seems to depend on the magnitude of the sequence duration that is under investigation.

Table 4.6: Analysis of the relationship between duration and evaluation paradigm used

		No. of research papers		No. of individual studies	
Timing of evaluation		prospective	retrospective	prospective	retrospective
Type of study	Duration				
Laboratory	Less than 180 sec	1	8	2	18
	3 min to 1 hour	-	1	-	1
	More than 1 hour	4	-	9	-
	Not mentioned	5	2	9	2
	Total	10	11	20	21
Questionnaire and survey	1 min to 23 hours	2	-	6	-
	1 day to 52 weeks	5	-	10	-
	More than 52 weeks	8	-	17	-
	Not mentioned	1	-	1	-
	Total	16	-	34	-

Note: The total number of research papers sum to more than the 29 reported in Table 4.1 (e.g. 10+11+16+0=37) because some papers contain multiple studies with different durations.

4.2.3 General discussion and recommendations

The purpose of the analyses in this section (4.2) is to provide a quantitative summary of the key parameters used in the literature, which can then be used to aid the selection of appropriate parameters for experiments in this thesis. The analyses identified the type of participants and average sample sizes used, and have categorised the experimental designs, other parameters used, and evaluation paradigm, according to the results obtained. The inference that can be drawn from the analyses is as follows.

On the selection of an appropriate choice of power levels to use, the studies in this thesis adopted Cohen's (1969) recommended power level of 0.8 (i.e. β=0.2). At this level, the corresponding sample sizes that should give a sufficient level of power are approximately 20 and 50 participants respectively for laboratory-based and questionnaire-based studies. All empirical studies in

this thesis reported in the next two chapters (5 and 6) have used sample sizes equal to, or larger than, the recommendation (i.e. 20 and 50 participants respectively).

On the selection of experimental methods to use, it is argued that laboratory and questionnaire/survey-based methods provide a practical, cost-effective and reliable alternative to having to test hypotheses using real-life experiences of sequences. The analyses however, points out the various merits and de-merits of using laboratory and questionnaire-based methods, which should be taken into account when designing future studies. Laboratory-based studies are often more realistic compared to questionnaires, because in a questionnaire, participants have to expend effort to imagine themselves experiencing a stimuli, which may increase the risk that the student misinterprets the question posed. In laboratory-based methods, however, such risks are less, because no imagination on the part of the participant is required because they actually experience the sequence stimuli. Another advantage of laboratory-based studies is that it is the only practical method capable of testing hypotheses that involves a retrospective evaluation paradigm. However, laboratory-based methods are less feasible for studies that involve durations that stretch for more than one hour, due to logistical reasons. Such methods may also potentially suffer from the problem of being able to recruit a sufficiently wide variety of people (i.e. type, not number) to participate in the study, due to cost and time constraints. Trying to obtain sufficiently large numbers of people who are within distance of laboratories may be too expensive, given that there is little advantage to be gained from doing so. Finally, based on the results presented in Table 4.6, it is suggested that an appropriate upper limit for the sequence duration for any retrospective laboratory-based study should not exceed 180 seconds. Any study that requires the use of a longer duration should consider using prospective methods, such as questionnaire and surveys, instead.

The use of interview-based methods is not widespread in the sequence literature. There is only one study [25] that uses a type of interview method, called the think-aloud protocol. In this method, participants are given the

stimuli, and encouraged to talk about their mental processes as they are considering and evaluating the stimuli. One of the advantages for adopting this method is that it allows the participant to give a range of answers that are not specifically limited to or pre-determined by the experimenter. This method is entirely appropriate for the hypothesis of the study cited, which was to explore people's reasons and thinking processes behind their preference for sequences. One of the reasons for the lack of popularity in adopting this method is because the hypotheses developed in most studies on pattern effects do not require such detailed descriptions of reasons for sequence preferences, when simpler methods to measure evaluation (such as questionnaires and surveys, and laboratory-based studies) are sufficient for the purpose.

4.3 Summary of the chapter

The purpose of this chapter was to extend the pre-1999 literature review reported in chapter 2, to incorporate more recent post-1999 developments. This post-1999 literature review was divided into two sections. The first section (4.1) described the theoretical developments since 1999. The second section (4.2) extended the analyses of methods used in the literature. In this second section, estimates of guide parameters were derived and key techniques used in the literature were summarised for the purposes of informing the future development of studies in this thesis. The next two chapters described the empirical section of this thesis.

Chapter 5

The role of task achievability in sequences

5.0 Introduction

The broad objectives of the studies in this chapter and the next chapter are to further refine the model used (i.e. Model 1) to predict how people evaluate sequences of delays in the Internet Delay Study (IDS) in chapter 3. This process involves a re-examination of Model 1's theoretical base and consideration of other factors that may influence but have not been incorporated into the model. This is held to be important for the research questions presented in this thesis, because its outcome would provide new knowledge on the extent to which gestalt theories can be used to predict behaviour in sequences of delays. The results from this research contribute to the literature by providing knowledge on the boundary conditions of the theories on sequence evaluations.

A model of how people evaluate sequences of delays was developed (Model 1) in the IDS. Based on evidence from the sequence literature, it was posited that the evaluation of delay sequences was not a simple un-weighted integral of the individual utilities from each delay component within the sequence, but instead had an end-loaded weighting. This meant that a sequence with aversive stimuli that is monotonically increasing in intensity would be evaluated more negatively than one with aversive stimuli that is a monotonically decreasing in intensity, even if the total level of aversion for both sequences were identical. The following notations below reproduce Model 1, which was previously derived in chapter 3.

Let the magnitude of delays be monotonically increasing (d1<d2<…<dn) or decreasing (dn>…>d2>d1), and the distribution of weights be monotonically increasing (w1<w2<…<wn). Therefore, Model 1 predicts that an increasing

delay sequence, V(d1, d2, …, dn), would be evaluated more aversively than a decreasing sequence, V(dn, …, d2, d1). This is written as:

V(d1, d2, …, dn)>V(dn, …, d2, d1) … [Model 1]

However, the result of the IDS in chapter 3 did not support the predictions of Model 1 and were inconclusive. It was suggested that sequence evaluations were bimodal, with one group of participants preferring an end-loaded distribution, while the other preferring a front-loaded distribution. This conclusion, however, could not be confirmed, because the possibility of measurement error could not be discarded from the analysis.

It is suggested that one of the reasons why the prediction of Model 1 was not supported is because the delays in the IDS did not significantly annoy people beyond a baseline level. For example, the delays in the IDS only obstructed progression for 110 seconds, which may not be sufficient to generate enough negative emotions, when compared to stimuli of more aversive nature reported in the literature, such as physical pain (e.g. Ariely, 1998; Kahneman et al., 1993). For the studies in the present chapter, it is therefore proposed that the level of aversion in delay sequences has to be increased to generate sufficient negative emotions for inducing an evaluation based on patterns, as per Model 1. The broad objectives of the studies reported in this chapter, therefore, are premised on testing this proposition.

The outline of the remainder of this chapter is as follows. The purpose of the first study (i.e. Memory Study (I) or MS1) in section 5.1 is to introduce a task manipulation that can generate sufficient negative emotions to induce an evaluation based on patterns. Briefly, the manipulation used comprised of delays that prevent participants from achieving the tasks set for them. The objective of the second Memory Study (MS2) in section 5.2 is to improve on the design and retest the hypothesis of MS1, given that there were concerns over the methodology used in MS1. The third study in section 5.3, Visual Search Study (I) or VS1, introduces an alternative method of manipulating obstruction in a sequence of tasks, by replacing delays that impede

progression with time constraints that reduces the amount of time allowed to complete a task. The reason for introducing this alternative method of manipulation is to demonstrate that an evaluation based on patterns is induced because the task manipulation prevents full task achievement. The purpose of the fourth and final study in section 5.4, Visual Search (II) or VS2, is to demonstrate that the manipulation of task achievability used in the previous studies – MS1, MS2 and VS1 – is indeed the determining factor that induces an evaluation based on patterns. The final section 5.5 summarises and concludes the key findings of this chapter.

5.1 Memory Study (I) or MS1

5.1.1 Introduction

The main purpose of this study is to test Model 1 using an alternative method of generating negative emotions. In the IDS, the manipulation used to generate negative emotions was delays that obstructed (but did not prevent) the tasks from being achieved. In the present study, however, the tasks were set up such that the delays would make the tasks difficult to achieve.

For example, imagine a task where people are required to memorise and recall a number of visual objects. The longer people have to keep objects in memory, the harder it will be for them when it is time to recall the objects. Therefore, delays between people's registration of the objects in memory and recall would be directly proportional to the difficulty of the task for them. It is posited that the longer the delays, the more difficult the task becomes, and the more frustrating it will get, because obstruction caused by the delays has increased (Lawson, 1965; Hornik, 1984; Sawrey and Telford, 1971). The memory task proposed in the present study would require more effort and attention from people compared to the previous delays in the IDS. In addition, there is also a factor of task achievability, where, for example, people are not always able to achieve the tasks required of them in the memory task, whereas they were previously able to complete all tasks in the IDS. The delays can be said to have a significant impact on people's performance if it impinges on people's ability to fully achieve the tasks required. If the performance in the memory task deteriorates as a result of the delays (i.e. they forget some of the objects), frustration is predicted to increase.

In addition to the increased levels of frustration, it is also posited that as the overall level of delay increases, the memory tasks become more difficult to achieve, and that people would rely more on heuristics such as patterns in the evaluation of a delay sequence. This is because the increase in the obstruction of the memory tasks places a greater information-processing burden on the decision-maker. This is predicted to increase the propensity of

the decision-maker to rely more on heuristics (e.g. the evaluation of a sequence based on its pattern) for sequence evaluation, as the decision-maker has to switch to simpler strategies for evaluation in order to cope with the increased processing requirements (Gigerenzer and Todd, 1999; Ariely and Carmon, 2003). As Kahneman et al. (1997) demonstrated, reliance on heuristics can lead to sequences with more overall negative utility receiving less aversive retrospective evaluations compared to sequences with less overall negative utility, because such evaluations are only based on a subset of the components of the sequence, such as its gestalt characteristics. To summarise, the present study is designed to evoke negative emotions by introducing delays that prevent participants from fully achieving the sequence of memory tasks set for them. It is predicted that the level of negative emotions evoked is sufficient to induce an evaluation based on patterns. The hypothesis for the Memory Study (I) is as follows:

H1: The decreasing delay sequence is less frustrating than the increasing delay sequence

5.1.2 Method

5.1.2.1 Participants

Seventeen members of the British public (non-students) volunteered to participate in the experiment. Their ages ranged from 21 to 32 years old. Ten of the participants were female.

5.1.2.2 Stimuli and experimental design

Several pilot tests (not reported here) using five other non-student volunteers (who did not participate in any of the studies reported in this thesis) were conducted to determine the appropriate choice of parameters for the Memory Study (I). These pilot studies tested various combinations of time and examined the performance in the memory task. The choice of parameters used was made on the basis that the memory tasks should be unachievable to most if not all participants. The stimuli were presented using a computer-based visual memory task.

Table 5.1: The parameters used in the experiment

	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6
Increasing time sequence						
Letter	X	T	O	H	X	T
Colour of letter	Black	Yellow	Green	Red	Blue	Brown
Duration of appearance (sec)	2	3	5	11	19	35
Decreasing time sequence						
Letter	Y	I	Q	A	Y	Q
Colour of letter	Black	Yellow	Green	Red	Blue	Brown
Duration of appearance (sec)	35	19	11	5	3	2

In this study, participants were asked to perform two sequences of memory tasks, once with increasing time, and the other, with decreasing time. The total duration for both sequences was equal. Each sequence had six objects that appeared consecutively, and participants had to commit them to memory as they appeared, whilst retaining the previous objects held in memory. Each object was made up of a coloured letter taken from the English alphabet. These objects remained on the screen for a fixed duration, before they vanished, and the next object appeared. The parameters used (e.g. the duration that the objects appeared on the screen, the object used, and the colour of the object) were reported in Table 5.1. The timings (in seconds) for the increasing time sequence are {2, 3, 5, 11, 19, 35} and for the decreasing time sequence {35, 19, 11, 5, 3, 2}. The total time (aggregate duration of all six pages) for both sequences is (2+3+5+11+19+35) = 75 seconds. Each object was placed amongst a light grey backdrop to make it more visible, and the font type and size used was Arial 48 point.

5.1.2.3 Measure of dependent variable

The instrument used to measure the first dependent variable (Q1) consisted of a 0-100 horizontal line (0=not frustrating; 100=very frustrating), where participants were required to make a mark along the line to indicate how frustrated they were in trying to remember the objects (see Figure 5.1.1a). The instrument used was similar to the one used in the IDS, except that the scale used in this study allowed for more variability in the responses (100 possible ratings instead of 10). In addition, the keyword of the dependent variable changed, where ‘annoyance’ from the IDS was replaced with ‘frustration’ for the current study. The replacement was made because the results of the pilot

studies (which were derived from a similar one used in Taylor, 1994) suggested that ‘frustration’ was a more appropriate word to describe the negative feelings induced by the waiting. The majority of participants (six out of seven) in a pilot study (not reported here) suggested that ‘frustrate’ was the most appropriate word to use to describe the negative affect caused by the delays, out of a list of possible synonyms that included words such as ‘annoy’, ‘impede’, ‘aggravate’, ‘irritate’, ‘upset’, and ‘disturb’. The second dependent variable (Q2) was used to record the objects that participants had managed to recall. Participants had to indicate both the correct letter and correct colour of the object (see Figure 5.1.1b).

Figure 5.1.1a: The instrument used to measure ratings of frustration (Q1)

How **frustrating** did you find it to remember the coloured letters?

Make a small mark on the line below.

0 100

Not frustrating at all Very frustrating

Figure 5.1.1b: The instrument used to record the letters remembered (Q2)

Circle the coloured letters you remember.

T	T	T	T	T	T
X	X	X	X	X	X
O	O	O	O	O	O
H	H	H	H	H	H
(Black)	(Yellow)	(Green)	(Red)	(Blue)	(Brown)

Figure 5.1.2: Illustration of one of the six pages in the Memory (I) study

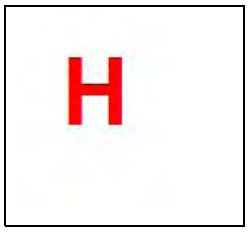
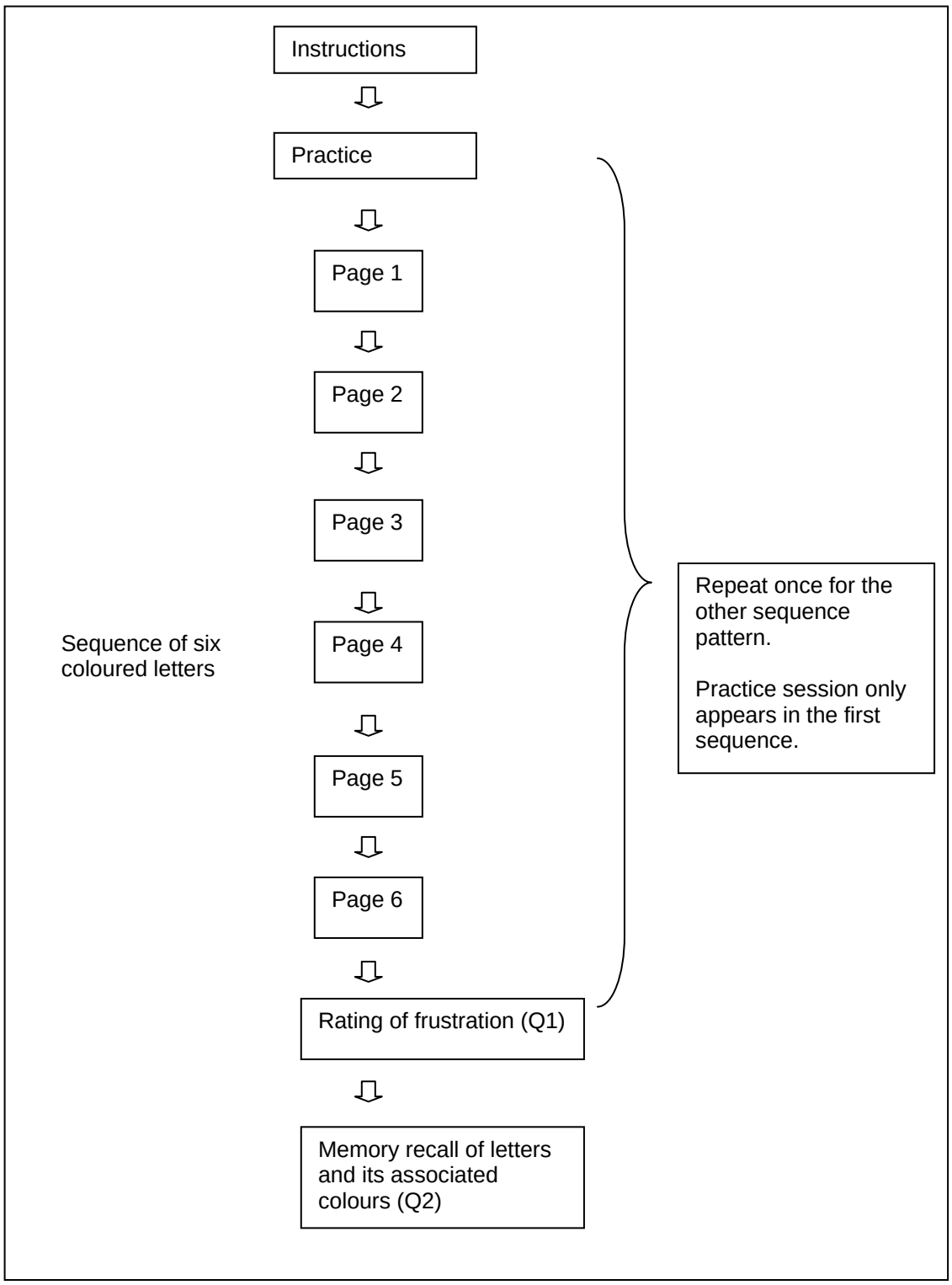


Figure 5.1.3: Schematic layout of the memory (I) experiment



5.1.2.4 Procedure

Figure 5.1.3 presents a flow diagram of the experiment, and Figure 5.1.2 illustrates one of the six pages from Figure 5.1.3 in which the object appeared on the computer screen. On the instructions page, participants were told that their task for the experiment was to commit to memory a sequence of objects, in the form of coloured letters that would appear in a consecutive manner. They then were required to evaluate the sequence, and to recall the objects committed to memory, once the computer had prompted them. After the reading the instructions, participants were then given an opportunity to familiarise themselves with the tasks required of them in a practice session. After the practice session was completed, the experiment started. Participants experienced the two sequences of delays and were required to evaluate them, as well as recall the objects. The order of sequence appearance (increasing before decreasing versus decreasing before increasing) was counter-balanced by randomly allocating the seventeen participants to one of the two conditions. Nine participants were allocated in the increasing before decreasing condition, and the rest to the other condition. After participants had completed the memory tasks in both sequences, they were then told that the study had ended, and were thanked for their efforts. The instructions for the study are reproduced in full in Appendix 5.1.

5.1.3 Results

The descriptive statistics for the dependent variables rating of frustration and performance in the memory task (i.e. number of correctly recalled letters and its associated colours) are reported in Table 5.2 below. The results indicated that participants evaluated the increasing pattern as more frustrating than the decreasing pattern, which was in the same direction as the prediction made in H1. Participants' achievements in the memory task, however, were similar across both sequence patterns, and reflected their inability to correctly recall all of the objects committed to memory. Based on the median result, participants only managed to correctly recall four out of the six objects, thus suggesting that the delay parameters had impeded their ability to fully achieve their task of recalling all of the objects. Statistical analyses were conducted to provide

further confirmatory evidence to support the observed hypothesised result, and to help discriminate between the mean versus median result of the memory task.

Table 5.2: Descriptive statistics of the dependent variables

	Mean	Median	Standard Deviation
Measure of frustration (Q1)			
Increasing time sequence	53.7	42.1	27.0
Decreasing time sequence	33.8	26.3	22.6
Number of correctly recalled letters (Q2)			
Increasing time sequence	4.06	4.00	1.60
Decreasing time sequence	4.53	4.00	1.18

Note 1: The figures for the measure of frustration are taken from the 0-100 scale used, where higher numbers indicate more frustration.
Note 2: The total number of letters for each of the sequences is 6.

Given that this study had adopted an identical ANOVA design to the IDS, a 2x2 mixed ANOVA with the between-participants factor 'order' (increasing before decreasing versus decreasing before increasing) and the within-participants factor 'pattern' (increasing versus decreasing delay) was calculated. The ANOVA did not reveal any significant effects of order ($F(1,15)=0.968$, $p>0.05$) nor of any significant interaction between order and pattern ($F(1,15)=0.000$, $p>0.05$). This result implied that order effects were not significant for this model, which is a similar result to the IDS. There was, however, a significant main effect of pattern ($F(1,15)=19.2$, $p<0.001$), which supported the hypothesis H1 that the decreasing pattern was less frustrating compared to the increasing pattern.

For 'task achievement', the single factor ANOVA supported the median observation in Table 5.2, where achievement was identical across both sequence patterns ($F(1,16)=0.968$, $p>0.05$). This result suggested that pattern does not influence participants' ability to recall the objects. Pattern should not influence the level of task achievement because it does not increase or reduce the total amount of delay in the sequence, but merely determines its distribution. However, there are some problems with the methodology used in this study that would negate this argument. These problems are discussed below.

5.1.4 Discussion

The findings from this study support the hypothesis that the decreasing delay sequence would be evaluated as being less frustrating than the increasing delay sequence. This result was observed under conditions where people were unable to fully achieve the tasks set for them. This finding is in contrast with the previous IDS, where no significant preference for a sequence pattern was observed. It was argued that the main conceptual difference in methodology between the two studies could be attributed to the achievability of tasks. If the heuristic framework discussed in the post-1999 literature review in chapter 4 (e.g. Ariely and Carmon, 2003; see also section 4.1.3) was used as an explanation of the phenomena observed, participants in this study had to rely on pattern for evaluation when they faced tasks that were unachievable, because they had to devote most of their cognitive resources to the challenging memory task. This meant that they had fewer resources available for evaluating the sequence of delays using the utility integration principle (UIP), which in turn increases their reliance on heuristics such as patterns that may result in asymmetric evaluations between sequences. In the IDS, however, the tasks set were always achievable, which suggested that participants might have had surplus resources that they could use to consider other non-pattern factors (e.g. domain) before evaluating the sequence of delays. To summarise, the manipulation of the participant-task interaction used in this study managed to induce a result that tentatively supported the predictions of Model 1.

The strength of the conclusions that could be drawn from this study, however, was limited, because the design of the study was confounded. This was an oversight on the part of the experimenter, which was only discovered at a later stage during this study. By this time, seventeen participants had already completed the study before the problem was recognised. Nevertheless, it was felt that, given the results of this study up to this point did not seem to be adversely affected by the confound, it was felt that there was value in reporting the results. In addition, this reporting would also provide a motivation for the

improvement in the current design for the next study in this chapter, Memory Study (II) or MS2.

The problem of confounding was caused by uneven memory loads of the two sequence patterns, whereby participants experienced a greater cognitive load for a longer period of time in the increasing delay sequence compared to the decreasing delay sequence. For example, in the increasing time sequence, participants have to keep three objects in memory for (2+3+5)=11 seconds (i.e. the first three tasks). They then have to remember another three objects for the next (11+19+35)=65 seconds (i.e. the final three tasks). For the decreasing time sequence, however, participants have to initially keep three objects in memory for (11+19+35)=65 seconds (i.e. the first three tasks). They then have to remember another three objects for the next (2+3+5)=11 seconds (i.e. the final three tasks). This meant that for the increasing time sequence, participants have to keep in their memory the 4th, 5th and 6th object in the longer half of the sequence (65 seconds), compared to that of the decreasing time sequence, where the 4th, 5th and 6th object is located in the shorter half of the sequence (11 seconds). Since participants had to memorise more objects for a longer period in the increasing compared to the decreasing time sequence, this implied that the results may have been contaminated by this confounded design.

Another potential limitation of the design in the present study is that the timing manipulation represents the time the object is displayed for. It is assumed that the time required for recognising and processing the objects in memory is brief. However, it could also be that the decreasing delay sequence was less aversive because participants have more time at the beginning of the sequence to commit objects to memory, therefore making them feel that they have achieved more before the sequence began to speed up. This is despite the final result where it was shown that participants did not manage to find more objects despite the less aversive experience rated.

Either way, it was not possible to separate out the portion of the sequence ratings that could be attributed to the greater cognitive load for one of the

patterns (i.e. the problem caused by the confounding) to the portion that could be attributed to the reliance on patterns for evaluation. While the analysis of task achievement in this study showed that achievement was unaffected by confounding (i.e. equal achievement in both increasing and decreasing sequences), it is nevertheless imperative to redesign the experiment to rectify the confound. This was the aim of the next Memory Study (II).

5.2 Memory Study (II) or MS2

5.2.1 Introduction

The main objective of this study was to improve on the confounded design of MS1 by eliminating the unequal memory task loads that caused the confound, while keeping the hypothesis and theory developed unchanged. The key improvement in the design of the current study is to position the memory task such that participants experience it before, but not during, the sequence of delays. The location of the memory task would then mean that both increasing and decreasing delay sequences have equal cognitive loads, as the objects committed to memory would be held for equal durations across both patterns before recall is permitted.

In this study, a second task, called the stroop-matching task (Stroop, 1935), was also introduced, in addition to the memory task. The positioning of this task was within the sequence of delays itself, where the purpose of this task is to segment the sequence, to create an increasing and decreasing pattern of delay. From a theoretical view, segmentation is important to the creation of a sequence, because it defines the pattern of delays. Without segmentation, the entire sequence would be perceived as one long delay, rather than a number of smaller set of delays that are related to each other based on its pattern. This argument is analogous to the discussion in Ariely and Zauberman (2000 and 2003), who found that the strength of pattern effects were highly conditional on the way in which the sequence was segmented. Ariely and Zauberman (2003) found that segmentation at strategic points could create the impression of an improving or deteriorating sequence.

In the previous study, the segmentation of the delays was achieved through the individual memory tasks that had varying durations of delay. The problem of that study, however, was that it did not control for equal memory loads. The stroop-task in the current study functioned in a similar fashion to the memory tasks in the previous study, but corrected for the memory loads across both sequence patterns by ensuring that the total memory load at any one time

during the sequence was only one additional object in memory. Previously, the increasing sequence was more difficult because more objects had to be held in memory for a longer period, compared to the decreasing sequence. In this stroop-task, participants were only required to keep a running total of the number of matches between the actual colour of the letters that they saw and the name of the colour that was spelt out. For example, a match was when the colour that participants saw was blue, and the letters also spelt blue. A non-match is when the colour seen was blue, but the letters spelt red. The stroop-task for both the increasing and decreasing delay sequences in this study was set such that the colour seen was always the colour spelt, and that participants only had to keep a running total of the number of matches that they had seen. Since the running total is only one number at any one time, therefore it is argued that the increase in memory load at any one time during the sequence has only increased by the factor of one additional object.

The graph of cognitive load versus time in Figure 5.2.1 (a and b) provided a visual illustration of the improvements in the design of the study. The top two graphs in Figure 5.2.1a illustrated the time profile of the memory load that participants experienced in MS1. The vertical scale is the cognitive load from the memory task at a particular moment in time, which is represented by a numbered box in the graphs. The numbers in each box represented the number of objects that participants have to keep in memory. The horizontal scale represented time in seconds, and represent the temporal progression of the sequence. At the end of 75 seconds, the memory loads for both sequences were equal, where up to six objects were held in memory. However, the MS1 was confounded because the memory loads were not equal during the sequence progression. The shaded area represented the difference in the memory loads between the increasing and decreasing sequence in the first Memory Study (MS1). The loads in this study (i.e. MS2) were represented in Figure 5.2.1b. The cognitive load for both the increasing and decreasing pattern was identical, because participants only need to keep in memory a maximum of eight objects (the seven objects from the first task plus keeping a running total of stroop-matches observed in the second task). The memory loads in the stroop-task would be equal across both patterns because

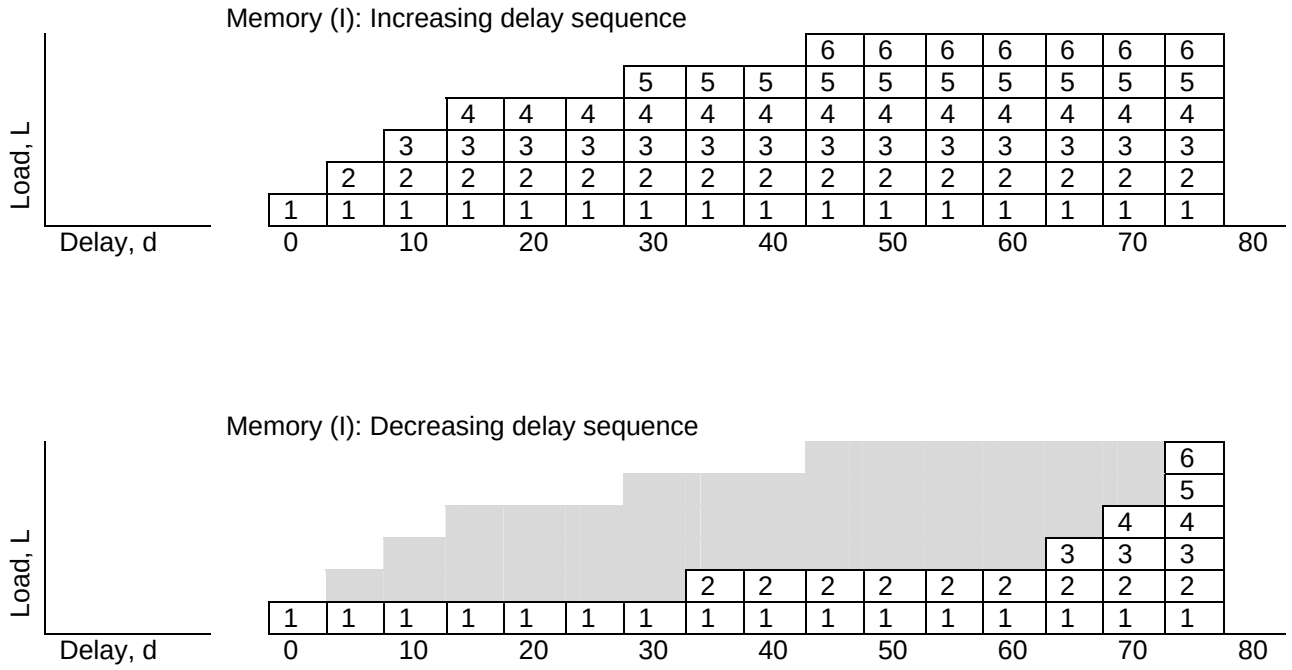
participants were not trying to remember all of the colours seen and spelt, but only needed to add the number of matches that they had counted to a running total. The parameters used to create the graphs (e.g. the duration of delays) were obtained from the method section of MS1 and the present study.

Figure 5.2.1b describes the progression of the sequence and its memory load. Here, participants are allowed a time of 15 seconds (t=-15 to t=0) to complete the first task, which is to commit the seven objects to memory. They have to then keep these objects in memory for 79 seconds, while performing a series of stroop-matching tasks (i.e. the second set of tasks).

The positioning of the stroop-matching tasks differ during the 79 second sequence duration, and is used to create an impression of an increasing or decreasing time sequence. In the increasing time sequence, the majority of stroop-matching tasks are positioned closer to the start of the sequence (indicated by ‘x’), while for the decreasing time sequence, the stroop-matching tasks are placed closer to the end of the sequence. This gives the participant the impression that the intervals between the stroop-matching tasks are either lengthening (increasing) or shortening (decreasing). It is assumed that the memory load, however, does not vary between the two sequences, because participants only need to keep one running total of the number of stroop-matches they find at any moment in time during the sequence.

Figure 5.2.1: Illustration of how the confound is removed

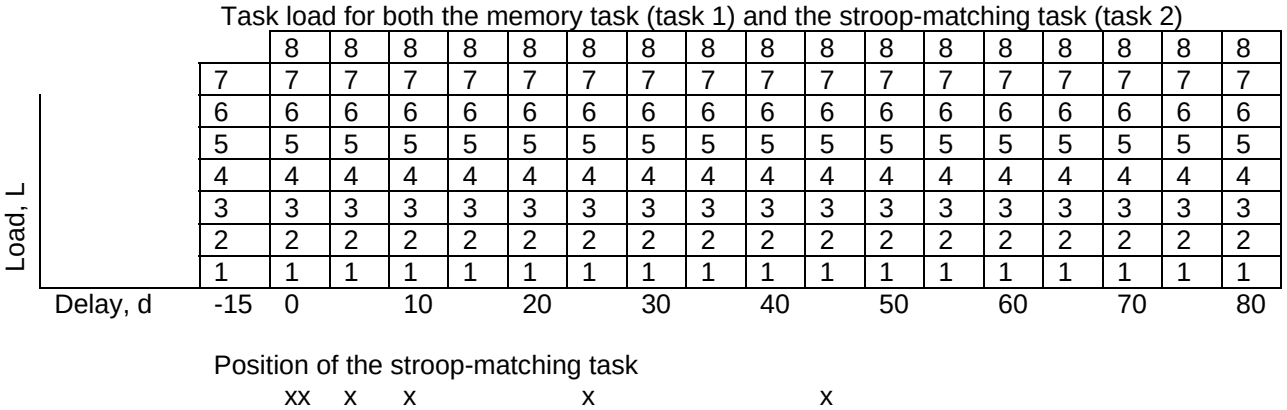
Figure 5.2.1a: Memory load versus time for Memory Study (I)



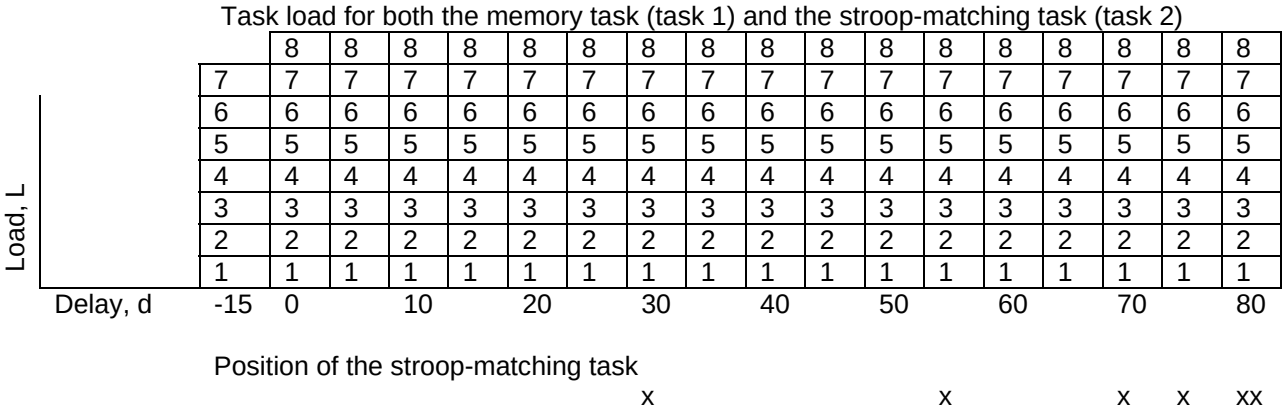
- The vertical axis represents memory load, while the horizontal axis represents the delay duration.
- The lightly shaded area represents the difference in cognitive load between the increasing versus decreasing delay sequence.
- The start of both sequences is indicated by t=0, and the sequence ends after 75 seconds.

Figure 5.2.1b: Memory load versus time for Memory Study (II)

Increasing delay sequence



Decreasing delay sequence



The hypothesis for this study is the same as in MS1. In addition, it is also expected that achievement on both the memory and stroop-matching tasks would be independent of pattern, because pattern does not increase or reduce the overall level of difficulty of both tasks. It is predicted that task achievement for the increasing sequence will be similar to that for the decreasing sequence, for both the memory and the stroop-matching tasks. In addition, it is also predicted that participants will not be able to fully achieve the memory tasks.

H1: The sequence of decreasing delays would be evaluated as being less frustrating than the increasing delay sequence

5.2.2 Method

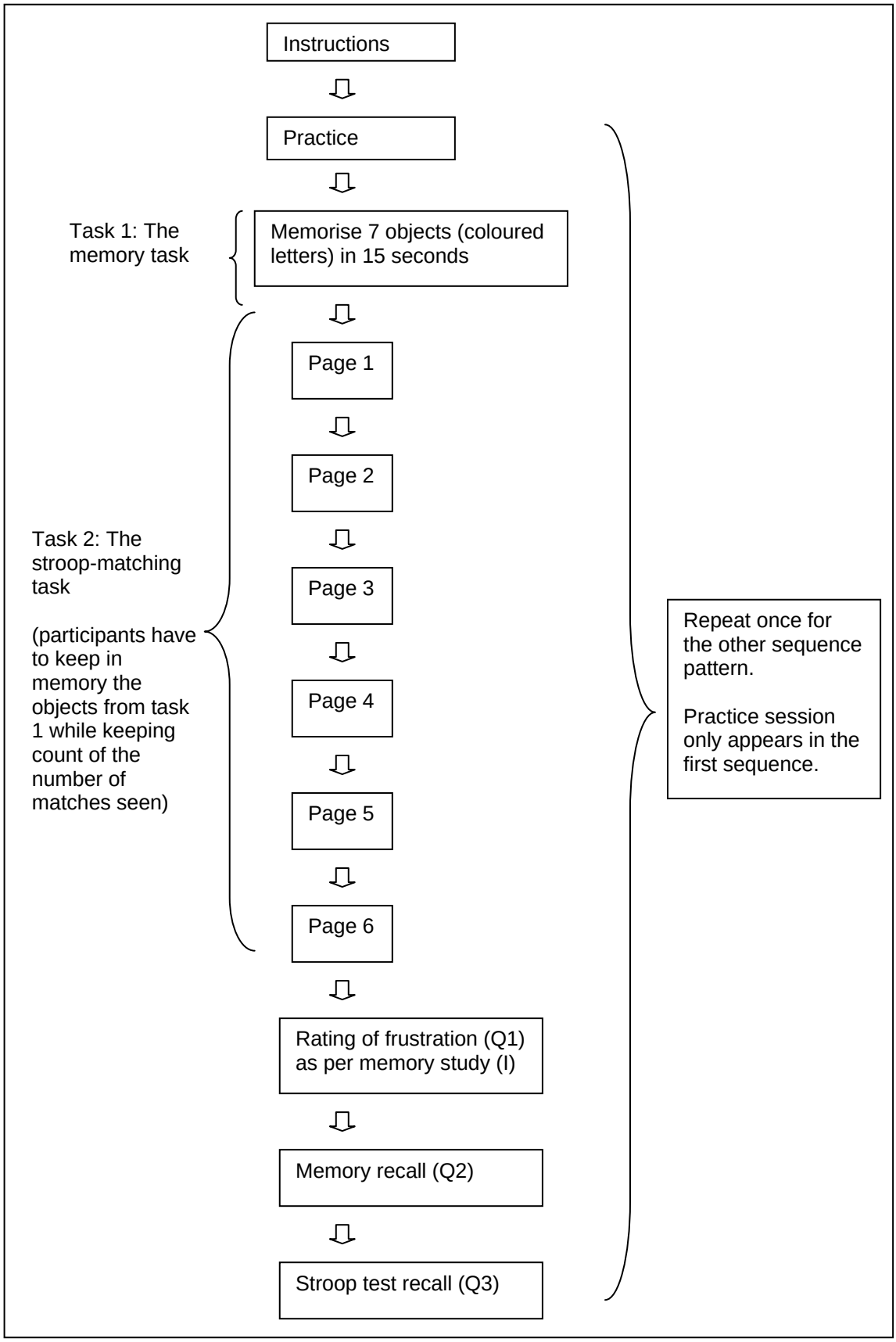
5.2.2.1 Participants

Twenty-two under-graduates (twelve male) at a large British university volunteered, and were each paid £4.00 as a participation fee. Their ages ranged from 18 to 25 years old. The entire study was computer-based.

5.2.2.2 Stimuli, dependent variables and procedure

Figure 5.2.2 below illustrates the flow of the study. Participants were required to complete a practice session to familiarise themselves with the study before starting. Participants were required to perform two sequences of tasks, one labelled the increasing delay sequence, and the other the decreasing delay sequence. Each sequence consisted of two consecutive tasks, the first of which was a memory task and the second was a series of stroop-matching tasks. Participants were allowed a time of 15 seconds to complete the first task, which was to commit the seven objects to memory (the objects used are reported in Table 5.3). Next, they have to these objects in memory for a further 79 seconds, while performing a series of stroop-matching tasks (i.e. the second set of tasks). There were six stroop-matching tasks (page 1 to page 6 on Figure 5.2.2), but the time allowed for each stroop-matching task differed according to the sequence pattern, increasing delay or decreasing delay. The parameters for these intervals are reported in Table 5.3.

Figure 5.2.2: Schematic layout of the memory (II) experiment



At the end of the sixth stroop-matching task, participants were required to evaluate the sequence using the rating instrument previously used in MS1 (see Figure 5.1.1a). Once this was completed, they were then required to recall as many of the objects as possible (see box Q2 in Figure 5.2.2), and record them using the instrument illustrated in Figure 5.2.3a. Next, they were required to write down (in box Q3 of Figure 5.2.2) the number of stroop-matches observed using the instrument presented in Figure 5.2.3b. A clock was included on each of the six pages in each sequence to make the time intervals between the six pages for both sequences more pronounced. The clock did not, however, indicated actual time, but instead rotated according to a fixed tempo.

Figure 5.2.3a: The instrument used to record the letters remembered (Q2)

Circle the coloured letters you remember.

X	T	O	H	(red)
X	T	O	H	(green)
X	T	O	H	(blue)
X	T	O	H	(black)

Figure 5.2.3b: Instrument used in the Stroop task (Q3)

How many coloured matches did you managed to count? ☐

Since each participant had to perform two sequences of memory and stroop-matching tasks in a within-participants design (i.e. one with increasing, and the other, with decreasing delay), the order in which the sequence patterns appeared in the study needed to be counter-balanced, where half the participants (selected at random) experienced the increasing delay before decreasing, while the other half, vice versa. After participants had completed the tasks for both sequences, they were then told that the study had ended, and were thanked for their efforts. The instructions of the study are reproduced in full in Appendix 5.2.

5.2.2.3 Experimental parameters

Table 5.3: The parameters used in the second (stroop-matching) task

	Page 1	Page 2	Page 3	Page 4	Page 5	Page 6
Increasing delay sequence						
Name of colour (task 2)	Grey	black	green	red	yellow	blue
Colour of letters (task 2)	Grey	black	green	red	yellow	blue
Duration of appearance (sec)	2	3	6	15	22	31
Decreasing delay sequence						
Name of colour (task 2)	Blue	black	grey	yellow	green	red
Colour of letters (task 2)	Blue	black	grey	yellow	green	red
Duration of appearance (sec)	31	22	15	6	3	2

The seven objects (coloured letters) of the memory task were based on one of the following two sets, counter-balanced between-participants: {blue T, red W, black X, green U, red Y, black Q, and blue Z} or {red Q, black T, green W, blue X, green Y, blue U and red Z}. The parameters of the stroop-matching task are reported in Table 5.3. The parameters were deliberately set such that the colours and name of colours in the stroop task always matched. This was because the primary function of incorporating this task into the study was to create the increasing and decreasing delay patterns in the sequences by varying the time allowed to complete each stroop-matching task. The stroop-matching tasks serve to segment the 79 second delay interval between memory (from the time task 1 ends in Figure 5.2.2) and recall (box Q1 in the same figure) of the seven objects to create a sequence of multiple delays. To expand on a point made earlier, the stroop-matching tasks were made trivial to ensure that this task does not confound the task loads for the two sequence patterns, increasing versus decreasing. It was assumed that by only requiring participants to keep a running count of the number of matches observed, this would ensure that the additional load on memory arising from the stroop-matching task are similar for both sequences. See Appendix 5.3 (a and b) for illustrative screenshots of the memory and stroop-matching tasks.

5.2.3 Results

The descriptive statistics for the three dependent variables measured, namely ratings of frustration, the performance in the first task (the memory of objects task) and the second task (the stroop-matching task) are reported in Table 5.4.

The first row of Table 5.4 presents the statistics (mean, median and standard deviation) for the measure of frustration from both sequence patterns. The second row presents the statistics for the correct number of objects recalled, while the third row presents statistics for the stroop-matches counted.

Table 5.4: Descriptive statistics of the dependent variables

	Mean	Median	Std Dev
Measure of frustration (Q1)			
Increasing delay sequence	62.5	70.0	30.2
Decreasing delay sequence	50.4	59.5	28.9
Number of correctly recalled letters and its associated colours (Q2)			
Increasing delay sequence	4.09	4.50	2.74
Decreasing delay sequence	4.27	4.00	2.64
Number of matches counted in the stroop-task (Q3)			
Increasing delay sequence	5.91	6.00	0.426
Decreasing delay sequence	5.73	6.00	0.550

Note 1: The figures for the measure of frustration are taken from the 0-100 scale used, where higher numbers indicate more frustration.
Note 2: For Q2, the total number of coloured letters is 7 for each of the sequences.
Note 3: For Q3, the total number of correct matches is 6 for each of the sequences.

The results suggest that the rating of frustration for the increasing pattern is higher than the decreasing, which is in line with the predictions made in H1. The results also suggest that task achievement for both the memory and stroop-matching tasks are independent of pattern, as predicted. In addition, achievement in the memory task was less than 100%, reflecting participants' inability to recall all of the objects, while achievement in the stroop-task was near 100%, reflecting the ability of most participants to correctly count all of the stroop-matches. These results were analysed further using ANOVA to provide statistical evidence to support these assertions.

A 2x2 ANOVA (pattern x order) was undertaken for all three dependent variables, ratings of frustration, achievement in the memory task, and achievement in the stroop-matching task. Pattern was a within-participants factor (increasing versus decreasing delays), where as order was a between-participants counter-balanced factor (increasing before decreasing delay versus decreasing before increasing delay).

The ANOVA calculations on the dependent variable frustration revealed a significant main effect of pattern ($F(1,20)=7.96$, $p<0.01$). There was, however, no significant main effect of order ($F(1,20)=0.017$, $p>0.05$), nor of an interaction between order and pattern ($F(1,20)=0.742$, $p>0.05$). These results supported the hypothesis H1 that participants would evaluate the increasing delay sequence more frustrating than the decreasing delay sequence.

The ANOVA calculations for the memory task did not reveal any main effects of pattern ($F(1,20)=0.111$, $p>0.05$) or order ($F(1,20)=0.198$, $p>0.05$), nor was the interaction between pattern and order significant ($F(1,20)=2.26$, $p>0.05$). These results implied that order was not a significant influence on performance in the memory task. More importantly, pattern did not influence task performance, which indicated that task achievement is independent of pattern. To test whether participants were able to fully recall all of the seven objects presented to them, t-tests were performed. The results of the tests showed that participants were unable to recall all of the objects (for the increasing sequence: $t=5.00$, $p<0.001$; decreasing; $t=4.87$, $p<0.001$), which suggested that the memory task was set at a level that exceeded participants' abilities to achieve it.

In the stroop-matching task, participants were required at the end of each sequence to provide the total number of stroop-matches that they had counted. As expected, the ANOVA calculations showed that the main effect of pattern on task achievement for the stroop-matching task was not significant ($F(1,20)=3.20$, $p>0.05$). In addition, neither the main effect of order ($F(1,20)=3.77$, $p>0.05$) nor of the interaction between order and pattern was significant ($F(1,20)=3.20$, $p>0.05$). Furthermore, participants, on average, reported that they had managed to count all six of the stroop-matches presented (i.e. a 100% record). This finding was supported by t-tests that compared the reported results with the actual number of stroop-matches presented ($t=1.00$, $p>0.05$ for the increasing delay sequence and $t=2.32$, $p>0.05$ for the decreasing delay sequence). This suggested that the level of the stroop-matching task was easy enough for the majority of participants to fully achieve the tasks required. From a statistical viewpoint, the fact that the

reported number of stroop-matches was not significantly different from the actual number demonstrated that the stroop-task did not interfere with pattern and had served its purpose of segmenting the sequence into increasing and decreasing patterns of delay. A summary of key results for Memory Studies (I) and (II) (MS1 and MS2) was reported in Table 5.5.

Table 5.5: Summary of key results for Memory Studies (I) and (II)

	F-test (ANOVA)	p-value
Memory study (I)		
Dependent variable: Frustration Main effect of pattern	F(1,16)=20.6	p<0.001
Dependent variable: Memory task Main effect of pattern	F(1,16)=0.968	p>0.05
Memory study (II)		
Dependent variable: Frustration Main effect of pattern	F(1,20)=7.96	p<0.01
Dependent variable: Performance in the memory task Main effect of pattern	F(1,20)=0.111	p>0.05
Dependent variable: Performance in the stroop-matching task Main effect of pattern	F(1,20)=3.20	p>0.05

5.2.4 Discussion for MS1 and MS2

The results of the memory studies supports the hypothesis presented in Model 1, where it was predicted that people would rate the decreasing delay sequence less aversively than the increasing sequence. This model was developed from the finding that the end-loaded pattern effects reported in the sequence literature could be used to predict behaviour in sequences of delays. Model 1 posited that because delays obstructed progression and task achievement, it would generate negative affect in the form of increased frustration with the service provider. By extrapolation, it was predicted that a sequence of increasing delays would be evaluated less favourably compared to a decreasing sequence. This was because people have asymmetric responses towards the pattern of delays, such as in end-weighting aversive sequences. This meant that a sequence of increasing obstruction will be rated

more aversively than one with decreasing obstruction, for sequences with identical integrated utility (i.e. equal total delays).

The Memory Study (I) (MS1) was designed to obstruct the memory task by making it impossible for people to recall all of the objects, through the inclusion of delays in the task. It was expected that the delays would obstruct the memory task, because people have to commit extra effort to keep the items in memory during the delay period, compared to a situation without delay. As a result of the increased difficulty, people then become more frustrated with the sequence.

The MS1 study, however, has only utilised one method to generate task obstructions that induces negative emotions in sequences. There may be other methods available that can also achieve the same task manipulation of inducing negative emotions. It is therefore important that further studies are undertaken to provide additional support for the model proposed in this study. This is because the model (i.e. Model 1) proposed in the current memory study presumes that the asymmetric sequence evaluations are a result of sequence evaluations being end-weighted. It cannot be concluded from this study, however, that people will always possess such characteristics when they evaluate other similar sequences that involve task obstructions.

The purpose of the next two studies, Visual Search (I) and (II) or VS1 and VS2, is to develop a set of tests to provide further supporting evidence of the hypothesis developed in Model 1. A confirmatory result from VS1 and VS2 would support the finding of the memory studies that for a sequence of time-sensitive tasks, it was the degree to which task achievement was obstructed (i.e. was full task achievement prevented?) that induces evaluation based on patterns. The role of time in this case is that of delays, which caused the obstruction.

5.3 The Visual Search Study (I) or VS1

5.3.1 Introduction

The results of the previous memory studies MS1 and MS2, and the IDS had suggested that the preference for an improving/decreasing sequence of delays would only be observed if and when the delays obstructed task achievement. If this observation is to be reconciled with the predictions of Model 1, this would mean that delays are only one of a number of ways in which obstruction of task achievement could be generated, and Model 1 had to be modified to take into account of this. Recalling the model:

V(d1, d2, …, dn)>V(dn, …, d2, d1) … [Model 1]

Since Model 1 only holds if the delay causes obstruction, therefore it is proposed that the delay in the equation is replaced by a more generic factor that may generate obstruction in a sequence of tasks, and one that also possesses the monotonic characteristics of delay (i.e. longer delays lead to more negative emotions, therefore more of the obstructing 'factor' should also lead to more negative emotions. Let O=any factor that obstructs task achievement, and that for O2>O1, V(O2)>V(O1). Therefore:

V(O1, O2, …, On)>V(On, …, O2, O1) … [Model 2]

Model 2 states that the evaluation of a sequence with increasing obstruction (larger numbers indicate greater obstruction) is more aversive than a sequence with decreasing obstruction. The objective of this study, VS1, therefore is to provide further empirical support for this observation (i.e. where the obstruction of task achievement, rather than delays per se, is the key factor that induces evaluation based on patterns). This study uses a different method from the memory studies to generate a sequence of obstructed tasks. Another way to generate obstruction in a sequence of tasks is to impose time constraints. The inspiration for this manipulation was from ideas originating from the time constraints literature, which is described below. Note that it is not the intention

of the study to test any hypotheses or make any claims relating to the time constraints literature, as this literature does not directly relate to the purpose of this study. The reference to the time constraints literature was only to demonstrate that the imposition of time constraints in a sequence of tasks make a viable tool of manipulating the degree to which sequences of tasks are obstructed.

Time constraints was defined as the difference between the amount of time available and the amount of time required to achieve a particular decision task (Benson and Beach, 1996; Rastegary and Landy, 1993), therefore the less time available to complete a task, the more time constrained a person would be. There are three characteristics that make time constraints as a method of manipulating task obstruction in sequences ideal. First, research had shown that time constraints can negatively affect the ability of individuals to make effective decisions or to achieve a task (Maule and Edland, 1997; Svenson and Maule, 1993), which suggests that it is possible to alter the level of task achievability by systematically varying time constraints. Second, time constraints have also been shown to induce changes in people's affective states in addition to feelings of time pressure, such as anxiety and energeticness, and in the information processing strategy they used for decision-making (Maule, Hockey, and Bdzola, 2000). The implication of this second characteristic suggested that time constraints could also be used to invoke negative emotions in people, which is an important consideration, given that the hypothesis developed in Model 1 was predicated on the evaluation of aversive sequences. Third, decision-makers react to the imposition of time constraints by applying simple heuristics for evaluation and choosing strategies that require less information processing (Gigerenzer and Todd, 1999). As suggested in the literature review in chapter 4 (see section 4.1.3), one such heuristic that might be used for the evaluation of time constrained sequences is the reliance on patterns for evaluation, which meant that sequences with more attractive end-loaded weights would be favoured over sequences with other weight distributions (see Ariely and Carmon, 2003). Given that these three characteristics described would make time constraints an ideal manipulation of task obstruction, Model 2 is re-written as:

V(TC1, TC2, …, TCn)>V(TCn, …, TC2, TC1) … [Model 3]

TC are time constraints, where for TC1>TC2, it follows that V(TC1)>V(TC2). This inequality says that an increase in the severity of the time constraints (i.e. less time to complete a task) would also increase the level of task obstruction, thus generating more (dis) utility, or negative emotions. The resultant evaluation would be where a sequence of increasing time constraints would have a more unfavourable evaluation than a decreasing time constraints sequence. The hypothesis from Model 2 was rewritten to take into account of the modifications to the model, which was presented in Model 3, and was restated as follows.

H1: The sequence of increasing time constraints would be rated more unfavourably than a sequence of decreasing time constraints.

5.3.2 Method

5.3.2.1 Participants

Twenty-three under-graduates from a large British university volunteered for the study and were each paid £4.00 for their participation. Their ages ranged from 18 to 25 years old. Eleven of the participants were females.

5.3.2.2 Introduction to visual search

This section provides a brief introduction to the visual search literature. This introduction is essential, because visual search tasks will be an essential part of the task manipulation used in this and the next study in this chapter. The objective of this introduction is to describe the various types of controls that can be achieved by manipulating the complexity of visual search tasks in order to make a visual search task easier or more difficult. Treisman (1982) is the major reference used to create the visual search manipulations in this study, and all findings described are attributed to this source.

Treisman (1982) found that visual objects can be more difficult to find if they share either a similar shape or colour to a group of objects next to them. She

called this effect the ‘conjunction illusion’. For example, a red X is more difficult to find compared to a red O if it is placed amongst green Xs. This is because the red X shares a similar shape with other objects. Additionally, the red X will also be more difficult to find compared to a blue X if it is placed amongst red Os, because of the similarities in colours between the target object (the red X) and the background of objects (red Os). In an extension to these basic findings, Treisman (1982) also found that objects ‘pre-attentively grouped’ are easier to find than those that are not. Pre-attentive grouping here is defined as the process where one object is made to stand out from other objects, either because it does not share the same shape and/or the same colour as objects that surround it. See array 1 in Figure 5.3.1 for illustration. The red X here stands out because it does not share the same shape (X versus H) or colour (red versus green) compared to the other objects that surrounds it, which are all green Hs. The green Hs surrounding the red X act as a reference to help the person group more quickly the objects that are similar, which then allows him/her to find the odd object more easily. In contrast, however, consider the objects in the top left hand quadrant of array 2 in Figure 5.3.1. The target object (the red X) is now embedded in an array of green Xs and red Hs. The search for the target object here is more difficult, for two reasons. The first is because the target object is not pre-attentively grouped, unlike array 1. The second is attributed to the ‘conjunction illusion’ described earlier, where the target object becomes more difficult to find because it shares either a colour or a shape with the objects that surround it.

In addition to the concepts of pre-attentive grouping and conjunction illusion, Treisman also reported a number of other controls that experimenters can manipulate to make the objects more prominent or more difficult to find. Some of these controls are obvious. For example, bigger objects are easier to find than smaller objects, and the more embedded the target object is with its background, the more difficult it will be to find it (i.e. it will require more time). The extent to which a target object is embedded into its background is determined by the number of other objects that surround a target object – the more the number of surrounding objects, the harder it will be to find the target object. In other words, difficulty in visual search tasks is a function of the time

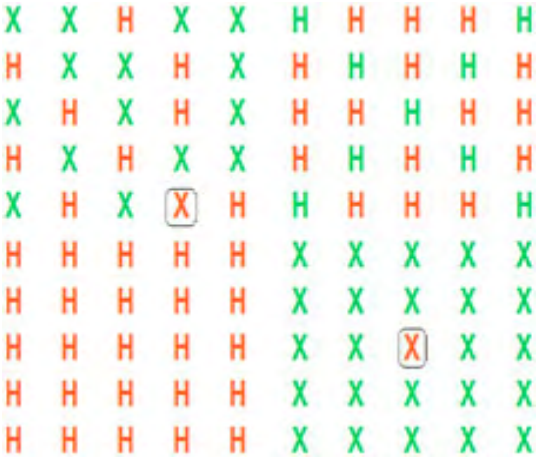
available to complete the tasks. The less time available, the more difficult the task is, and vice versa. Pre-attentive grouping of objects make the task easier, because it speeds up the visual search process. In contrast, conjunction illusion makes the task harder, because the more embedded the objects are among similar objects, the more it becomes necessary to scan each array of objects on a row-by-row or column-by-column basis, which is time-consuming (Treisman, 1982). This means that for embedded objects, the search becomes more difficult as more rows and/or columns are added to the array of surrounding objects. This is illustrated by comparing array 4 with array 5: Array 4 has eleven rows by eleven columns, whereas array 5 has fifteen rows by fifteen columns. Both arrays require participants to use a line-by-line scan strategy (whether on a row-by-row or column-by-column basis), because the objects are not pre-attentively grouped (unlike in array 1, for example). It is more difficult to find the target objects in array 5, compared to array 4, because the line-by-line scan for the objects would take longer due to the larger number of objects in each row and column.

5.3.2.3 Stimuli and experimental design

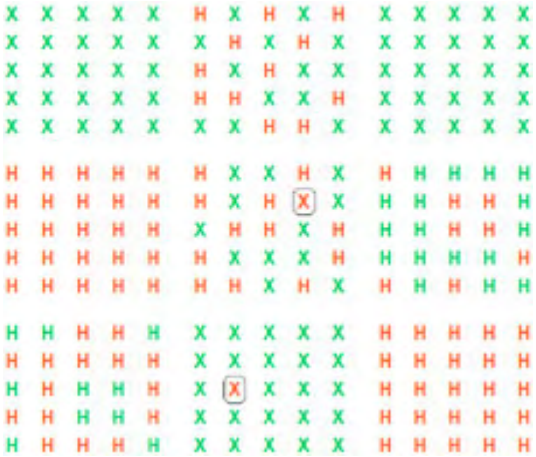
The time-constrained task that participants were required to perform was described as follows. Participants have to search for target objects that were embedded in an array of similar objects. The difficulty of the visual search task was manipulated by utilising a number of controls described above, such as by varying the size of the objects, by introducing pre-attentively grouped objects, and by varying the density of the background (i.e. number of other objects surrounding the target object). Apart from these controls described, task difficulty is also a function of time. The more time allowed completing each task, the easier the task and vice versa versa. For the purposes of this study, only the first set of controls (size, pre-attentive grouping, and background density) was used to manipulate task difficult to create sequences of increasing versus decreasing time constraints. It was not necessary to also utilise time to manipulate task difficulty, as the first set of controls is sufficient for the purposes. The use of time as a manipulation was only utilised in the next study, visual search (II).

Figure 5.3.1: The complete set of stimuli used for visual search (I)

Array 1

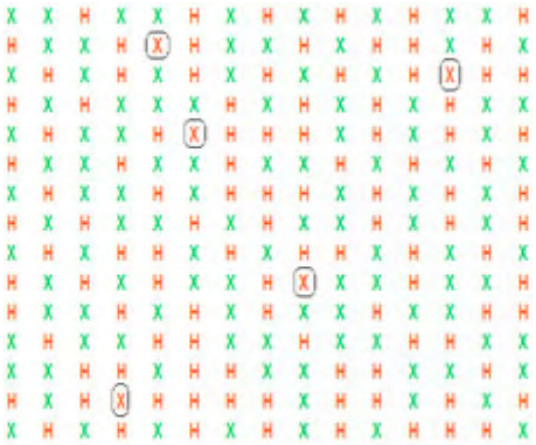


Array 2



Array 3

Array 4



Array 5

Note:

The sizes of the arrays displayed on the actual computer test screen had to be compressed and slightly distorted for it to fit on this page.

The task manipulations to create the sequences of increasing and decreasing time constraints are as follows. The time allowed for task completion for each array is fixed at 5 seconds. The arrays are increasing in difficulty if the sequence progresses from array 1 to array 5, but decrease in difficulty if the sequence progression is from array 5 to array 1. Because time is fixed for all arrays, however, therefore any increase in difficulty is not compensated by an increase in the time allowance for task completion, which means that the tasks would become increasingly time constrained. The opposite is also true, where a sequence of tasks that is decreasing in difficulty is becoming less time constrained because of the additional time allowed to complete the next task, relative to the current or previous tasks. Therefore, the sequence {Array 1, Array 2, Array 3, Array 4, Array 5} is a sequence of increasing time constraints, and {Array 5, Array 4, Array 3, Array 2, Array 1} is a sequence of decreasing time constraints.

The task difficulty manipulation was undertaken as follows. The task in array 2 was more difficult than array 1 because the target object was pre-attentively grouped in array 1, the size of objects were smaller, the background was more dense (i.e. filled with more surrounding objects), there were more target objects to find, and one of the target objects was surrounded by similar objects, creating a conjunction illusion. The task in array 3 was more difficult than array 2, because it was denser. Array 2 had four groups of 5x5 objects arranged in a square grid, while array 3 had nine groups of 5x5 objects in a square grid. In addition, each object in array 3 was also smaller in size than those in array 2. The task in array 4 was more difficult than array 3, because the entire 11x11 grid was not pre-attentively grouped, unlike array 3, where seven out of the nine groups of 5x5 objects was pre-attentively grouped. This made it more difficult to find the target object due to the stronger conjunction illusion effect. Finally, array 5 was more difficult than array 4, because it was denser. Array 4 had an 11x11 grid of objects, while array 5 had a 15x15 grid of objects.

There were 13 target objects (red X) in total embedded into all of the arrays. The choice of objects and the inspiration for the grids and conjunction illusion effect between objects were based on Treisman's (1982) work. The final design was pilot tested using four volunteers who did not participate in this study. The parameters (i.e. time allowed and degree to which the target object was embedded) were chosen such that participants would not be able to fully achieve the task of finding all the target objects embedded. This manipulation was identical to that used in the previous memory studies. The layout of the experiment was illustrated in Figure 5.5 (see Appendix 5.10 for an illustration of the stimuli with the computer-based interface).

It is highlighted that this form of manipulation used to generate task obstruction in this study is completely different from the previous memory studies, where time was used as a manipulation to delay task completion. In the context of the present study, however, it is the type of task that participants have to perform during the time interval that is being manipulated, because the intervals are all of the same length. This is a novel approach to manipulating task obstruction in sequences that involve time.

Obstruction increases when the degree of camouflage increases, while time allowed to achieve the task was constant. Therefore, the sequence of increasing obstruction was represented by the sequence with the order {Array 1, Array 2, Array 3, Array 4, Array 5}. Similarly, a sequence of decreasing obstruction was created by re-ordering the sequence such that the level of camouflage decreases, which was the sequence order {Array 5, Array 4, Array 3, Array 2, Array 1}. The time that was allowed for the completion of the task within each array was 5 seconds. There were 13 target objects (red X) in total embedded into all of the arrays. The parameters, such as the number of target objects, colours and shapes to use, sizes on the computer screen and spacing, were all developed from the findings of Treisman (1982), and was pilot tested using four volunteers who did not participate in this study. The parameters (i.e. time allowed and level of camouflage) were chosen such that participants would not be able to fully achieve the task. This manipulation was identical to that used in the previous memory studies. The layout of the experiment was

illustrated in Figure 5.3.3 (see Appendix 5.10 for an illustration of the stimuli with the computer-based interface).

The overall design of the study was a 2 (pattern) x 2 (order) mixed model. Pattern was within-participants, and consisted of an increasing versus decreasing sequence of obstruction. Order was a counter-balanced factor, with twelve participants randomly assigned to the increasing before decreasing obstruction condition, and the rest to the decreasing before increasing condition.

5.3.2.4 Dependent variable used

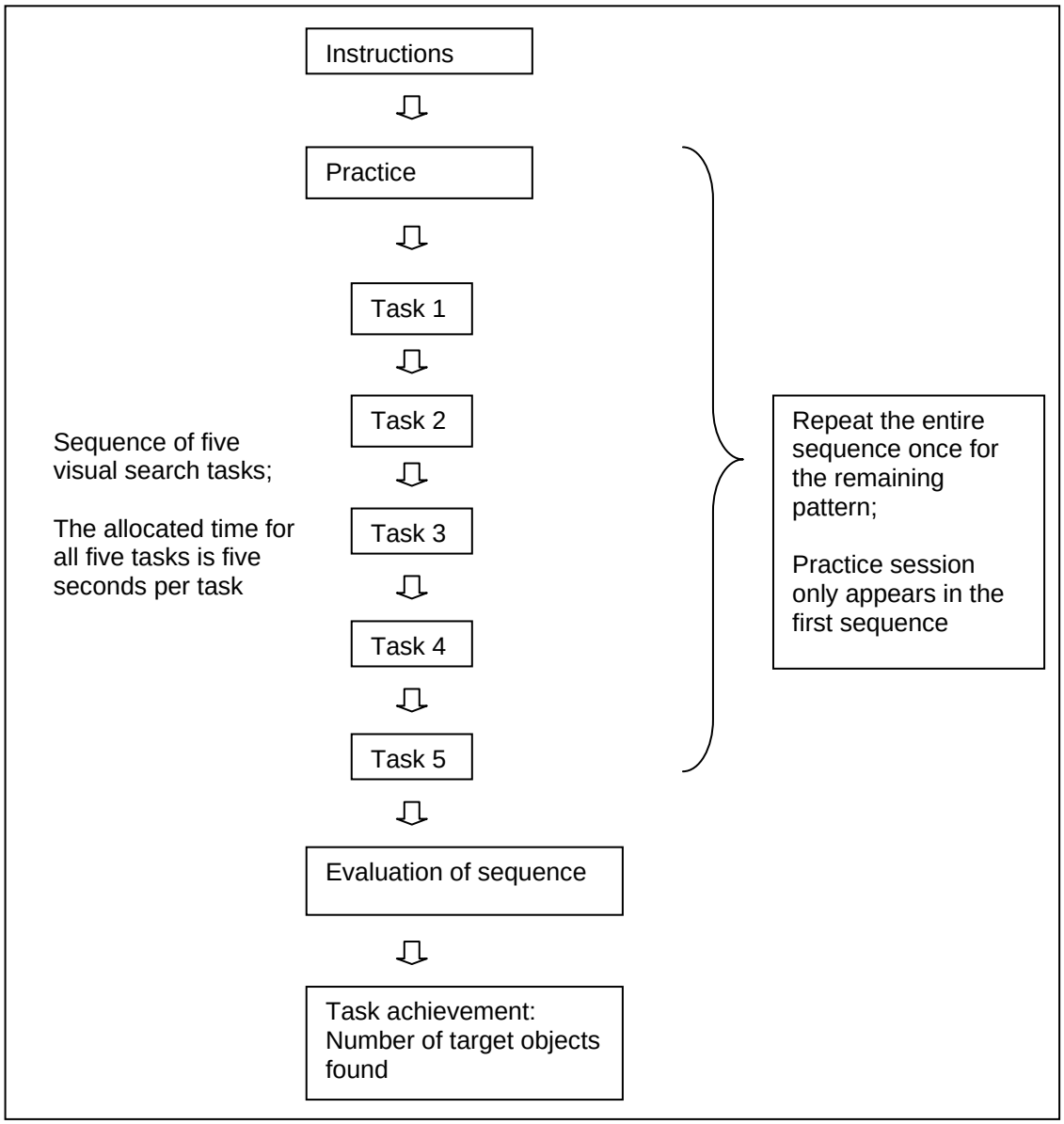
There were two dependent variables recorded, which were shown in Figure 5.3.2. The first dependent variable involved participants evaluating how annoying the (visual search) task was, on a 0-100 scale, with 0=not annoying at all and 100=very annoying. This dependent variable was very similar to that used in the memory studies, except for a change in terminology, from frustration to annoyance. Annoyance was the term first used in the IDS. This change was made based on the findings from a pilot study which suggested that annoyance was the most suitable term to use as a measure of the negative emotion caused by the visual search task. The current scale and use of the emotional term 'annoyance' replicates that used in Ariely and Loewenstein's (2000) study. The second dependent variable required participants to recall the number of signals detected, once they had evaluated their annoyance from the visual search tasks.

Figure 5.3.2: The rating instruments used to measure the dependent variable for both visual search studies

Q1. Overall, how ANNOYING was the task?
Draw a small vertical line on the scale. 0 100 Not annoying at all Very annoying
Q2. Write the estimated total number of red X here.

Note: The length of the 0-100 line above was 15.2 cm in the actual experiment.

Figure 5.3.3: The experimental layout and flow of visual search (I)



5.3.2.5 Experimental procedure

In Figure 5.3.3, each box represented the experimental interface, which was a web page. The first page that participants encountered was the instructions page, which explained to them the procedures for the study, which was to find the target objects (i.e. red Xs). Participants were then given an opportunity to familiarise themselves with the tasks required before starting the actual visual search. Once participants had read and understood the instructions, they were then told that they could begin the study. The full details of instructions for this study were presented in Appendix 5.5. After participants had completed the

sequence of visual search tasks, they were then required to provide their evaluations using the dependent variables described above, before they were allowed to move on to the next sequence of tasks. At the end of the final sequence, they were then told that they had completed the entire experiment, and were thanked for their involvement. At this point, the attendance fee was paid to them, and they were encouraged to give some informal feedback as to what they thought about the experiment.

5.3.3 Results

The presentation of the results of this study is as follows. First, the descriptive statistics of the study are reported in Table 5.6. Second, tests are performed to determine if there are any order effects. Third, tests of the hypothesis H1 are conducted.

Table 5.6: Descriptive statistics of annoyance ratings and the number of target objects found (task achievement)

	Dependent variable measured			
	Annoyance ratings		Number of target objects found	
	Increasing obstruction	Decreasing obstruction	Increasing obstruction	Decreasing obstruction
Mean	49.9	33.5	8.17	8.65
Std Dev	26.0	20.3	2.08	2.79

Note 1: The figures for the measure of frustration are taken from the 0-100 scale used, where higher numbers indicate more annoyance.
Note 2: The total number of target objects embedded in each sequence was 13.

The direction of the findings reported in Table 5.6 support hypothesis H1 with the rating of annoyance higher for the sequence of increasing time constraint (49.9) than for the sequence of decreasing time constraint (33.5). To determine the significance of the findings, a 2x2 (order) x (pattern) mixed model ANOVA was first performed, as in the previous IDS, MS1 and MS2. Pattern was within-participants and order was between-participants. The results did not reveal any significant main effect of order ($F(1,21)=0.373$; $p>0.05$) nor any interaction between order and pattern ($F(1,21)=0.674$; $p>0.05$). There was, however, a significant main effect in the hypothesised direction ($F(1,21)=16.5$; $p<0.001$) for the dependent variable annoyance. For

the dependent variable task achievement, however, no significant main effect nor interaction was present (pattern: $F(1,21)=0.566$; $p>0.05$; order: $F(1,21)=0.28$, $p>0.05$; pattern x order: $F(1,21)=0.392$, $p>0.05$). Table 5.7 presented some of these results in tabular format for ease of future reference. These results supported the general finding and argument from the memory studies (MS1 and MS2) that pattern does not materially affect task achievability, as achievability in the visual search task should only be influenced by the total amount of time available to complete the task, which was identical for both sequences in this case.

Table 5.7: Summary of test results: main effects

	F-test (ANOVA)	P
Dependent variable: Annoyance	$F(1,21)=16.5$	$p<0.001$
Dependent variable: Task achievement	$F(1,21)=0.566$	$p>0.05$

Given that the manipulation of the degree to which time constraints obstructed achievement in the visual search tasks was an important consideration in the development of Model 3's hypothesis, a t-test on task achievement was calculated to test whether participants were able to fully achieve the visual search tasks. The results of the test showed that participants were not successful in finding all of the embedded target objects (increasing obstruction sequence: $t=11.1$, $p<0.001$; decreasing; $t=7.47$, $p<0.001$). This demonstrated that the manipulation of task achievability in this study was successful in creating a task that exceeded participants' ability, where the number of target objects that participants had found was significantly less than the number of target objects embedded. This result is important, because the negative emotions that are induced in the sequences of tasks in this study is contingent on the tasks being difficult (see the conditions of Model 2 in section 5.3.1).

5.3.4 Discussion

The results of this study support the hypothesis that in a sequence of time-constrained tasks, the pattern of decreasing time constraints was rated less aversively than an increasing pattern. The time constraints in the sequence of

tasks were organised in such a way that they obstructed task achievement, and it was posited that this obstruction led to the generation of the negative emotion annoyance. The concepts used to develop Model 3 argued that the evaluation based on patterns was induced because people perceived the sequence of decreasing obstruction as an improving sequence of aversive events, while the sequence of increasing obstruction was perceived as a deteriorating sequence, leading to improvement being preferred.

The results of this study have suggested that it is the obstruction and changing levels of task difficulty, rather than delays per se, which induces people to evaluate a sequence of tasks based on its patterns. This assertion was based on the participant-task manipulation used in this study, where time constraints were used to obstruct, and in so doing invoke negative emotions. In the previous memory studies, task obstruction and the subsequent negative emotions invoked were generated from the manipulation of delay parameters. If it is accepted that obstruction and changing levels of task difficulty are the main factors that induces people to evaluate a sequence based on an end-loaded weighting of its pattern (i.e. prefer sequences with the most attractive peak-end or trend characteristics), this could partially explain the findings of the IDS where the main effect of pattern was not significant. The task achievability in the IDS was not in any way obstructed by the delays, or by any other factor, and was always achievable for all participants. This in turn allowed participants to consider a wider range of factors when evaluating the sequence, for example, domain-specific characteristics such as the initial shopping experience and expectations, where front-loaded as well as back-loaded weighting of pattern was a possibility.

There is, however, one caveat to the conclusions reached in the Memory Studies (MS1 and MS2) and the current study (VS1). While it was demonstrated that evaluation based on patterns could be induced by obstructing the tasks to create sequences with different patterns of task difficulty, such as increasing and decreasing difficulty, there is an implicit assumption made in these three studies (MS1, MS2, and VS1) regarding people's evaluations of tasks where the patterns of task difficulty remains at a

constant level. It was assumed that if the level of task difficulty of the memory tasks in MS1, MS2 or VS1 were simple enough to the extent that most/all participants will fully achieve all tasks within each sequence of tasks, people would not perceive the delays or time constraints as an obstruction. This means that the level of negative emotions invoked by the task obstructions may be insufficient to induce evaluations based on patterns. The purpose of the next and final study in this chapter is to test the hypothesis that task achievability is indeed an indispensable factor that leads to sequences of tasks being evaluated based on patterns.

5.4 The Visual Search Study (II) or VS2

5.4.1 Introduction

This study tests the assumption implicit in the previous three studies that an evaluation based on patterns is in response to changing levels of task difficulty. There are two plausible reasons for why this is so. The first is that when the tasks are made simple enough such that the task difficulty does not vary with the changing durations (i.e., people will always have enough time to complete the tasks), people experience little or no negative emotions as a result. This means that the previous models developed (model 1, 2, and 3) would not be applicable, as these models are dependent on people experiencing a sequence of increasing or decreasing negative emotions that are induced by task obstructions. The second suggests that the disruptions caused by more difficult tasks (compared to easier tasks) may have made it harder to evaluate a sequence, leading people to use a different strategy to evaluate sequences, from a more complex strategy such as utility integration, to a simpler heuristic-based strategy, such as evaluation based on patterns. The switch to a heuristics-based strategy can occur in situations where people have little incentive or cognitive ability to work out the utility integral of a sequence, as argued in Ariely and Carmon (2003). It is suggested that memory tasks and visual search tasks are examples of such situations.

In the current study, it is proposed that the condition that would trigger the switch of evaluation strategies would be task achievability, where evaluations based on pattern would only be induced when the tasks are difficult to achieve (i.e., where participants are not given sufficient time to complete the task). The stimuli used to test the hypothesis in this study were adapted from the stimuli of the previous visual search study, which is further described below. The design of the second visual search study was a two-factor (2x2) repeated-measures model, with pattern (increasing versus decreasing obstruction) being one of the factors, and task achievability (simple versus difficult) the other. Ariely and Loewenstein (2000) (see also Ariely et al., 2000, and the previous literature review in section 4.1.1.2 of chapter 4) had argued that a repeated-

measures procedure is a better experimental design compared to a between-participants procedure for research in sequences, because people are better able to evaluate sequence duration when they can compare between different sequences. Ariely and Loewenstein (2000) stated that comparability between sequences is important for sequence evaluations in general, because such evaluations are often made on the basis of prior experiences. The hypothesis of the study is as follows.

H1: The increasing obstruction sequence will only be evaluated more aversively than the decreasing sequence when the tasks are difficult.

5.4.2 Method

5.4.2.1 Participants

Ninety-six under-graduate students volunteered for the study. Half of the students were from a large a British university, and the other half from a Malaysian university. British participants were paid £3.00, while Malaysian participants were paid RM$10 (approximately £1.80 at the time when the study was conducted) as a participation fee. Their ages ranged from 18 to 25 years old. In addition, performance incentives in the form of music vouchers and cash prizes were also offered to the best four participants in both groups to motivate participants to perform to the best of their ability. Performance was judged on the accuracy of the search for the target objects in the visual search task, with the responses that were closest to the actual answers getting the prizes.

5.4.2.2 Stimuli and parameters

The task that participants were required to perform was to search for target objects that were embedded in an array of similar objects, and to do this under time constraints. The visual search tasks were presented as a sequence of six activities, with varying parameters, in terms of the degree of camouflage of the target object, and the time constraints. All participants were required to perform four sequences of visual search tasks. The four sequences, however, were different in terms of their parameters, and represented the conditions in the 2x2 repeated-measures factorial design. The first two sequences had

increasing and decreasing time constraints with the task level set at 'simple.' The next two sequences also had identical increasing and decreasing time constraints, but where the task level was set at 'difficult.' These four sequence conditions were coded {increasing-simple}, {decreasing-simple}, {increasing-difficult}, and {decreasing-difficult}}. Since participants had to undertake all four sequence conditions, the order in which the four sequences appeared had to be counter-balanced using a rectangular arrays procedure, of which there are (2x2)!=24 possible permutations of order (See Table 5.8 for an illustration). For participants in each country, a total of 48 participants were recruited and randomly allocated to make up the two (48/24=2) complete sets of rectangular arrays. This gave four complete rectangular arrays when the participants from both countries were combined.

Table 5.8: Rectangular array design and codification for the Visual Search (II) study (VS2)

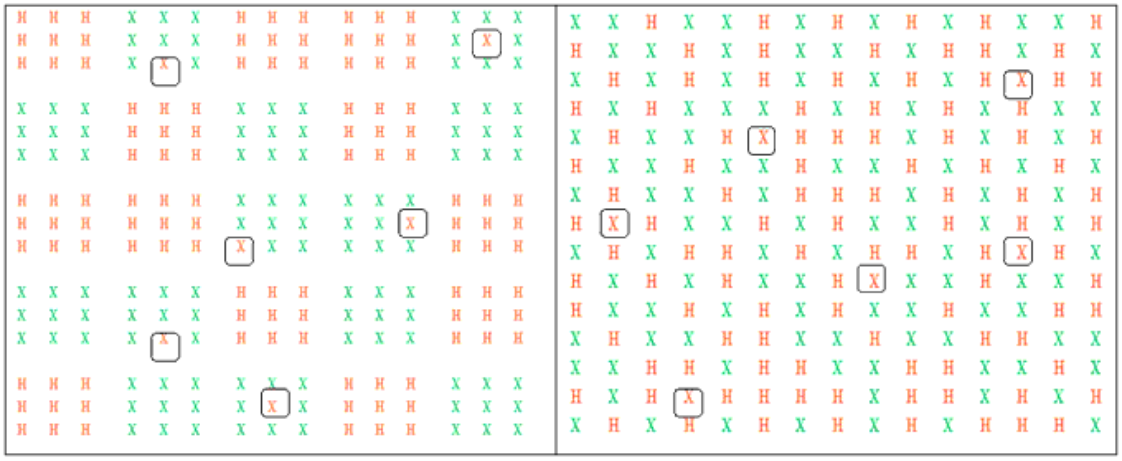
Code	Order of presentation 1^{st} 2^{nd} 3^{rd} 4^{th}	Code	Order of presentation 1^{st} 2^{nd} 3^{rd} 4^{th}
1	ABCD	13	CABD
2	ABDC	14	CADB
3	ACBD	15	CBAD
4	ACDB	16	CBDA
5	ADBC	17	CDAB
6	ADCB	18	CDBA
7	BACD	19	DABC
8	BADC	20	DACB
9	BCAD	21	DBAC
10	BCDA	22	DBCA
11	BDAC	23	DCAB
12	BDCA	24	DCBA

Key to four sequence conditions:
A = increasing time constraints & simple task level
B = decreasing time constraints & simple task level
C = increasing time constraints & difficult task level
D = decreasing time constraints & difficult task level

The sequence of time-constrained visual search tasks that participants were required to perform was similar to that used in the previous visual search

study. However, a number of adaptations were made to incorporate the manipulation of the additional factor task achievability (i.e., simple versus difficult level). The variations in the complexity of the camouflage, which was used to create five different levels of task difficulty in the previous VS1, was now used to manipulate the factor of task achievability, which only required two different levels. The manipulation is illustrated in Figure 5.4.1. The stimuli for the simple and difficult condition was constructed using the visual search manipulations previously used VS1 (i.e. pre-attentive grouping and the conjunction illusion), and described in section 5.3.2.2 and 5.3.2.3.

Figure 5.4.1: Illustration of task level manipulation in VS2



Note: The left hand panel demonstrated the stimuli interface in the simple condition, while the right hand panel demonstrated the interface in the difficult condition. The target objects were red X (circled).

There are two ways in which task difficulty for the current visual search study (VS2) was manipulated. The first manipulation varied the complexity of the stimuli, while the second manipulation varied the time allowed to complete the task. The complexity of the visual search stimuli was varied to construct the two different levels of task difficulty, simple and difficult. The left hand panel in Figure 5.4.1 represented an illustration of one of the six tasks under the simple task condition, while the right hand panel represented the difficult task condition. Both panels consisted of similar number and size of target and background objects in a 15 row by 15 column grid. However, the target objects in the left hand panel are much easier to find compared to those in the

right hand panel, as they are pre-attentively grouped (e.g. red X embedded among green X), while the target objects in the right hand panel are mixed with other objects that share similar characteristics such as shape or colour, thus creating a conjunction illusion (c. Treisman, 1982).

The second manipulation was also based on Treisman's (1982) finding that task difficulty in visual search is inversely proportional to the amount of time available to find the target object(s). The manipulation of increasing versus decreasing time constraints was achieved by varying the time allowed for task completion. In VS1, this parameter was set constant at 5 seconds for all individual tasks within a sequence, but for the current study, this was varied to create patterns of increasing and decreasing time constraints. For the sequence of increasing time constraints, the time allowed to find the target object(s) for each of the six tasks were {22, 15, 10, 6, 4, 3} seconds (i.e. people have less and less time to perform the task as the sequence progressed, thus becoming increasingly time constrained). For the sequence of decreasing time constraints, the parameters used were {3, 4, 6, 10, 15, 22} seconds.

These parameters, which were derived from the same pilot study as used in the previous visual search study, were chosen because in the simple task condition, the stimuli were designed such that participants were able to find all of the target objects even at its most time constrained moment (i.e. 3 seconds). In the difficult task condition, however, the stimuli was designed such that participants would not be able to find all of target objects even at its least time constrained moment (i.e. 22 seconds). For each visual search task within the sequence, the distribution of embedded target objects (red X) per task within each sequence ranged from 3 (minimum) to 5 (maximum), but the total number of embedded signals for each sequence always added up to 24 (i.e. an average of 4 red Xs * 6 tasks in each sequence gives 24).

5.4.2.3 Dependent variable used

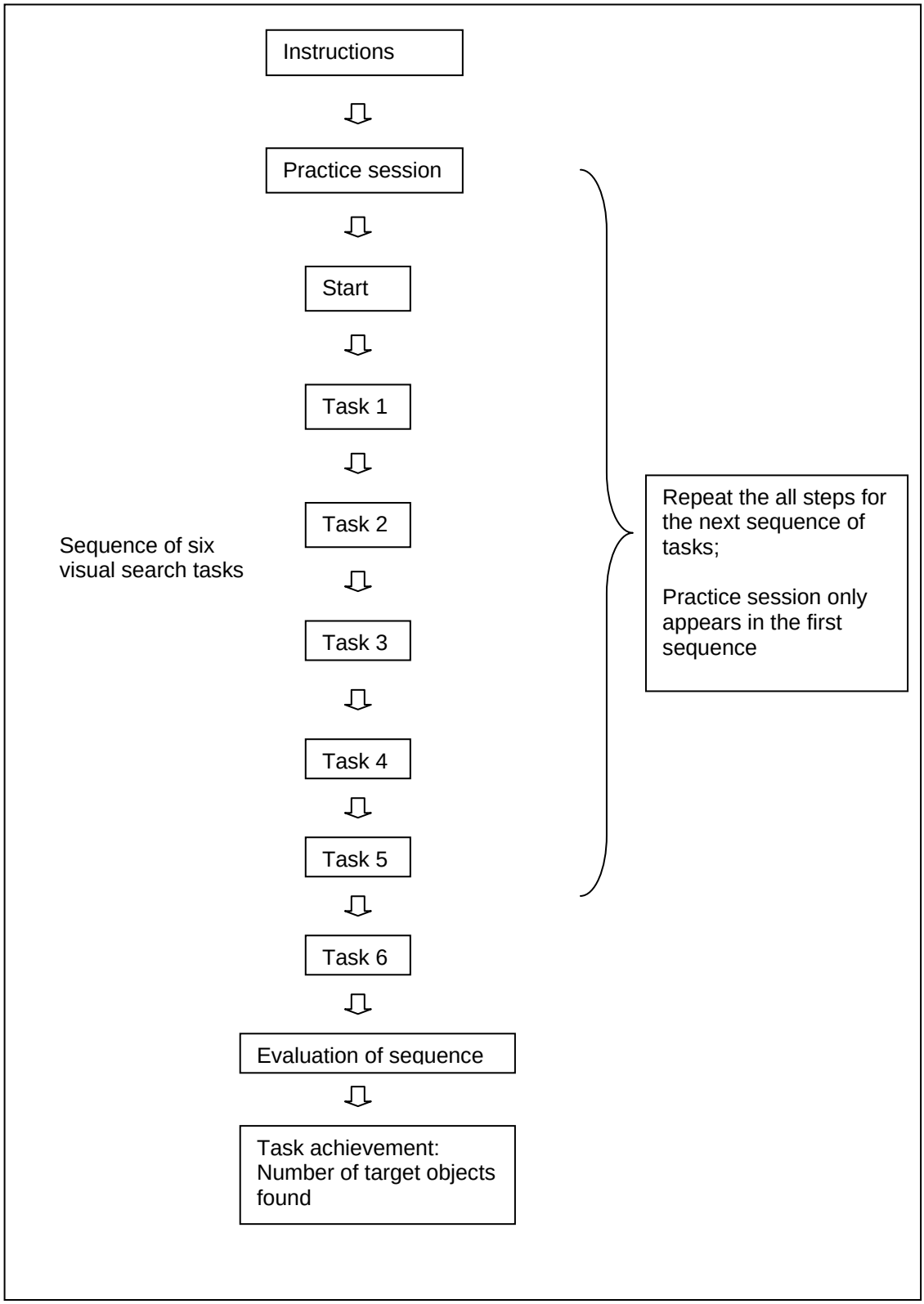
These were identical to that used in the first visual search study (Figure 5.3.2).

5.4.2.4 Experimental procedure

Figure 5.4.2 illustrated the layout and flow of the VS2 (see Appendix 5.4 for an illustration of the stimuli on the computer-based interface). The procedure is exactly the same as the one used in the previous visual search study, except that in this study (VS2), participants were required to perform the visual search tasks for four sequences.

In Figure 5.4.2, each box represented the experimental interface, which was a web page. The first page that participants encountered was the instructions page, which explained to them the procedures for the study, which was to find the target objects (i.e. red Xs). Participants were then given an opportunity to familiarise themselves with the tasks required before starting the actual visual search. Once participants had read and understood the instructions, they were then told that they could begin the study. The full details of instructions for this study were presented in Appendix 5.5. After participants had completed the sequence of visual search tasks, they were then required to provide their evaluations using the dependent variables described above, before they were allowed to move on to the next sequence of tasks. At the end of the final sequence, they were then told that they had completed the entire experiment, and were thanked for their involvement. At this point, the attendance fee was paid to them, and they were encouraged to give some informal feedback as to what they thought about the experiment.

Figure 5.4.2: The layout and flow of Visual Search (II) study (VS2)



5.4.3 Results

A preliminary 2x2x2 ANOVA (pattern x task achievability x country) was calculated for both dependent variables (annoyance ratings and number of target objects found). The factor 'country' denoted the country in which participants' universities were located in, and was a between-participants factor with two variables UK and Malaysia. This factor was included to test if there was any significant difference arising from the use of participants from two different host countries. The ANOVA calculations, however, did not show any main or interaction effect of country. The between-participants variable country was not significant ($F(1,94)=0.423$, $p>0.05$). In addition, there were also no interaction effects between country with pattern ($F(1,94)=1.08$, $p>0.05$), country with task achievability ($F(1,94)=0.029$, $p>0.05$), or interaction between country, pattern and task achievability ($F(1,94)=0.283$, $p>0.05$). The factor 'country' therefore was removed from all subsequent analyses.

Table 5.9: Descriptive statistics of the ratings of annoyance

N=96	Simple task condition		Difficult task condition	
	Decreasing time constraints	Increasing time constraints	Decreasing time constraints	Increasing time constraints
Mean	27.5	29.5	49.1	57.7
Median	27.3	26.3	49.4	63.5
Std Dev	18.9	20.9	25.0	24.9

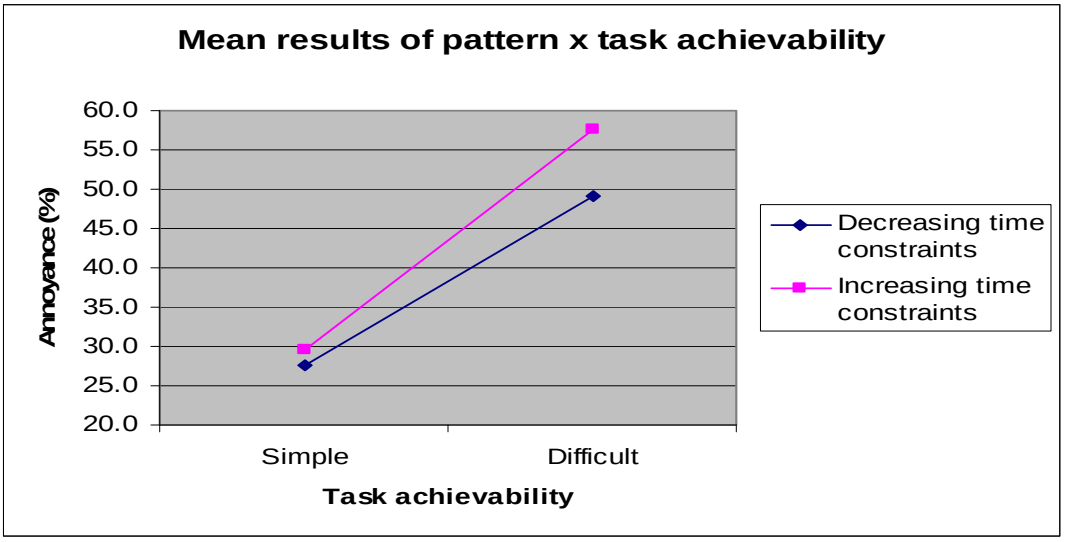
Note: The figures for the measure of frustration are taken from the 0-100 scale used, where higher numbers indicate more annoyance.

Table 5.10: Descriptive statistics on the number of target objects found

N=96	Simple task condition		Difficult task condition	
	Decreasing time constraints	Increasing time constraints	Decreasing time constraints	Decreasing time constraints
Mean	23.1	23.0	16.2	15.3
Median	24	24	15	15
Std Dev	2.57	3.13	8.00	5.30

Note: The total number of target objects embedded in each sequence is 24

Figure 5.4.3: The main and interaction effects of visual search (II)



The descriptive statistics for the dependent variables rating of annoyance was reported in Table 5.9 and illustrated in Figure 5.4.3, while task achievement was reported in Table 5.10. These results represented the combined results from studies conducted in both countries. The analyses of the results are divided as follows. Section 5.4.3.1 examines the results for possible order effects. The next two sub-sections investigate the effects of pattern and task achievability manipulations on ratings of annoyance (section 5.4.3.2) and performance in the visual search tasks (section 5.4.3.3).

5.4.3.1 Analysis of order effects

As this study was conducted using a repeated-measures design, there is a need to determine whether the order in which participants experience the four sequences (i.e. {increasing-simple}, {decreasing-simple}, {increasing-difficult}, and {decreasing-difficult}) has any significant influence over sequence evaluation and task achievement. It is expected that order effects will not be present in this study, as other studies using repeated-measures (e.g. Ariely and Loewenstein, 2000) had not reported any order effects resulting from the use of a repeated-measures design, nor have they provided a theoretical justification for order effects to be present.

The test for order effects is an ANOVA calculation on the rectangular arrays design of the study. The 24 different permutations of the rectangular arrays (which were labelled as 'order' for purposes of the ANOVA calculation) were previously presented in Table 5.8. The orders for the 24 permutations were as follows: The first permutation was {ABCD}, the second {ABDC}, third {ACBD} … and so on, until the twenty-fourth permutation {DCBA}. The tests for order effects in the rectangular arrays design was carried out using the procedures set out in Rosenthal and Rosnow (1991; pp. 405-406). A 4x24 (sequence condition x order) mixed design ANOVA was calculated for both dependent variables measured (annoyance and number of objects found). Sequence condition was a within-participants factor, while order was between-participants. The results of the calculations were presented in Table 5.11.

Table 5.11: Analysis of the rectangular arrays counter-balancing design

Model: 2x2 (pattern x task achievability) repeated-measures ANOVA	F-test	p-value
Dependent variable: Annoyance		
Sequence condition (main effect)	F(3,216)=80.5	p<0.001
Order (main effect)	F(23,72)=0.709	p>0.05
Interaction (sequence condition x order)	F(69,216)=0.959	p>0.05
Dependent variable: Number of objects found		
Sequence condition (main effect)	F(3,216)=83.7	p<0.001
Order (main effect)	F(23,72)=1.11	p>0.05
Interaction (sequence condition x order)	F(69,216)=1.72	p<0.05

Note: The factor 'sequence condition' consisted of four levels A, B, C, and D ({increasing&simple}, {decreasing&simple}, {increasing&difficult}, and {decreasing&difficult}). The between-participants factor 'order' represented the 24 permutations of sequence order listed in Table 5.8.

For the dependent variable ratings of annoyance, the calculations revealed a significant main effect of sequence condition ($p<0.001$), which demonstrated that the annoyance ratings were higher in the difficult task condition compared to the simple condition, as observed in Figure 5.4.3. This result is expected, because it implied that people were more annoyed with tasks they could not fully achieve compared to tasks they could fully achieve. Further analyses of the main effect of sequence condition are presented in the next section 5.4.3.2. From Table 5.11, the main effect of order, and the interaction between sequence and order was, however, not significant (both $p>0.05$). This meant

that there was no significant effect of order on people's ratings of annoyance, which is an important result; because it demonstrated that the adoption of a repeated-measures design in this study had not inadvertently introduced unexpected order effects.

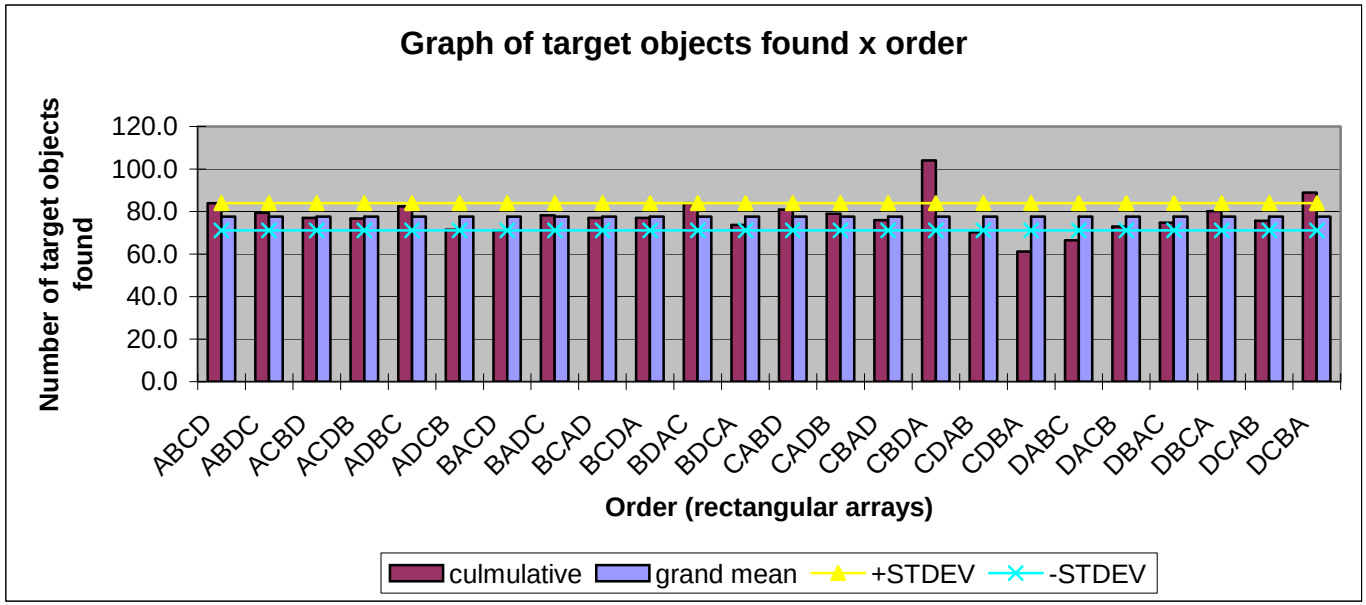
For the dependent variable number of objects found, the calculations revealed a significant main effect of sequence condition ($p<0.001$) and a small, but significant interaction between sequence and order ($p<0.05$). There was no main effect of order ($p>0.05$). The main effect of sequence and the lack of a main effect of order were expected, but the interaction between sequence and order was not. The main effect of sequence demonstrated that participants performed poorer in the difficult task condition compared to the simple task condition. The small interaction effect, however, was unexpected, because the purpose of the rectangular arrays design was to ensure that there was a full counter-balance to reduce the possibility of such order-related effects. It is proposed that this interaction may not have any theoretical significance, but merely reflected minor irregularities or outliers in the data. Two line graphs were plotted in Figure 5.4.4 to shed further light on this interaction effect.

In Figure 5.4.4, the grand mean was compared with the number of objects found per order. The number of objects found per order was calculated by adding up all of the objects that each participant had found in all four sequence conditions (i.e. increasing-simple, …, decreasing-difficult), and then averaging that total by the number of participants per order. The grand mean then averages the number of objects found per order across all 24 different orders. The relationship between the grand mean and number of objects found per order is presented in Figure 5.10. In addition to the grand mean, a one-standard-deviation boundary was also calculated, and presented in the same figure. The purpose of including this error band is to provide a visual reference for any potential outliers that may have been the source of the interaction effect calculated in the ANOVA.

From Figure 5.4.4, the biggest deviations, or outliers, were in the CBDA and CDBA positions respectively. Since there were no other outliers that

significantly deviated from the one standard deviation boundary apart from CBDA and CDBA, it is estimated that the main source of the interaction effect between sequence and order could largely be attributed to these two large outliers. On a closer re-examination of individual participants' data on the number of objects found for CBDA and CDBA (there were four participants for each order within the rectangular arrays design), the results suggested that some of the responses that participants had provided seemed large or small compared with the average results for the other 22 orders.

Figure 5.4.4: Analysis of the interaction effect of order in the number of objects found



Summary statistics:
Grand mean = 77.6 objects
Total number of objects embedded = 96
Standard deviation = 6.4 objects
N=96

Outliers:
Number of objects in CBDA: 104
Deviation from grand mean = 26.4 objects or 4.1 standard deviations

Number of objects in CDBA: 61.3
Deviation from grand mean = 16.3 objects or 2.5 standard deviations

Key to sequence order conditions:
A = decreasing time constraints & simple tasks
B = increasing & full
C = decreasing & difficult tasks
D = increasing & partial

Notes:
The purpose of the bar chart is to allow for a visual identification of the source of the interaction calculated in analysis of order effects in the rectangular arrays design.
The vertical axis of the graph represented the number of target objects found, while the horizontal axis represented order, in the form of the 24 different permutations from the rectangular arrays (see Table 5.8).
The bars represented the grand mean and cumulative mean. The horizontal lines represented one standard deviation boundaries over and below the grand mean.
The between-participants conditions with the largest deviation from the grand mean (and therefore the largest source of the interaction between sequence condition x order) were CBDA and CDBA. The rest of the deviations did not significantly deviate from the one standard deviation boundary.

For example, in CBDA, the average number of target objects found across all four sequence conditions was 104, which is larger than the total of 96 objects embedded. In addition, one participant had reported the total number of objects found as 148, which was more than 50% difference from the total of 96 objects embedded. In CDBA, the average number of target objects found was 61, which is less than two-thirds of the total number, when the grand mean is more than 80% (i.e. 77.6/96) of the total number of objects embedded. In addition, one participant had only found a total of 36 target objects, which is low. These results suggested that some of the participants may have lost count or concentration halfway through the visual search task sequence, and therefore had significantly over-estimated or under-estimated the number of objects present.

If the results of the two participants described above (i.e. the one who found a total of 148 target objects and the one who found 36) are substituted using the median results for the number of objects found (24*2 for the simple tasks + 15*2 for the difficult tasks = 78 – see Table 5.10 for medians), the interaction effect between sequence condition and order disappears immediately ($F(69,216)=1.06$, $p>0.05$). The implication of this analysis was that the interaction effect reported in the analysis of the rectangular arrays design could be attributed largely to the two outliers described. No further action was taken, as the outliers were seemingly random, not systematic. The ANOVA for the rectangular arrays design was also particularly sensitive to large outliers, because only four participants were used for each order. It is expected that sensitivity to outliers would decrease as the number of participants per order increased.

5.4.3.2 Analysis of the dependent variable ratings of annoyance

Figure 5.4.3 illustrated the mean results of pattern x task achievability for the ratings of annoyance. A 2x2 (pattern x task achievability) repeated-measures ANOVA was calculated, which revealed significant main effects of pattern ($F(1,95)=17.1$, $p<0.001$), task achievability ($F(1,95)=125$, $p<0.001$), and an interaction effect between the two ($F(1,95)=7.45$, $p<0.01$). From Figure 5.8, it

seems that the main effect of pattern was limited to the difficult task condition. To test the prediction that the source of the main effect of pattern and interaction was from the difficult task condition, the data was separated into two; the first represented the annoyance ratings for the simple task condition, while the second represented the annoyance ratings for the difficult task condition. ANOVAs were then calculated for each the two conditions, using pattern as a factor. The results of the post-hoc ANOVA tests revealed that for the full task achievability condition, the main effect of pattern was not significant ($F(1,95)=1.37$, $p>0.05$). The ANOVA for the difficult tasks, however, was significant ($F(1,95)=20.9$, $p<0.001$), which indicated that the difference between the increasing and decreasing sequences were different. These calculations supported the prediction that the source of the main effect of pattern calculated in the 2x2 (pattern x task achievability) ANOVA above was from the difficult task condition. In other words, evaluation based on patterns was only observed when people could not fully achieve the tasks. These results were consistent with hypothesis H1.

This result implied that the criteria or rules that participants used to evaluate the sequences were conditional on the achievability of the tasks. The significant main effect of pattern and the subsequent post-hoc tests suggested that the evaluation for the increasing time constraints pattern was more aversive than the decreasing time constraints only in the difficult task conditions. The significant main effect of task achievability demonstrated that participants were more annoyed with difficult tasks, compared with simple tasks. This result provides additional validity for the dependent variable measure used, as it is able to demonstrate that the task achievability manipulation does lead to different levels of annoyance being induced.

5.4.3.3 Analysis of the dependent variable number of objects found

A 2x2 (pattern x task achievability) repeated-measures ANOVA for the number of target objects found revealed a significant main effect of task achievability ($F(1,95)=165$, $p<0.001$). The main effect of pattern ($F(1,95)=1.18$, $p>0.05$) and its interaction with task achievability ($F(1,95)=1.00$, $p>0.05$), however, were not significant. This result was expected, because the main effect of task

achievability demonstrated that the experimental manipulation of task achievability into two distinct conditions (i.e. simple versus difficult) was successful. The non-significant effects of pattern (main and interaction) were also highly consistent with the results obtained from the previous memory studies and visual search study, where pattern did not materially affect participants' ability to find the target objects. A summary of the most important results was presented in Table 5.12.

Table 5.12: Summary of test results for visual search (II)

Model: 2x2 (pattern x task achievability) repeated-measures ANOVA	F-test	p-value
Dependent variable: Annoyance		
Pattern (main effect)	F(1,95)=17.1	p<0.001
Task difficulty (main effect)	F(1,95)=125	p<0.001
Interaction (pattern x task difficulty)	F(1,95)=7.45	p<0.01
Dependent variable: Number of objects found		
Pattern (main effect)	F(1,95)=1.18	p>0.05
Task difficulty (main effect)	F(1,95)=165	p<0.001
Interaction (pattern x task difficulty)	F(1,95)=1.00	p>0.05

5.4.4 Discussion

The results of this study supported the hypothesis that an evaluation based on patterns is only induced when the tasks are not simple. More generally, this result suggests that evaluations based on patterns may not be universally applicable to all sequences of tasks that are obstructed (as a result of delays or time constraints being placed on the tasks), but is, instead, conditional on whether the obstruction experienced substantially affects the achievability of the tasks. People are more likely to evaluate based on patterns if the level of obstruction materially affected the levels of task achievability in the sequence.

It was previously argued that one of the reasons why sequence pattern was relied upon for evaluation under such circumstances was because it represented a type of heuristic in which people could use when more sophisticated means of evaluating sequences is out of their reach. This view represented the 'information processing capacity' perspective on people's motives for evaluating sequences based on patterns (e.g. see Simon, 1981;

Gigerenzer and Todd, 1999; Ariely and Carmon, 2003). Under this perspective, people were assumed to be limited in their ability to process all available information required for an optimal evaluation or judgement to be made. Once their processing capacity and/or motivation were exceeded, people would then resort to using other, simpler, strategies for evaluation. Simplification strategies included shifting the focus of attention to more important information (e.g. see Wallsten and Barton, 1982; Payne, Bettman and Johnson, 1988), or changing the information processing priorities, with certain types or categories of information weighed more heavily (Wright, 1974), and/or processed to a greater extent than others (Ben-Zur and Breznitz, 1981). In using the analogies just described to explain the results of the current study, evaluation based on patterns in sequences is an important alternative strategy that people use to aid evaluation, because patterns are easier characteristics to retrieve from an experience compared to other factors (such as the total utility of the sequence) when making retrospective evaluations (e.g. see Ariely, 2001). In this study, it was argued that pattern might be easier to retrieve because people have to be attentive to the time parameters (and by extension, its sequence pattern), as time is a crucial factor that would have a direct impact on their ability to find the embedded target objects.

A second, alternative but related, explanation to the heuristic perspective explored above, it is proposed that emotional evaluations is a function not only of the pattern of annoyance caused by the time constraints, but also of the relative changes in task achievability. Under this perspective, it is suggested that in a sequence where the tasks are simple (i.e., always achievable regardless of time constraints), people's moment-by-moment evaluation is relatively constant or flat over time. This is because the emotional evaluation of a sequence of tasks is largely a function of task achievement (i.e. their inability to achieve the tasks and their annoyance that results from it), rather than of the changing patterns of obstruction. This means that for tasks at each stage of a sequence that is always achievable, the expected level of annoyance for both increasing and decreasing sequence patterns is constant throughout the sequence and similar between the two patterns. However, if the tasks were time-constrained to the point that people could not complete the

some/all of the tasks within the sequence, the emotional evaluation of the sequence over time would then track the variations in achievability, where decreasing task achievability induces an emotional pattern of increasing annoyance, and vice versa for increasing task achievability. It is predicted that the increasing annoyance sequence would be evaluated more negatively compared to the decreasing, because the weights in the sequence evaluations are end-loaded. To summarise this interpretation of the result, the pattern of emotions is dependent on the severity of the time constraints. Time constraints that obstruct but does not prevent full task achievement may not be sufficient to induce an evaluation based on patterns, because the pattern of emotions over time does not vary with the changing severity of the time constraints.

If these explanations are accepted, then further restrictions have to be placed on the proposed Model 2 and Model 3 to account for the results of the present study. From Model 3, where TC=time constraints and V(.) is the negative utility function for a sequence of time-constrained tasks, it does not always follow that if a sequence of time constrained tasks, where TC1>TC2>…TCn and if the weights end-loaded, that the inequality V(TC1, TC2, …, TCn)>V(TCn, …, TC2, TC1) would hold. The results of the visual search study (II) suggested that the inequality would hold only if the obstruction, On, as defined in Model 2 (and of which Model 3 was a derivative of), was redefined such that the obstruction materially affected task achievement. In notational form:

Let O=obstruction and V(.) the negative utility function for a sequence of obstructed tasks. For O2>O1, V(O2)>V(O1) is true only if On reduces people's ability to achieve a task, otherwise V(O1)≈V(O2), and V(O1, O2, …, On)≈V(On, …, O2, O1) … [Corollary 1]

V(TC1, TC2, …, TCn)≈V(TCn, …, TC2, TC1) … [Corollary 2]

Corollary 1 represents the constant or flat emotional profile of the sequence when task obstruction does not materially affect achievement (Model 2 applies

when the degree of obstruction is material). Note that when V(O1)≈V(O2) ≈…V(On), Corollary 1 would always hold, even when people recognise the differences in sequence patterns. Corollary 2 is an extrapolation from Corollary 1, implying that the manipulation of time constraints to generate obstruction would not induce evaluations based on patterns unless the obstruction has a material impact on perceived task achievement.

A third and final interpretation of the results of the present study posits that the sequence evaluation comprises of two stages. In the first stage, people evaluate the variations in the pattern of emotions. Under this view, the emotional profile does vary with changes in the parameters of the time-constrained tasks, unlike Corollary 1 above. In other words, emotions increase when time constraints become more severe, and decrease when time constraints become less severe. In the second stage, however, people consider their perceived performance in the tasks. If people perceive that the tasks are simple enough to be always achievable, this may lead them to conclude that the variations in their emotions are negligible and not worth differentiating, because the task has already been achieved. However, if people perceive that they were unable to complete the tasks within the given time constraints set, they may prefer a sequence pattern that is seemingly less emotionally unattractive (i.e. the decreasing time constraints sequence) to compensate for the increase negative emotions caused by their inability to complete the tasks.

The difference between this third interpretation and the first (i.e. heuristics) interpretation is that under the heuristics interpretation, it was assumed that people choose to follow a rule such as evaluation based on patterns because they were unable or unwilling to work out the total negative utility, and therefore needed a tool that helps them to quickly summarise their experience. Under the third interpretation, however, it was assumed that people choose to use such a rule for evaluation because they are trying to manage their emotional experiences by providing evaluations that indicate their preference for what they perceive as the more attractive sequence, which can help them to minimise the negative impact of the unachievable tasks. This is also the

case for the second interpretation, except that the pattern of emotions in the second interpretation does not vary with the profile of time constraints, unlike the third interpretation.

A general point to note from the results of the studies reported in this chapter is that task achievability is independent of the emotional evaluation of a sequence. It is important that the only factor that would influence task achievement is the manipulation of different task levels (i.e. full versus partial achievability). If pattern were also shown to have affected task achievement, it would then become impossible to separate and determine the factors that influenced the emotional evaluation of a sequence of tasks. Consider a situation in the VS2, when people evaluated the decreasing time constraints sequence as being less annoying than the increasing time constraints sequence, and had also performed better in the decreasing (instead of equal achievement across both sequence patterns found in the four studies of this chapter, MS1 and MS2, and VS1 and VS2). There are three possible explanations for this hypothetical result.

The first is that people could have evaluated the decreasing time constraint sequence less aversively because they interpreted the sequence pattern to be improving and therefore preferred improvement. The second is because they had performed better in the decreasing time constraint compared to the increasing sequence, and therefore evaluated the sequence based on task achievement, which meant that the decreasing would be less aversive because of the better task achievement. The third is that they interpreted the decreasing as an improvement, and had also performed better in the decreasing; therefore it would only seem logical to evaluate the decreasing as being less aversive than the increasing. In the first explanation, evaluation is based strictly on pattern, rather than task achievement. In the next two explanations, however, evaluation is based on task achievement, or a combination of pattern and achievement. These explanations compete against each other, because it could not be known whether pattern, task achievement or its combination induce people to evaluate based on patterns. Therefore, the independence of sequence evaluation from task achievement is a necessary

condition to ensure that differences in the emotional evaluation of annoyance measured were actually induced by sequence patterns, rather than by task achievement.

5.5 Summary and conclusions

The broad objectives of the studies in this chapter were to further refine the methods used to test Model 1, which predicted how people would evaluate sequences of delays. The results of the studies in this chapter showed that the original specifications of Model 1 had to be modified to incorporate an additional condition (task achievability) that related to the degree of obstruction caused by the delays. The contribution made by the studies in this chapter to the sequence literature is to demonstrate that gestalt theories may be applied to model behaviour associated with sequences of tasks that are affected by delays, albeit only if certain conditions are met. Previous research on how gestalt theories could be applied to model the evaluation of delay sequences was limited to the documentation of end effects in a sequence of queues (i.e. Carmon and Kahneman, 1993). The literature, has not, however, developed models that links delays with obstruction and negative emotions, or acknowledged and addressed the boundary issues of the model. The primary contribution of the research reported in this chapter is to present data to support in the development of such a model. The remainder of this section summarises the rationale for and contributions of the specific studies in this chapter, and suggests possible future avenues of research.

The original purpose of MS1 was to test Model 1 using an alternative task manipulation, given that the results of the previous study, IDS, were inconclusive. The participant-task interaction in MS1 was redesigned such that the delays materially affected task achievement (i.e. participants could not fully achieve the tasks because of the delays). The results of the study supported the hypothesis that the decreasing delay pattern was less aversive than the increasing pattern. However, the results of MS1 were confounded. MS2 was introduced to rectify the confounded aspect of the design, by using a different participant-task manipulation from the one used in MS1. The results of MS2

supported the hypothesis in MS1, which had predicted that the decreasing pattern was less aversive than the increasing one. The objective of VS1 was to introduce an alternative form of participant-task manipulation to test the proposition that obstruction in a sequence of tasks could also be generated using alternative methods, such as time constraints. The final study, VS2, was conducted to test the implicit assumption made in the previous three studies that evaluation based on patterns would only be induced if the tasks had become more difficult due to obstruction caused by the delays or time constraints. The results of both VS1 and VS2 supported the hypotheses made for it.

The design of the memory and visual search studies required that Model 1 be modified to take into account the different participant-task interaction used across all studies, including the IDS. Model 2 proposed that the evaluation of delay sequences developed in Model 1 could be extrapolated to incorporate any manipulations that would cause obstruction in a sequence of tasks. This was because VS1 had demonstrated that delay was only one of a number of ways in which people's achievement of tasks could be obstructed. Model 3 represented a more specific model of obstruction that was derived from Model 2; by proposing that time constraints could be used as a manipulation of obstruction. The most important finding of the studies in the present chapter was the demonstration that an evaluation based on patterns was only induced when tasks were difficult, which suggested that evaluation based on patterns represented a heuristic strategy that people use when the evaluation of the utility integral is difficult or unachievable due to an overload in information processing demands. There were two alternative explanations, both of which suggested that the emotional evaluation of a sequence was itself driven by task achievement, where if full achievement is possible, this meant that the negative emotions generated would be insufficient to induce an evaluation based on patterns. If, however, the tasks were made difficult, this would then trigger an evaluation that traces the pattern of task achievement, which means that a sequence of increasing task achievability would have a less aversive evaluation compared to one with decreasing achievability. The difference between the two explanations was in the prediction of the profile of emotions

over the sequence. The first suggested that the profile was flat for the sequence of tasks that were simple (i.e., always achievable regardless of the level of obstruction). The second suggested that the profile of emotions did vary with changes in the intensity of the aversive stimuli, but because the overall level of aversion was so small when the tasks were always achievable, people did not feel that they needed to differentiate between the sequence patterns.

To conclude, the studies in this chapter have refined Model 1, which predicted people's evaluation of a sequence of delayed tasks. It was shown that the predictions of Model 1 were only applicable when the delays were significant enough to make the tasks difficult (i.e., the point where the delays prevent full task achievement). The studies in the next chapter investigate the applicability of Model 1 for sequences with longer delays and influences from other factors such as expectations.

Chapter 6

Extending the model of delays to longer sequences

6.0 Introduction

The studies in the previous chapter were motivated by the idea that gestalt theories could be applied to model behaviour in delay sequences. It was demonstrated that delays could induce the evaluation of a sequence based on its patterns if the delays materially affected the sequential experience, such as by preventing full task achievement. There is, however, a need to know whether the models proposed in chapter 5 can also be applied to sequences with longer durations of delays (where duration is defined as the time interval from start to the end of the sequence), given that previous studies had only investigated sequences with short durations. For example, the durations used in all four studies of chapter 5 did not exceed 180 seconds (3 minutes), yet one can think of many instances in real life where the duration of delays would last for a considerably longer period. For example, it is not unusual to expect to have to wait for more than 60 minutes when flights are delayed at an airport, or for more than a few hours when queuing to buy tickets for tennis matches at Wimbledon during the tournament.

The objective of the studies reported in this chapter is to broaden the research topic, by exploring other reasons that can affect and/or influence people's evaluation of delay sequences, and sequences more generally. For example, in the discussion of the Internet Delay Study (IDS) in chapter 3, it was proposed that domain-specific factors, in addition to delay, might also influence sequence preference. Domain is defined as the environment or context in which a delay is experienced. For example, the delay could be experienced when checking out of a hotel, or when queuing for train tickets. While the delay duration for both events might be identical (e.g. 60 minutes each), the experience of the event, however, may be different because people's expectations may vary for different events. For example, 60 minutes queuing

to buy a train ticket at a station is more realistic compared to a 60 minute delay when checking out of a deluxe hotel. Given that expectations were shown to be a significant driver of preferences in sequences of health and money (Chapman, 1996a and 2000; Read and Powell, 2002), it is important that further research is carried out to determine whether such findings also extend to sequences of delays.

The outline of the remainder of this chapter is as follows. There are three studies in this chapter, the Transport Delay Study (TDS) in section 6.1, the Domain Study (DS) in section 6.2, and the Expectations Study (ES) in section 6.3. Briefly, the objective of the Transport Delay Study is to test the predictions of Model 1 developed in chapter 5 for delay sequences of longer durations. The objective of the Domain Study is to explore the effects of domain on preferences for delay sequences. The Expectations Study presented a formal test of the relationship between preferences and expectations for delay sequences. Finally, section 6.4 summarises and concludes the chapter.

6.1 The Transport Delay Study (TDS)

6.1.1 Introduction

The studies in chapter 5 have indicated that Model 1 could only be used to predict sequence evaluation when the delays prevented full task achievement. One reason given was that delays on their own were unable to generate a sufficiently large negative emotion to induce an evaluation based on patterns. It is suggested that this could be due to the length of sequence durations used (e.g. less than 3 minutes in total), which means that such short delays will only induce sufficient negative emotions when the delays have a material affect on the sequence outcome, such as by preventing full task achievement. In other words, it is the inability to achieve the task(s) that generates negative emotions which is sufficient to induce evaluation based on patterns, rather than from the experience of the delays. However, the multitude of studies reported in the (non-sequence) delay literature (e.g. Lawson, 1965; Sawrey and Telford, 1971; Larson, 1987, Taylor, 1994; and Dube, Leclerc and Schmitt, 1991; see section 2.1 of chapter 2) has shown that the impact of longer delays (for the purposes of this study, longer delays are defined as more than 3 minutes) does have a material affect on consumers' experience with, and the evaluation of, a service. It is suggested that the negative emotions evoked from delays in the studies reported in the delay literature is high, which is why they have a significant affect on preferences. It is therefore proposed that, in the context of delay sequences, longer delays can evoke sufficient negative emotions in people to induce evaluations based on patterns. This prediction is based on Model 1 developed in the previous chapters, which is now briefly revisited.

Let V(d1, d2, …, dn) be the evaluation of a sequence of delays, which is equal to a weighted average of the utility from individual delays.

V(d1, d2, …, dn) = w1.V(d1) + w2.V(d2) + … + wn.V(dn), with (w1, w2, …, wn) being the weights.

Let the magnitude of delays be such that (d1<d2<…<dn), and that the weights are end-loaded (e.g. w1<w2<…<wn). Therefore, Model 1 predicts that an increasing delay sequence, V(d1, d2, …, dn), would be evaluated more aversively than a decreasing sequence, V(dn, …, d2, d1). Notationally, this is written as:

V(d1, d2, …, dn)>V(dn, …, d2, d1) … [Model 1]

The present study adopted a scenario where people experience a sequence of delays while waiting for public transport. This scenario was used so that longer delay durations can be introduced to test Model 1, given that the delay duration (d1+d2+…+dn) did not exceed 180 seconds for all four of the previous studies reported in chapter 5. The delay duration used in the present study was increased to 2 hours. This means that a questionnaire rather than a laboratory-based method of experimentation would be a more appropriate choice of experimental method, because of the longer duration involved (see discussion in section 4.2.2.3 of chapter 4 on the relationship between sequence duration and experimental method chosen).

The dependent variable used required participants to indicate which of two possible scenarios, increasing or decreasing sequence of delays (the increasing delay sequence was {5,15,40,60} minutes, and the decreasing delay sequence was {60,40,15,5} minutes) would make them feel more frustrated. They were also given the option of indicating that both sequences were equally frustrating. The design of the dependent variable adopted in this study (i.e. choice of increasing, decreasing or indifferent) was adapted from one used in Ross and Simonson's (1991) studies, because the experimental design of this study (i.e. comparison between two sequences) is similar to the one used in Ross and Simonson. One advantage of including an indifference option is that it is able to capture preferences where the majority of participants are indifferent, which is not possible with a forced choice option (i.e., choose either increasing or decreasing delays).

In addition to the original Transport Delay Study, a replication study that used an alternate dependent variable was also tested. The replication study was a precautionary measure used to test for the potential effects that negatively phrased questions might have on preferences, as previous studies on choices had demonstrated the sensitivity of decision-makers to such framing effects. For example, Tversky and Kahneman (1986) showed that people had different preferences depending on whether the question was positively or negatively phrased.

This replication study, however, was only carried out once the main study was completed and the results known. The main purpose of the replication study was to provide further confirmatory evidence of the results of the main study, but using a different method to capture preferences. In the main study, the question was negatively framed, where participants were asked to indicate their frustration. This meant that the dependent variable captured an emotive response from participants. In the replication study, however, participants were asked to indicate their preference for one of the two patterns, with no option of indifference. Given that participants were now asked to choose between the two sequence patterns, the reframing of the question gave participants the possibility of incorporating other non-emotive factors into their decision if they so wished to. The reason why the indifference option was removed in the replication study was because it became clear from the results of the main study that a significant majority of participants were not indifferent between the patterns. This meant that it was not strictly necessary to offer the indifference option anymore. All other stimuli and procedures used in the replication study were exactly identical to the main study. It is expected, however, that the responses to the replication study is similar to the original study, as there is no theoretical reason to expect differences given that the stimuli is identical, and that the purpose of the replication study is primarily a method of triangulating the results obtained. The hypotheses of the studies are therefore:

H1(original study): The increasing delay sequence is more frustrating than the decreasing delay sequence.

H1(replication study): The decreasing delay sequence is preferred to the increasing delay sequence.

6.1.2 Participants and method

One hundred and ten undergraduate students from several large British universities volunteered to participate in the study. Their ages ranged from 18 to 25 years old. No payment was made. Fifty undergraduate students participated in the study with the original dependent variable and the remainder in the replication study. Both studies, which were in a questionnaire format, were administered independently.

Participants had to imagine that they have to get from one place to another using public transport, while experiencing a sequence of four delays when they have to change transport. They were shown two sequences of delay parameters, where the increasing pattern of delays was {5,15,40,60} minutes, and the decreasing pattern was {60,40,15,5} minutes. They were then asked which, if any, of the pair of sequences were more frustrating (original study), or which of the two sequences they would choose (replication study). The actual questionnaires used were reproduced in Appendix 6.1. As a final manipulation, the order in which the sequence pattern appeared (increasing before decreasing, or decreasing before increasing) was also counter-balanced. For both studies, half the number of participants was presented with the scenario where the increasing delay sequence was presented before the decreasing, and the other half, the decreasing before increasing.

6.1.3 Results and discussion

In the original study, 7 participants (14.0%) indicated that the increasing delay sequence was less frustrating, 39 participants (78.0%) indicated that the decreasing sequence was less frustrating, while 4 participants (8.0%) indicated that they were equally frustrated with both sequence patterns. A chi-squared test that grouped the evaluations into two categories (category 1 is where participants had evaluated one sequence less frustrating than the other versus category 2 of indifference) found that the indifference option was not significant

(chi-squared(df=1)=35.3, p<0.001). A subsequent chi-squared test was then calculated with the indifference option removed to determine the significance of the observed differences between the increasing versus decreasing delay sequence. The results of this test supported the hypothesis that the increasing delay sequence was more frustrating than the decreasing delay sequence (chi-squared(df=1)=22.3, p<0.001). For the replication study, 13 participants (21.7%) chose the increasing delay sequence, while 47 participants (78.3%) chose the decreasing delay sequence. A chi-squared test on the responses of the replication study revealed similar results to the original study (chi-squared(df=1)=19.3, p<0.001), which showed that people preferred the decreasing to the increasing delay sequence. In addition, the results of both studies also implied that the type of dependent variable used (original versus replication) did not seem to have affected preferences, which were both significant in the hypothesised direction.

It is implied from the results of this study that Model 1 can also be used to predict preference for sequences of delays with longer durations. People prefer a decreasing delay sequence over an increasing one because the distribution of weights in their evaluation is end-loaded. This means that sequences with patterns that have more favourable end-characteristics, such as a decreasing delay trend, will have a more favourable evaluation compared to other patterns with less favourable end-characteristics (such as an increasing delay trend).

However, it is argued that a distinction has to be made between the sources of the end-loaded weights for retrospective versus prospective studies. In retrospective studies, an end-loaded or monotonically increasing set of weights represents a heuristic that a decision-maker can use to quickly summarise an experience. The literature has shown that the use of such heuristics (i.e. heavily weighing the final trend and the end of a sequence) is highly prevalent in sequences with very intense and aversive stimuli, such as exposure to cold water, loud sounds, or painful medical procedures (e.g. colonoscopy). In contrast, however, it is argued that there are two reasons why people have end-loaded weights in prospective studies. The first is because of its function

as a heuristic. The second, however, is that such preferences are derived from the context or domain in which people experience the sequence, which may involve deliberate and careful reflection. For example, it had been demonstrated in the literature review in chapter 4 that in a non-delay domain, people often prefer an increasing or improving sequence of monetary payments, because this pattern provided a signal of the future career prospects of the decision-maker (Loewenstein and Sicherman, 1991). Matsumoto et al. (2000) and Read and Powell (2002) had also provided numerous examples where people were clearly able to articulate good reasons for preferring a sequence pattern, which does not suggest that such preferences are reflective of its function as a heuristic for decision-making.

Therefore, the question that arises is to what extent does research that used a prospective paradigm translate to a retrospective paradigm and vice versa? The analysis of methods used in the literature, which was previously reported in section 4.2.2.2 of chapter 4, and more generally the review of the literature in chapter 2 and section 4.1, did not show any discernable bias in the theories developed using a prospective compared to a retrospective study. For example, Varey and Kahneman (1992) developed their peak-end theory in a study that used a prospective design, where they asked participants to provide evaluations of imagined aversive and painful stimuli. These results were replicated in subsequent studies that used retrospective designs (Kahneman et al., 1993; Fredrickson and Kahneman, 1993; Redelmeier and Kahneman, 1996 etc.; Kahneman et al., 1997; Schreiber and Kahneman, 2000). On this basis, therefore, it is argued that the extension of the theory developed in Model 1 to explain the results in the present study is justified.

In addition to providing additional support for Model 1, the results of this study also provides implicit support for the idea proposed in the discussion of the second Visual Search study in chapter 5 (VS2) that there may be different sources of negative emotions that are generated from sequences of delays. In VS2, it was suggested that the first source was from the delays itself, where people were frustrated because the delays took up a significant proportion of their time and obstructed their progression. The second source is from the

prevention of task achievement. People were frustrated here because they could not achieve the task that was set for them, rather than because the delays obstructed their progression. The results of VS2 had shown that negative emotions were only large enough to induce an evaluation based on patterns when the delays reduced people's ability to fully achieve the tasks. On the basis of this argument, it was concluded that, on its own, very short delays (in the order of seconds) were unable to generate sufficient negative emotions to induce an evaluation based on patterns unless it prevented full task achievement. However, for sequences of longer delays, such as those used in the present study, it is argued that people felt that the delays were long enough to lead them to evaluate a sequence based on patterns.

There is, however, a need to conduct further tests on the general applicability of the gestalt concepts embodied in the extension of Model 1 to longer delay sequences. This is because it was suggested that the source of the end-loaded weights observed in the present study might be different from those of shorter studies investigated in chapter 5. The weights could be reflective of the domain in which the sequence is experienced rather than from its function as a heuristic. For example, participants in the present study may have felt that, in the current domain of the delays, it was appropriate to use end-loaded weights as a basis on which to evaluate the sequence patterns. The next study in this chapter, however, revisited some of the earlier literature on non-sequential delays that was discussed in chapter 2 to explore the effect of domain on preferences for delay sequences.

6.2 The Domain Study (DS)

6.2.1 Introduction

The objective of this study is to further explore the effects of domain on preferences for delay sequences. In the introduction to this chapter, domain was defined as the environment or context in which a delay is experienced. It is conjectured that preferences for delay sequences may be conditional on domain, based on evidence from the delay literature reported in section 2.1 of chapter 2. Briefly, this literature had shown that in addition to the duration of delay, domain might also play an influential role in evaluation. For example, Larson (1987) demonstrated that American consumers get upset if people were seen to be cutting the queue, and therefore was more willing to tolerate a longer queue without queue-jumpers, to a shorter queue where queue-jumping is rife. In another study, this time on airport queues, Taylor (1994) found that consumers were less frustrated with delays caused by events outside of the service providers' control, such as force majeure, compared with events that were perceived to be within the service providers' control, such as poor management and organisation.

These studies showed that the delay duration was not the only consideration in the evaluation of a delayed service, which suggested that the domain, or event in which the delay is experienced, is also an important consideration for research on sequences of delays. The potential implications for sequences of delays and the predictions of Model 1 are that the presence of multiple domains can affect the distribution of weights in sequence evaluation. This means that a sequence pattern with attractive characteristics from a gestalt viewpoint (e.g., improving trend, better end) may not always receive a more favourable evaluation compared to other sequence patterns, because the weights are not end-loaded in evaluation, as postulated in Model 1.

The manipulation of domain in the present study was operationalised as follows. In the previous Transport Delay Study, the events that caused the delays were all identical, that is, people were waiting for public transport to

arrive to take them to their next destination. The present study, however, utilised a scenario where people experience a sequence of delays with identical parameters to the previous study, but where the events that caused the delays were varied. The dependent variable here was also identical to that used in the Transport Delay Study. Participants were asked which of two possible scenarios, increasing or decreasing sequence of delays, would make them feel more frustrated. They were also given the option of indicating that both are equally frustrating.

Given that the objective of the present study is to explore people's preferences for delay sequences with different domains, however, no specific hypothesis is made. Instead, the study aims to determine whether people's preferences for delayed sequences are significantly affected by domain, and to provide a basis on which to develop a more formal model to explain the effects of domain (if any) on preferences for delay sequences. Such a model is presented in the next study, which is the Expectations Study.

6.2.2 Participants and method

Sixty-nine undergraduates from one large British university volunteered to participate in the study. Their ages ranged from 18 to 25 years old. No payment was made. Participants had to imagine that they are leaving a hotel to catch a train. In the process, they experienced a sequence of four events: checking out of a hotel, waiting for a taxi to arrive, travelling in the taxi, and then queuing at the station to buy a train ticket for their onward journey. The parameters for the delays caused by these events were {5,15,40,60} minutes respectively for the increasing delay sequence, and {60,40,15,5} minutes for the decreasing delay sequence. The dependent variable required participants to indicate which, if any, of the pair of sequences was more frustrating. The format and presentation of the dependent variable was identical to the one used in the original condition of the previous Transport Delay Study. The full questionnaire used in this study is reproduced in Appendix 6.2. As a final manipulation, the order in which the sequence pattern appeared (increasing before decreasing, or decreasing before increasing) was also counter-balanced. Thirty-five participants were presented with the scenario where the

increasing delay sequence was presented before the decreasing, while the remainder were presented with the decreasing before increasing.

6.2.3 Results and discussion

The results showed that 39 participants (56.5%) evaluated the increasing delay sequence less frustrating compared to the decreasing sequence, 20 participants (29.0%) evaluated the opposite, while 10 participants (14.5%) indicated that they were equally frustrated with both sequence patterns. Since the purpose of the study was to explore the effects (if any) of domain, no specific hypothesis was tested. Instead, tests were performed to determine the significance of the differences observed. A chi-squared test that grouped the evaluations into two categories (category 1 is where participants had evaluated one sequence less frustrating than the other versus category 2 of indifference) found that the indifference option was not significant (chi-squared(df=1)=34.8, $p<0.001$). A subsequent chi-squared test was then calculated with the indifference option removed to determine the significance of the observed differences between the increasing versus decreasing delay sequence. The results of this test showed that the increasing delay sequence was less frustrating than the decreasing delay sequence (chi-squared(df=1)=6.12, $p<0.05$), which indicated that preferences were in the opposite direction of those from the previous Transport Delay Study.

This result suggests that domain is a significant consideration in sequences of delays. To determine the significance of the preference reversal observed between the results of the present study compared to the previous study, a Mann-Whitney test was calculated with preferences as the dependent variable and the two studies as the grouping (independent) variable. It is possible to perform a Mann-Whitney test on the responses from both the present and the previous study because the dependent variable and delay parameters for both studies are identical. The test revealed that preferences for the present study were significantly different from those of the Transport Delay Study ($Z=3.44$, $p<0.001$). This result provided support for the assertion that domain has a significant effect on preferences for delay sequences, because variations in the domains of the events was shown to have reversed the preference for the

delay sequences. Figure 6.1 provided a visual illustration of the sequence preference from both the Transport Delay and the present study, while Table 6.1 presented the results.

Figure 6.1: Preferences in the Transport Delay and the Domain Study

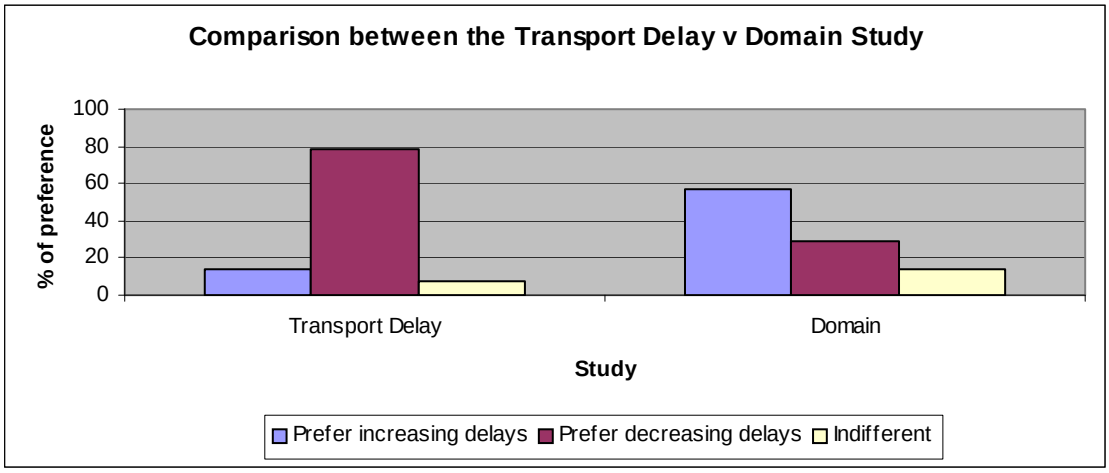


Table 6.1: Summary of results from the Transport Delay and Domain Study

Study	Descriptive statistics							Analysis
	Increasing delay sequence preferred		Decreasing delay sequence preferred		Indifference		N	Chi-squared test of significant differences
Transport Delay Study (TDS)	7	(14.0%)	39	(78.0%)	4	(8.0%)	50	$\chi^2(1)=34.8$, $p<0.001$
Domain Study (DS)	39	(56.5%)	20	(29.0%)	10	(14.5%)	69	$\chi^2(1)=6.12$, $p<0.05$
Mann-Whitney comparing TDS with DS: $Z=3.44$, $p<0.001$								

It is proposed that the observed preference reversal (where the increasing delay sequence was preferred to the decreasing delay) and the effects of domain on preferences for delay sequences could be explained using the expectations theory developed by Chapman (1996a; 2000). In applying Chapman's ideas to interpret the results of this study, it is conjectured that the preference reversal was observed because the increasing delay sequence was more expected compared to the decreasing delay sequence. Expectations here were defined as the degree of realism between the delay parameters with

the domain of events. In other words, is the duration of delay realistic given the event that causes it?

For example, the expected time taken to check out of a hotel is usually prompt, because hotels have to earn their keep based largely on the quality of their service, of which speed is one important factor. Therefore, a 5 minute checkout (i.e. from the increasing delay sequence) would be a more expected or realistic parameter, compared with a 60 minute wait (i.e. from the decreasing delay sequence), which would seemed somewhat unusual and perhaps rather tardy, especially if one was paying for, and therefore expecting, a quality service. Similarly, assuming that one could compare individual delay events between the increasing and decreasing sequences, a 15 minute (from the increasing delay sequence) wait for a taxi may be perceived as a more reasonable and expected duration to wait compared to 40 minutes (from the decreasing delay sequence). As for lengthy queues at the train station, while 60 minutes (from the increasing delay sequence) might seem a long duration to wait, it is, however, not an entirely unexpected proposition, especially at peak travel seasons and times.

Based on the comparisons above, it is argued that the sequence of increasing delays was evaluated less frustrating because each of the individual delays and events that causes it was more expected compared to the decreasing delays sequence. Therefore, it is proposed that the different domains for the delays has lead people to evaluate delay sequences based on a combination of two factors, the realism of the individual events that causes the delay when compared with its delay parameters and the pattern of delay. People prefer a pattern of decreasing to increasing delays, but only if the parameters is realistic for the events. In this exploratory study, domain was a stronger influence than pattern, because the preference for the decreasing delay sequence was reversed.

To summarise, the present study has found that the predictions of Model 1 do not always hold (i.e. people do not always evaluate the decreasing delay sequence less negatively) when the domain of delays within a sequence

varies. It is suggested that the effect of varying domains on preferences was to alter the distribution of weights to reflect the different expectations arising from different combinations of event and its duration of delay. A limitation of the present study, however, is that it has only used one form of manipulation of domain. This implies that the current results may not generalise to other delay domains. The next study in this chapter presents a more formal model to expand the test of the conjecture proposed in this study.

6.3 The Expectations Study

6.3.1 Introduction

In the discussion of the previous study, it was suggested that the expectations theory developed by Chapman (2000) could be used as a framework to explain the observed domain effects on preferences for delay sequences. It was argued that in addition to patterns, people's preference for delay sequences was also influenced by their expectations of whether the duration of delay matches the event that caused the delay. The primary objective of this study therefore is to test the conjecture that the mechanism in which domain influences sequence preference for delay sequences is through expectations.

In the literature on expectations previously reviewed in section 4.1.2.1 of chapter 4, Chapman (2000) explained that expectations influences preference through the reference point effect, whereby decision makers compare sequence outcomes to a reference sequence that is based on expectations of how such outcomes usually unfold over time. An earlier paper by Chapman (1996a) showed that people prefer sequences with patterns that are expected over ones that have unexpected patterns (assuming the integrated utilities of both sequences are equal). This is because unexpected sequences generate net negative utilities (i.e. the difference between the sum of all individual positive and negative utilities), which make such sequences less favourable when compared with expected sequences. The net negative utility arises because the sum of individual negative utilities in a sequence with an unexpected pattern outweighs the positive ones, due to negative outcomes weighing more in evaluation compared to positive outcomes (as demonstrated by the asymmetric value function theorised in Kahneman and Tversky, 1979). The reference point itself, which represents people's expectations, however, is derived from people's observations of, and experiences with, their environment (Chapman, ibid.).

In an application of Chapman's (2000) theory to sequences of delay, it is assumed that people's attitudes towards delays vary across different events

that cause the delay, because of expectations developed from personal experience. For instance, as previously argued in the Transport Delay Study, people may expect delays from hotel check outs to be short, but expect that a delay from queuing for a train ticket to be more unpredictable, and possibly longer in duration. For example, people may prefer to wait 5 minutes for a hotel check out and 60 minutes for a train ticket rather than the other way round. The previous study has shown that people do not always prefer a sequence of decreasing delays, because changes in expectations as a result of changes in the event that caused the delay may also affect the degree to which sequence evaluation is influenced by pattern alone. Two related predictions could be made based on this observation. The first is that people use their own experiences as a reference point to determine what is expected and what is unexpected. People will then prefer the expected sequence to the unexpected. The second is that the influence of pattern vis-à-vis other factors on sequence evaluation varies according domains – for some domains, pattern dominates expectations, but in others, it is the other way around. The first prediction (where pattern dominates evaluation) had been investigated in the four studies reported in chapter 5, and the first two studies of the present chapter. The model that is being developed in this study considers the second prediction, where expectations dominate evaluation. The development of the model is now presented in mathematical form.

In Model 1, V(d1, d2, …, dn) was defined as the evaluation of a sequence of delays, which is equal to a weighted average of the utility from individual delays, or:

V(d1, d2, …, dn) = w1.V(d1) + w2.V(d2) + … + wn.V(dn), with (w1, w2, …, wn) being the weights.

Model 1 also assumes that the weights are end-loaded (e.g. w1<w2<…<wn), which leads to the prediction that the increasing delay is more aversive than the decreasing delay sequence [i.e. V(d1, d2, …, dn)>V(dn, …, d2, d1)] when the magnitude of delays are such that (d1<d2<…<dn).

In this study, however, it is argued that expectations determine the distribution of weights. The magnitude of the individual weight (||wn||) is a function of the expectations, f(expect), of the duration for the event. Notationally:

||wn|| = f(expect) … proposition 1

This means that the distribution of weights (w1, w2, …, wn) may take different forms, depending on the expectations for the individual events within the sequence. The evaluation of a sequence of delays, V(d1, d2, …, dn), therefore, will also be conditional on expectations.

V(d1, d2, …, dn) = w1.V(d1) + w2.V(d2) + … + wn.V(dn)

Combining proposition 1 with the weighted utility model above gives:

V(d1, d2, …, dn) = f(expect1).V(d1) + f(expect2).V(d2) + … + f(expectn).V(dn) … Model 4

Model 4 predicts that the evaluation of a sequence of delays is a function of the individual expectations for each event, in addition to the negative utility from the individual delays. The implication of Model 4 for the evaluation of delay sequences is that when the pattern of delays are in the form (d1<d2<…<dn), it does not always follow that V(d1, d2, …, dn)>V(dn, …, d2, d1). In other words, the evaluation of a decreasing delay sequence is not always more favourable than an increasing delay sequence. The reason is because the weights from expectations may not have an end-loaded distribution, which is a necessary condition for an evaluation predicted by Model 1 to hold. Instead, it is proposed that the evaluation of, and preference for, a delay sequence is conditional on expectations. Therefore, the first hypothesis of this study is written as:

H1: Preferences for delay sequences track expectations

This hypothesis is an adaptation of the one used in Chapman's (2000) study. Chapman (2000) uses the word 'track' to mean that when people consider a decreasing sequence to be more realistic than an increasing sequence, the direction of this expectation will also be observed in preferences, where the decreasing is preferred to an increasing pattern. The test of this hypothesis was operationalised as follows. The general template of the stimuli, dependent and independent variables used in the present study was largely derived from the previous Domain Study. The present study also used a similar stimuli and independent variable to the Domain Study, where participants were required to indicate which of the two sequence patterns, increasing or decreasing delay, was more frustrating. In addition to measuring preferences, however, the present study also included an additional dependent variable that measured expectations. This dependent variable, which was designed by Chapman (2000), measured expectations by requiring participants to indicate on a Likert scale which of the sequence patterns (increasing or decreasing delay) was more realistic. As in Chapman's (2000) study, a between-participants design was used to measure the two dependent variables used, to avoid the possibility of preference unduly leading, or being lead by, expectations, if a within-participants design is used.

In addition to the comparison between preferences and expectations, a within-participants factor, 'combination', which systematically manipulates the individual delay durations with the event that causes the delay, was also introduced. The purpose of this manipulation is to test the proposition that people's preferences for different combinations of delay duration and event vary. This proposition seeks to demonstrate that different combinations of delay duration and event evoke different expectations, which then leads to different preferences. This result implies that variations in preferences are observed because some combinations of event and delay duration are more realistic than others. The hypothesis is written as:

H2: Preferences differ when the combination of event and duration in a delay sequence is varied

In addition to varying the combinations of event and delay duration, a between-participants factor, which uses a completely different set of four events to the set used in the Domain Study, was also introduced. To differentiate between the two sets of events, the first set (i.e. the one taken from the Domain Study) was called Set1, and the new one introduced, Set2. It is predicted that expectations, and therefore preferences, would be different for the two sets of events, because the combinations of event and delay duration that is more realistic for one set may be less for the other, and vice versa. This prediction was based on Chapman's (2000) results, which showed that expectations and preferences also varied between different sub-domains (e.g. athletic ability vs. facial complexion) within the wider domain of health sequences. In the context of this study, Set1 and Set2 represented different sub-domains or scenarios of delay within the wider domain of delay sequences. The hypothesis is as follows:

H3: Preferences for delay sequences vary for different sets of events

6.3.2 Method

The methodology of the present study was largely based on Chapman's (2000) design, but adaptations were made where necessary to suit the different stimuli used and hypotheses developed.

6.3.2.1 Participants

One hundred and twenty-eight undergraduates at a large Malaysian university volunteered to participate in the study. Their ages ranged from 18 to 25 years old. Sixty-nine participants were female. No payment was made to them.

6.3.2.2 Design, stimuli and procedure

There were four questionnaires in this study, each of which was independently administered to thirty-two randomly selected participants. The first questionnaire represented the condition that measured preferences for the set of events, Set1. The second questionnaire measured expectations for Set1, while the third and fourth questionnaires measured preference and expectations for Set2 respectively. This was represented in Figure 6.2 below.

Figure 6.2: The four questionnaires administered in the study

Set of events	Dependent variable measured: Preference	Dependent variable measured: Expectations
Set1	Questionnaire 1	Questionnaire 2
Set2	Questionnaire 3	Questionnaire 4

Figure 6.3: The event-delay combinations used for both sets

Set1					
Delay (minutes)		Combination			
increasing	decreasing	(a1,b1,c1,d1)	(b1,c1,d1,a1)	(c1,d1,a1,b1)	(d1,a1,b1,c1)
5	60	Hotel checkout	Taxi wait	Trip to station	Ticket queue
15	40	Taxi wait	Trip to station	Ticket queue	Hotel checkout
40	15	Trip to station	Ticket queue	Hotel checkout	Taxi wait
60	5	Ticket queue	Hotel checkout	Taxi wait	Trip to station

Set2					
Delay (minutes)		Combination			
increasing	decreasing	(a2,b2,c2,d2)	(b2,c2,d2,a2)	(c2,d2,a2,b2)	(d2,a2,b2,c2)
5	60	Post office	Try clothes	Dentist	Concert tickets
15	40	Try clothes	Dentist	Concert tickets	Post office
40	15	Dentist	Concert tickets	Post office	Try clothes
60	5	Concert tickets	Post office	Try clothes	Dentist

For each of the four questionnaires in Figure 6.2, participants were asked to imagine that they face four consecutive delays. Unlike in the Domain Study, however, participants were not told that the delays were experienced as part of a wider scenario of leaving a hotel to catch a train. The reason for not providing a broader context to describe the delays is to keep the four delays separate. This was done to allow for the sequence order of the four events to be varied using a Latin-squares procedure, so that H2 can be tested. The scenarios for the four events used in Set1 were: 1) to check out of a hotel, 2) to wait for a taxi to arrive, 3) to travelling to the train station, and 4) to queue for a train ticket. For future referencing purposes, these four events were labelled a1, b1, c1, and d1 respectively (see Figure 6.3). The four events used in Set2 were: 1) delays at a post office, 2) delays waiting to try clothes, 3) delays at a

dentist, and 4) delays when buying concert tickets, labelled a2, b2, c2 and d2 respectively. The events for Set1 were chosen on the basis that they were similar to the events used in the Domain Study, which enabled comparisons to be made between the two studies. The events for Set2 were selected on the basis that they are commonly experienced events that most people should have experience of. The events of Set2 were also different from the events in Set1, to enable comparisons to be made between preferences for different events. The parameters used for the increasing and decreasing delay sequences were {5,15,40,60} minutes and {60,40,15,5} minutes respectively for all four questionnaires. The delay parameters were identical to the ones used in the previous Transport Delay and Domain studies.

While the emphasis of the scenarios used in the present study assumed that participants were able to imagine experiencing delays as described in Set1 and Set2, it is argued that the utilisation of such a method is an essential and perhaps indispensable tool for testing new ideas in sequence research, as had been demonstrated by the literature. For example, the tasks in Varey and Kahneman (1992) required participants to imagine that a fictitious person, who was in an uncomfortable state (such as sitting in a vibrating room, being exposed to loud drilling noises or standing uncomfortably), was relaying to them (i.e. participants in Varey and Kahneman's study) moment-to-moment responses of their emotions. Participants then had to evaluate and provide a global retrospective evaluation of these responses. In another case, one of the experiments in Loewenstein and Prelec (1993) required participants to provide their evaluations of a sequence of imaginary weekends that varied in the intensity of pleasure (i.e. it is boring, moderately pleasurable, very pleasurable etc.). In the next example, Chapman's (1996a) study required participants to imagine that they are currently 20 years old and would go on to live for another 60 years. As a final illustration, Diener et al.'s (2001) study presented a series of vignettes that described the fictitious life of people who lived consistently happy lives or depressed lives, which participants had to provide evaluations of. While all of these prospective studies may have relied on participants being able to imagine themselves in the hypothetical situations described above, it is argued that the validity of the conclusions drawn from the

results of these studies was high, because the theories developed in these studies were also found to be applicable in many subsequent studies. For example, as previously mentioned, the peak-end theory was first conceptualised in Varey and Kahneman (1992), which were applied in other studies (e.g. Kahneman et al., 1993; 1997; Schreiber and Kahneman, 2000). Loewenstein and Prelec (1993) was another seminal paper that provided the theoretical foundations of the preference for an improving sequence, which was cited in all subsequent studies on sequences that was reported in Table 2.1 and 4.1 in chapters 2 and 4 respectively.

The remainder of this section describes the dependent variables used and the organisation of the study. The dependent variable used for questionnaires 1 and 3, which measured participants' preference for either the increasing or decreasing delay sequence, was adapted from the one used in the previous studies in this chapter (see Appendix 6.3), which was also similar to those used by Chapman (2000). The dependent variable in questionnaires 2 and 4 measured participants' expectations by asking them to rate on a 7-point Likert scale which of the two sequence patterns (increasing or decreasing delay) was more realistic. On the scale, a rating of 1, 2 or 3 indicated that the increasing delay sequence was more realistic; 4 indicated that both sequences were equally realistic (or unrealistic); and 5, 6 and 7 indicated that the decreasing sequence was more realistic. The scale used was adapted from the one used by Chapman (2000) to fit the context of delays in the present study.

To test H2, different combinations of delay durations and the events that caused the delays were introduced in a within-participants factor, labelled 'combination', for each of the four questionnaires in Figure 6.2. Since there were four different events in each sequence, a Latin-squares procedure was used to generate four separate combinations of events and duration. The combinations were (a1,b1,c1,d1), (b1,c1,d1,a1), (c1,d1,a1,b1) and (d1,a1,b1,c1) for Set1, and (a2,b2,c2,d2), (b2,c2,d2,a2), (c2,d2,a2,b2) and (d2,a2,b2,c2) for Set 2 (see Figure 6.3). In other words, all participants were presented with four different scenarios consisting of the Latin-square combination of event and duration that they were required to evaluate either

their preference or expectations for the delay sequences. These four scenarios were presented consecutively. Since participants were required to evaluate all four scenarios, this would reveal whether participants have different preferences and expectations for the various combinations of event and duration. For example, participants might prefer the decreasing delay sequence for the combination (a1,b1,c1,d1), because they expected the decreasing delay sequence to be more realistic than the increasing delay sequence, but could prefer the opposite sequence pattern for (d1,a1,b1,c1) and/or (a2,b2,c2,d2), because of differing expectations.

To counter-balance the order of presentation of the scenarios, two further between-participants procedures were carried out. The first, labelled 'order of sequence pattern', varied the order of appearance of the sequences, where half of the participants for each questionnaire were given the increasing before decreasing delay condition, while the other half, the opposite. For the second procedure, labelled 'order of scenarios', a Latin-square design was introduced to balance the order in which the scenarios appeared. Given that there were four scenarios for each questionnaire in Figure 6.2, the Latin-square procedure generated four different conditions. Let w1=(a1,b1,c1,d1), x1=(b1,c1,d1,a1), y1=(c1,d1,a1,b1) and z1=(d1,a1,b1,c1) for Set1, and similarly for Set2, w2 to z2. The order of appearance for each scenario is from left to right, e.g. w1=first scenario, to z1=last scenario. The four between-participants conditions were (w1,x1,y1,z1), (x1,y1,z1,w1), (y1,z1,w1,x1) and (z1,w1,x1,y1) for Set1, and similarly for Set2.

To summarise, the overall design for each of the four questionnaires in Figure 6.2 is a 4 ('combination', within-participants) x 2 ('order of sequence pattern', between-participants) x 4 ('order of scenarios', between-participants) factorial model, where the last two factors were counter-balancing considerations. Briefly, the test for the first hypothesis, H1, is to compare the responses to questionnaire 1 with questionnaire 2, and questionnaire 3 with questionnaire 4 respectively. The test of H2 is to determine the significance of the differences in the factor 'combination'. Last but not least, the test of H3 is to compare the

responses to questionnaire 1 with questionnaire 3. Appendix 6.3 presented the original questionnaires used in the present study.

6.3.3 Results

6.3.3.1 Description of results

Table 6.2: Results of the Expectations Study

Set1	Combination			
	(a1,b1,c1,d1)	(b1,c1,d1,a1)	(c1,d1,a1,b1)	(d1,a1,b1,c1)
Preferences *	25.0%	34.4%	59.4%	40.6%
Expectations **	-0.53 (1.78)	-0.16 (1.63)	0.34 (1.79)	-0.28 (1.76)
Rescaled preferences ***	-50.0%	-31.3%	+18.8%	-18.8%
Rescaled expectations ***	-17.7%	-5.2%	+11.5%	-9.4%
Chi-squared test ****	$\chi^2(1)=8.0$, $p<0.01$	$\chi^2(1)=3.1$, $p>0.05$	$\chi^2(1)=1.1$, $p>0.05$	$\chi^2(1)=1.1$, $p>0.05$

Set2	Combination			
	(a2,b2,c2,d2)	(b2,c2,d2,a2)	(c2,d2,a2,b2)	(d2,a2,b2,c2)
Preferences *	31.3%	53.1%	65.6%	53.1%
Expectations **	-0.53 (1.95)	0.56 (1.87)	0.22 (2.09)	-0.41 (1.64)
Rescaled preferences ***	-37.5%	+6.3%	+31.3%	+6.3%
Rescaled expectations ***	-17.7%	+18.8%	+7.3%	-13.5%
Chi-squared test ****	$\chi^2(1)=4.5$, $p<0.05$	$\chi^2(1)=0.13$, $p>0.05$	$\chi^2(1)=3.1$, $p>0.05$	$\chi^2(1)=0.13$, $p>0.05$

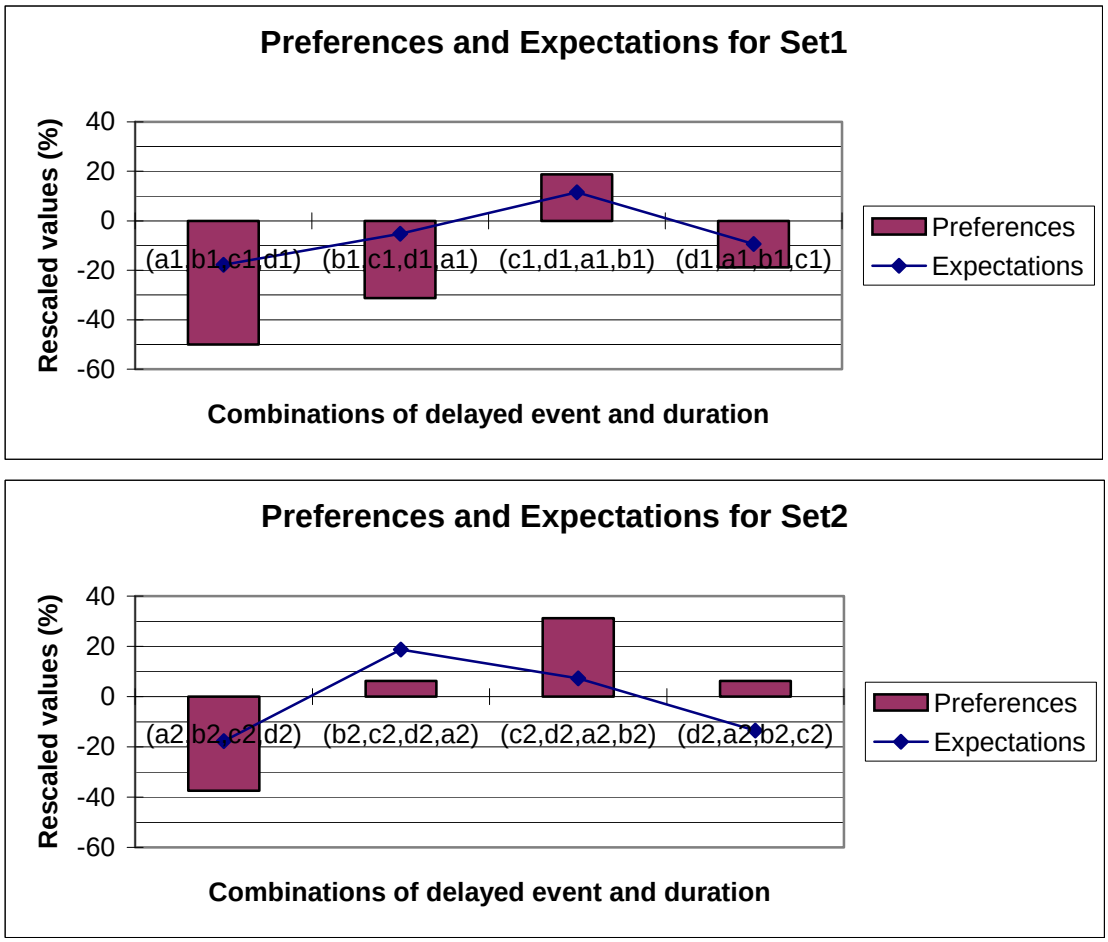
Key to labels for 'combination':				
Set1	'a1'=check out of a hotel	'b1'=wait for a taxi	'c1'=travel to the station	'd1'=queue for a ticket
Set2	'a2'=queue at a post office	'b2'=queue to try on clothes	'c2'=queue for dental treatment	'd2'=queue for concert tickets

Notes:
* Percentage of participants who preferred the decreasing delay sequence (N=32).
** Mean (standard deviation in brackets) ratings of expectations on a –3 to +3 scale, where positive numbers indicated that the decreasing delay sequence was more realistic, while negative numbers indicated that the increasing delay sequence was more realistic (N=32).
*** Preference and expectations ratings expressed as a –100% to +100% scale to enable the direct comparison of the magnitude and direction of the responses. Positive percentages meant that the decreasing time sequence was preferred or more realistic and vice versa for negative numbers. The direction of preference tracked expectations if they were both of the same signs.
**** Chi-squared test of significance for preferences (but not expectations), with 50% used as the expected value.

Table 6.2 reported the preferences and expectations for the four questionnaires. The first row (preferences) for both Set1 and Set2 reported the percentage of participants who preferred the decreasing delay sequence for each 'combination'. The values were reported in such a way that values under 50% indicated that the majority preference was for the decreasing delay sequence; values over 50% indicated that the majority preference was for the increasing delay sequence. The second row (expectations) reported the mean of expectations ratings, which was converted from the 1 to 7 scale used to a –3 to +3 scale, where positive numbers indicated that the decreasing delay sequence was more realistic, while negative numbers indicated that the increasing delay sequence was more realistic. The purpose of converting this scale is to make it easier to compare expectations with preferences. For example, a positive expectation rating and a preference value of less than 50% would indicate that participants preferred and expected the decreasing delay sequence.

The third and fourth rows for Set1 and Set2 in Table 6.2 reported the results of the rescaled preference and expectations values presented in the first two rows. The purpose of this second rescaling exercise is to present the reader with a clearer impression of the differences in magnitude, as well as direction, of the preferences and expectations for the various combinations of events and durations. The rescaling converted preference ratings reported in the first row from a 0-100% scale, where the mid-point 50% represented equal preferences for the patterns, to a –100% to +100% scale, with 0 representing the rescaled mid-point. For expectations ratings, the re-adjustment converted the scale from one of absolute deviations from 0 (i.e. a –3 to +3 scale), to one that represented percentage deviations from 0 (i.e. a –100% to +100% scale). The effect of the rescaling exercise is such that all positive values reflected the preference and expectation for the decreasing delay sequence, while all negative values reflected the preference and expectation for the increasing delay sequence. Values with larger magnitudes indicate stronger preferences or expectations. The rescaled values were then used to create a combination line and bar chart of preferences and expectations for Set1 and Set2, which were presented in Figure 6.4.

Figure 6.4: Re-scaled results of preferences and expectations for sequences



Notes:
Set1:
'a1'=check out of a hotel; 'b1'=wait for a taxi; 'c1'=travel to the station; 'd1'=queue for a ticket
Set2:
'a2'=queue at a post office; 'b2'=queue to try on clothes; 'c2'=queue for dental treatment
'd2'=queue for concert tickets

Positive values indicate preferences/expectations for the decreasing delayed sequence, while negative values indicate preferences/expectations for the increasing delayed sequence.

The use of bar and line combinations for the charts were used to illustrate the mean preferences and expectations respectively. This combination of graphics was chosen because a chart consisting of solely of coloured bars may appear too cluttered due to the number of means that have to be presented for both preferences and expectations. The use of bars and lines are purely for illustrative purposes and do not imply any relationship between preferences and expectations that have not been previously reported in the analyses of results.

Figure 6.4 compares the preferences with expectations for all four combinations of event and duration. The values for expectations appear to be a better fit for preferences in Set1 compared to Set2. In Set1, preferences were in the same direction as expectations for all four combinations of event

and duration, whereas in Set2, only three out of the four were in the same direction. The statistical significance of the goodness-of-fit would be calculated in the next sub-section 6.3.3.2 using a Pearson correlation test. The remainder of this sub-section, however, describes the numbers presented in the fifth and final row of Table 6.2. In this row, a chi-squared test of the significance of preference for a sequence pattern was calculated. The results showed that the preference for a sequence pattern, whether for the decreasing or increasing delay sequence, was only significant for two out of the eight possible combinations in both Set1 and Set2. In Set1, the preference for the increasing delay sequence was significant (chi-squared(df=1)=8.0, $p<0.01$) only for the combination (a1,b1,c1,d1), while for Set 2, the preference for the increasing delay sequence was also significant (chi-squared(df=1)=4.5, $p<0.05$) for the combination (a2,b2,c2,d2).

This result suggested that people did not have a significant preference for the decreasing delay sequence for any of the combinations, which is in contrast with the results of Transport Delay Study, where preference for the decreasing delay sequence was shown to be significant. This suggested that the preference for a sequence of decreasing delays was moderated by other considerations, such as domain (represented by the various combinations of events and duration). The results of the present study also provided further evidence that the preferences measured were consistent with the previous Domain Study. This was shown when the significant preference for the increasing delay sequence in the combination (a1,b1,c1,d1) of Set1 of the present study replicated the results of the Domain Study, which it shares a similar stimuli, dependent and independent variable with. The ability to replicate the results of the Domain Study is important because it provides evidence that preferences are consistent across participants recruited from different geographical locations. This consistency provides a basis to further generalise the theories developed in this thesis.

6.3.3.2 Tests of the first hypothesis, H1

To test the hypothesis H1 that preferences for delay sequences track expectations, Pearson's correlation coefficient, r, which measures the similarity

between the distributions of preferences and expectations, was calculated. Distribution here is defined as the collection of preferences and expectations from the four combinations. This was a method that Chapman had developed with Loewenstein to provide a formal test of the link between preferences and expectations (see Chapman, 2000). Across the eight combinations of event and duration for both Set1 and Set2, the mean expectation rating was significantly correlated ($r=0.741$, $N=8$, $p<0.05$) with preferences (i.e. row 1 of Table 6.2). This result supported the hypothesis H1 that preferences for sequences of delay tracked expectations. The results for Set1 and Set2 were combined together in the calculation of the Pearson correlation in the present study because this enabled the widest possible comparison between the distributions of preferences and expectations, as the comparison covers all of the combinations of events and duration investigated in the present study. The effect of combining the results of Set1 and Set2 is similar to increasing the sample size, N, and thus power, used to test H1. This was a method also used by Chapman (2000) to describe the overall effect of expectations on preferences for all four different sub-domains of health sequences that she had examined.

In addition to the Pearson test performed on the combined data above, individual Pearson tests were also performed on Set1 and Set2 respectively, to obtain a measure of the degree to which preferences track expectations for each set. Across the four combinations of event and duration for Set1, the mean expectation rating was significantly correlated with preferences ($r=0.952$, $N=4$, $p<0.05$). Similarly, a Pearson correlation coefficient was also calculated for Set2, which represented a different set of events from that used in Set1. Across the four combinations of Set2, the correlation between the mean expectations rating with preferences, however, was not significant ($r=0.629$, $N=4$, $p>0.05$), despite r being greater than 60%. The implications of the result of the Pearson correlation for Set2 suggest that preferences are not always influenced by expectations for every combination of event and duration.

It is further suggested that the non-significance of this result could, to a large extent, be attributed to the divergence between preferences and expectations

for combination (d2,a2,b2,c2). Figure 6.4 showed, for the combination (d2,a2,b2,c2), that people prefer the decreasing delay sequence, but expected the increasing delay sequence. It is, however, not uncommon for preferences to diverge from expectations, as Chapman's (2000) results had shown. The reason that Chapman gives is that expectations are only one (albeit an important one) of a number of factors that can influence preference. This reason, along with other possible explanations, will be expanded in the discussion section (6.3.4) of the present study.

6.3.3.3 Tests of the second hypothesis, H2

To test the hypothesis H2 that preferences for delay sequences differ when the combination of event and duration is varied, a non-parametric Friedman test on the preferences for various combinations of event and duration was calculated for both Set1 and Set2. For Set1, the test result was significant (N=32, chi-squared(df=3)=8.73, $p<0.05$), meaning that preferences were significantly different across the four different combinations of event and duration. For Set2, however, the test was only marginally significant at the 10% level (N=32, chi-squared(df=3)=6.67, $p=0.083$). These results showed that preferences for delay sequences were not uniform, but sensitive to the domain or events that caused the delays. This is important because it implies that domain (through expectations, or other factors) has an influence over preference.

6.3.3.4 Tests of the third hypothesis, H3

To test the hypothesis H3 that preferences vary for different sets of events, a Mann-Whitney test that compared the preferences for Set1 with the preferences for Set2 was calculated. The Mann-Whitney test calculates the mean and sum of ranks of the preferences in Set1 (mean rank=115.5; sum of ranks=14784) and contrasts them with those of Set2 (mean rank=141.5; sum of ranks=18112). The result of the test was significant ($Z=3.24$, $p<0.001$), indicating that preferences for Set1 were different from Set2, which supported the hypothesis H3.

Additional tests were also carried out to determine whether expectations were also different between the two sets. A one-way ANOVA was performed on the

mean expectations ratings of Set1 and Set2, but the results indicated that the overall mean expectations for both sets were not significantly different ($F(1,255)=0.26$, $p>0.05$). This result, however, does not suggest that expectations across both sets were similar, because the ANOVA test only measures overall mean differences, but does not discriminate between the differences in the distribution of expectations for individual combinations of event and duration. For example, the ANOVA was unable to discriminate between a distribution with expectations ratings of {0,0,0,0} with a distribution of {-3,-1,+1,+3}, because the overall means for both distributions were identical.

It is, however, important to be able to discriminate between the individual distributions of expectations ratings to determine whether people's expectations vary with the order in which the events appear (i.e. the factor 'combination'). To increase the level of discrimination, a Pearson correlation coefficient, or r, was calculated by comparing the mean expectations ratings for Set1 with Set2. A significant test result would suggest that the distribution of expectations for both sets were similar. The Pearson test, however, was not significant ($r=0.628$, $N=4$, $p>0.05$), which suggests that the distribution of expectations for Set1 was dissimilar to Set2. In other words, while the overall mean expectations across all four combinations of event and duration might be similar for both sets, their distributions were shown to be different. This suggested that expectations, like preferences, also varied across different domains. Table 6.3 summarises the hypotheses and main results of the present study.

Table 6.3: Summary of hypotheses and main results

H1: Preferences for delay sequences would track expectations The Pearson correlation coefficient comparing the distribution of preferences and expectations for both Set1 and Set2 was significant, supporting H1 ($r=0.741$, $N=8$, $p<0.05$)
H2: Preferences differ when the combination of event and duration in a delay sequence is varied For Set1, the Friedman test was significant, supporting H2 ($\chi^2(3)=8.73$, $p<0.05$) For Set2, the Friedman test was marginally significant, providing some support for H2 ($\chi^2(3)=6.67$, $p=0.083$)
H3: Preferences vary for different sets of events The Mann-Whitney test comparing the preferences of Set1 to Set2 was significant, supporting H3 ($Z=3.24$, $p<0.001$)

6.3.4 Discussion

The primary objective of the present study was to test the conjecture that the mechanism in which domain influenced preferences for delay sequences is through expectations. This conjecture was first proposed in the Domain Study, where it was found that people did not always prefer a decreasing delay sequence as predicted by Model 1. It was then suggested that the observed preference reversal in the Domain Study could be attributed to differences in expectations for individual combinations of event and duration, where expectations is defined as the degree to which the duration of delay is realistic for a given event that caused the delay. The prediction was that people would prefer the more expected to the less expected sequence. This conjecture was then developed into a theory to explain the findings of the previous Domain Study, which was represented by Model 4. This model postulated that the overall evaluation of a delayed sequence is a function of individual expectations for each event. Three related hypotheses were developed to test Model 4.

The first and main hypothesis, H1, predicted that preferences would track expectations. The results of the Pearson correlation test supported this hypothesis, but the analyses also showed that the degree to which preferences

tracked expectations varied. It was shown that the direction and magnitude of preference was highly correlated to expectations for only some, not all, of the combinations of events. For example, it was observed in the combination (d2,a2,b2,c2) of Set2 that the decreasing delay sequence was preferred, even though the increasing delay sequence was more realistic. This result should not be construed as being an anomaly, because, as Chapman (2000) has argued, expectations were only one (albeit important) factor that influenced preference for sequences. This analysis implies that the influenced of expectations over the preference for a delay sequence would be stronger for some combinations of event and duration than others.

Other psychological factors that may also influence preference was suggested by Chapman, and include, in no particular order, wishful thinking (Mascaro, 1969; Babad and Katz, 1991), and ego bias (Dawson and Arkes, 1987). In wishful thinking, preference influenced expectations, instead of the other way around, where people impose their desired outcomes on their expectations, which then makes expectations unrealistic. The effect of ego bias was similar, in which people distorted the likelihood of an event occurring to fit their personal goals or preferences. For example, surgeons over-rate their surgical skill by judging the mortality of their operations as being lower than other surgeons (Detmer et al., 1978). In the context of delays, ego bias could materialise when people under-estimate the amount of delay they may face, because they over-estimate the levels of service that they would receive. For example, an arrogant and deluded person may under-estimate the amount of delay that he/she faces, and consequently, the negative utility from that delay. This is because he/she assumes that he/she will always be served first, ahead of anyone else, for all situations, even if this is not the reality.

The second hypothesis, H2, predicted that preferences differ when the combination of event and duration in a delay sequence varied. The results of the tests, however, were mixed. The preferences for the combinations were significantly different for Set1, but only marginally significant for Set2. These results demonstrate that preferences are generally sensitive to differences in the combinations of event and duration, but that the degree of sensitivity can

also vary between the combinations. The variability of preferences across the different combinations in both Set1 and Set2 provide evidence to support the assertion of Model 4 that the distributions of weights vary according to the expectations of each event and duration combination. The results also show that the preference for a sequence pattern is stronger for some combinations of event and duration, but less so for others. The highly variable nature of preferences for delay sequences suggests that it is difficult to construct a model that could predict preferences for all delay sequences. This is because the variety of combinations of events and duration is too wide to be covered by a single model alone. Therefore, the predictability of preferences for delay sequences, as implied by Model 1, is limited to sequences of delays that are experienced under specific circumstances, such as in the identical domains used for each stage of delay in the Transport Delay Study reported earlier in section 6.1, where all delays are from waiting for public transport to arrive.

The third hypothesis, H3, predicted that preferences would vary for different sets of events, is an extension of the second hypothesis. The results of the tests that compared the preferences and expectations of Set1 to Set2 found that both preferences and expectations differed across the sets, supporting H3. This result provided more evidence to support the prediction that even for events within a similar domain (health for Chapman's (2000) study, delays for the present study), expectations of, and preferences for, individual events that make up a sequence vary enormously. As concluded in the previous paragraph, the implications of the results for H3 suggests that it would be difficult to generalise the preference for a delay sequence, especially when the events that causes the delays are different across the sequence.

To summarise the results of all three hypotheses tested, it is concluded that, in the domain of delay sequences, expectations has a significant influence over preferences. Expectations were shown to significantly vary for different combinations of event and delay duration, which made it more difficult to predict the direction of preference for delay sequences. This is because sequence evaluations are based on a combination of the individual expectations of each event and the negative utility generated from the delay.

Chapman (1996a) proposed that the reference point effect is a mechanism that explains why expectations influence sequence preferences. The reference point effect predicts that people prefer a sequence that is expected to one that is unexpected, because the unexpected sequence induces net negative utilities due to the asymmetry of the value function.

A research topic worthy of further consideration is to better understand the basis on which expectations are formed. The anecdotal explanation offered for the present study is that expectations are formed on the basis of people's observations of real-life delays. As suggested in the introduction, tennis fans that want to watch live matches at Wimbledon can expect to queue for hours, if not longer, for tickets. Therefore, it is argued that because people observe that such long queues are commonplace for Wimbledon, the observed length of delays then become the norm (Kahneman and Miller, 1986) or expectation for other similar events. In other words, people consider it normal to queue for hours if they want to watch such events, and their reference point for what is a realistic duration for queuing changes accordingly. In addition, norms can also be based on outcomes that occur naturally. For instance, Kahneman and Ritov (1994) found that environmental damage from natural sources are viewed less negatively than identical damage from human causes. In a separate example, this time in the domain of delays, Loewenstein (1988) found that the willingness of people to pay a higher price to speed up a delayed outcome was less than the price demanded to delay an immediate outcome. This implies that people have a preference for experiencing a particular outcome at its expected time. Both of these examples provide further evidence to show how expectations can influence preference.

On a final point, the finding that the results of the combination (a1,b1,c1,d1) in the present study replicated that of the Domain Study is noteworthy, because this replication provided further support for the basis on which results obtained from one sample of participants could be generalised to others. While the present study was not set up for, nor intended, to investigate cross-cultural aspects of preferences for delay sequences, the fact that preferences from the British-based sample in the Domain Study were replicated in the present study,

which were obtained from participants based in Malaysia, suggests that there are trans-cultural commonalities in the rationales and values that influence the preference for delay sequences. Future studies that explicitly focus on cross-cultural values could perhaps analyse the extent to which expectations influenced people's preference for sequence patterns for various everyday experiences common across different cultures. The conclusions of such a study would have useful implications for marketers, who could assess various tolerances for sequences of delays across cultures, and tailor the consumer experience accordingly. More importantly, such a study would also provide further evidence on the degree to which expectations influence preferences.

To conclude, the present study has shown that expectations have a significant influence over preferences for delay sequences. The mechanisms by which expectations influenced preference was explained using a model based on those developed in previous studies in this thesis. There are some exceptions, however, to this conclusion. Topics for future studies that would build on the theories developed in the present study were also discussed. The next and final section of this chapter provides an overall summary of the themes that link all three studies, Transport Delay, Domain, and the present study, and concludes the chapter.

6.4 Chapter summary and conclusions

The broad theme of the three studies in this chapter was to extend the model developed in the previous chapter, Model 1, to sequences of delays with longer durations. Model 1 postulated that because people place more emphasis or weight on the end-characteristics of an aversive sequence, therefore a sequence with less aversive end-characteristics, such as a decreasing delay sequence, would receive less negative evaluations compared to a sequence with more aversive end-characteristics, such as an increasing delay sequence. The studies in chapter 5 found that the predictions of Model 1 hold only if, for a sequence of time-sensitive tasks, the delays had significantly obstructed task achievement. It was suggested that on its own, the delays, as utilised in the four studies of chapter 5, was unable to generate

sufficient negative emotions to induce evaluation based on patterns, because they were too short in duration to have really mattered in people's experiences (short durations were defined as being less than 3 minutes for the purposes of the present chapter).

Therefore, the purpose of the first study in the present chapter (the Transport Delay Study) was to test Model 1 for longer durations of delay. A duration parameter of two hours was used. The results of the study supported Model 1's prediction that people prefer a decreasing delay to an increasing delay sequence. This finding supported the assertion that there are two mechanisms in which delays can significantly influence sequence evaluation and preference. The first is through the length of the delays, where if the duration of delay is long enough, people will become frustrated when they experience it. The second is through the obstruction of a time-sensitive task, such as a memory recall or visual search task. Frustration here is generated because the delay has impeded people's ability to fully achieve the task.

The purpose of the second study (Domain Study) was to explore preferences for delay sequences when the domain of delay was varied. In the non-sequence literature review on delays in chapter 2, it was demonstrated that there were many factors other than duration of delay that could also influence the evaluation and preference for delayed experiences. It is important to understand the effects of having different domains for each delay on the overall preference for a sequence of delays, given that preferences for individual delays can vary, depending on the domain in which it is experienced. Because this study was exploratory in nature, however, no hypothesis was developed. The results of this study showed that people evaluated the increasing delay sequence less aversive than the decreasing, which was opposite to the prediction of Model 1. This implied that variations in the domain of delays introduced as a manipulation had significantly influenced preference. It was suggested that the mechanism through which domain influences preferences for delay sequences is through expectations, which is the subject of the next and final study of this thesis.

In the Expectations Study, a formal model, which was developed based on Model 1, was introduced to explain how preferences are affected when a delay sequence comprises of delays from different domains, as observed in the Domain Study. In Model 4, it was posited that people's expectations vary with the domain. Domain in this model is defined by two parameters, the event that causes the delay and the duration of delay. Expectations were defined as the realism of the duration for the event that causes the delay. Given that some combinations of event and duration might be more realistic than others, therefore it was predicted that expectations vary when domains vary. In Model 4, these expectations were represented by the weights of the individual utilities from each delay within a sequence. This meant that the distribution of weights was conditional on the individual expectations of delays, rather than being exclusively end-loaded, as in Model 1. Since the sequence evaluation is a function of the weighted average of the utilities from the delays, therefore preferences will also vary when expectations vary. The prediction that preference vary with expectations was supported by the results of the study. The significant effect on preferences of varying domains makes it more difficult to predict the direction and strength of the preference for a delay sequence. It also suggests that the preference for a decreasing delay sequence would only be observed under specific circumstances, such as when the domains of delays within a sequence are all identical.

The overall conclusion that is drawn from the results of three studies paints a complex picture of the preference for delay sequences. People prefer a decreasing delay pattern when the domains of delays are similar, but when domains vary, preferences were also shown to have varied. Expectations were demonstrated to be one of the significant factors that explain the domain dependency of preferences. The challenge for future studies would be to explore in greater detail other factors that can also affect preferences for delay sequences.

Chapter 7

Summary and conclusions

7.0 Introduction

The primary purpose of this chapter is to summarise the key contributions that this thesis makes to the literature and to conclude the thesis. The material in this chapter is organised as follows. The first section (7.1) of this chapter provides a synopsis of the key contributions made. These contributions are then discussed in greater depth in the next the section (7.2), which summarises key developments and findings in the other chapters of this thesis. This summary includes an analysis of the key contributions based on four broad themes. The first and second themes demonstrate the theoretical contributions to the delay literature and to the sequence literature, the third theme outlines the methodological contributions, and the final theme provides the practical implications of the findings to retailers. While the contributions described in these themes are interwoven into the broader summary of the section, there would be signposting for the readers in areas where a key contribution is made. The rest of the sections of this chapter outline the limitations (section 7.3), future progression of the thesis (section 7.4), and a conclusion (section 7.5).

7.1 Synopsis of key contributions

The main contribution of the research in this thesis has been to provide a better understanding of the factors that influence the evaluation of delay sequences. It was initially recognised that while there is a growing commercial need to understand how sequences of delays affect consumer experiences, there was little research to date on delay sequences, in either the delay literature or in the sequence literature. In addition there has previously been only one attempt (i.e., Carmon and Kahneman, 1993) to model the evaluation of a series of delays using the theories developed in the sequence literature,

which suggested that further development was needed. This lead to the development of a model (Model 1) on the evaluation of delay sequences in the Internet Delay Study in chapter 2, which was further refined in subsequent studies in chapters 5 and 6. Briefly, Model 1 predicts that pattern effects, such as the preference for an improving sequence, are also present in sequences of delays. The derivation of Model 1 and its refinements (Models 2 to 4) is revisited in the next section (7.2).

The studies in chapter 5 and 6 have shown that the theories developed in the sequence literature, as exemplified by the four models (Model 1 to 4), can be used to model sequences of delays. The theoretical formulation of the models and its empirical testing represent a major contribution to both the sequence literature and the delay literature. The demonstrated applicability of theories developed in the sequence literature to the domain of delays has enriched and expanded the sequence literature in two ways. First, it shows that the theories in the sequence literature can be extrapolated to the delay domain. Second, it provides support for the theoretical themes in the post-1999 literature, which has focussed on the factors that moderate pattern effects.

At a more specific level, the thesis has shown, through the studies reported in chapter 5, that the pattern of delays is important in the evaluation of sequences with short delay durations – people prefer an improving sequence of delays if the delays obstruct their ability to achieve their goals. The main finding here demonstrates that theories from the sequence literature can be applied to model sequences of delays, but only in situations where the delays significantly obstruct progression and achievement. At a broader level, the highly specific conditions needed to induce pattern effects in the Memory and Visual Search studies in chapter 5 is also in line with the broader findings of the post-1999 literature reported in chapter 4, where the moderation of pattern effects was one of the key topics discussed at a theoretical level. In summary, pattern effects also affect sequences of delays, but these effects are moderated by the extent to which the delays obstruct achievement.

The next major finding of this thesis was that, for sequences of longer delays (i.e., in terms of minutes and hours rather than seconds), the situation is more complex than is the case for sequences with shorter delays. In the three studies reported in chapter 6, it was shown that pattern effects were strongest when the delay domains are homogeneous across the sequence (i.e., all delays are from the same environmental context, such as when a person experiences a sequence of four delays, all from waiting for public transport), but weaker when they are heterogeneous (when the delays are caused by a series of different events, such as queuing at the post office, then queuing for concert tickets, then queuing to try clothes in a department store etc.). The reason is that the heterogeneity raises different expectations of the delays, which in turn, influences preferences for delay sequences. This finding, which illustrated the effects of domain heterogeneity on sequence preference, is new. Previous studies that investigated sequence preference for different domains, such as Chapman (2000), only examined inter-sequence domain heterogeneity (i.e., comparing the effects of preference between sequences with different domains), but not intra-sequence domain heterogeneity (i.e., where the domains vary within the sequence itself).

A comparison between the findings of the studies in chapter 5 with those of chapter 6 demonstrate that it is important to distinguish between different durations of delay, because it suggests that different psychological mechanisms (which would be discussed in greater depth in the next section 7.2) underpin the observed pattern effects for different durations. For very short delay sequences (e.g., studies in chapter 5), pattern effects were only observed when the delays were significant enough to obstruct achievement. For longer delays (e.g., studies in chapter 6), however, pattern effects were stronger, because they occur regardless of whether they obstruct achievement of goals.

The implications of this finding for the sequence literature suggest that the relationship between the delay duration and sequence evaluation varies, meaning that findings from a study conducted using one set of parameters might not easily extrapolate to other studies using a different set of

parameters. At present, no study in the sequence literature has explicitly researched the relationship between parameter choice and sequence evaluation, therefore this argument represents a new development for the sequence literature, because it claims that the choice of parameters is an important determinant of the strength of pattern effects in an evaluation of a delay sequence. Prior studies in the literature had often only utilised a narrow band of parameters as stimuli (e.g., Kahneman et al., 1993, only looked at a narrow range of temperature gradient, of 1 degree Celsius).

In addition to contributing to the knowledge of the sequence literature, the theories developed in this thesis (i.e., the four models described in chapter 2, 5 and 6) have also added a new dimension to the existing literature on delays. In particular, these models provide a theoretical base to understand how people evaluate sequences of delays, which, as in the sequence literature, is also inadequately researched in the delay literature. For example, previous research on delays in sequences has largely focussed on single periods (e.g. Dellaert and Kahn, 1999) or on just one type of sequence pattern (Carmon and Kahneman, 1993). Therefore, the extension of the knowledge base to a sequence of (multiple) delays has filled a missing gap in the delay literature.

In terms of the contributions of the findings to the world of commerce, retailers, whether internet-based or otherwise, could use these findings to help them predict the pattern of delay sequences that is most likely to generate the highest/lowest levels of annoyance to their customers, and decide on where best to position the delays if the delays could not be engineered out of the retailing experience. For example, if an internet retailer finds that people prefer a website to start quickly, but do not mind waiting at the check out, where their orders are being processed (i.e., this represents a preference for an increasing delay sequence pattern), the retailer could engineer the website such that all of the slower pages occur towards the end of the shopping experience. In the internet retailing business, the ability of a retailer to influence the positioning of delays within the retail environment is crucial, because the ‘last mile problem’ means that slow connections on the consumer side of the internet connection could generate delays and irritate the consumer, which the retailer might not

know about, as the retailer's web sites might be operating at very high speeds. In addition, online retailers should also ensure that the delays are not excessive to the point that it significantly obstructs task achievement, as pattern effects sequence might accentuate any negative experiences with delays.

In summary, the key contributions of this thesis to the literature are as follows. The first is the development of a new theoretical framework in which to analyse sequences of delays – the delay literature did not have sufficient coverage of how people would evaluate multiple delays that occur in a sequence. This theoretical framework culminated in the development of Model 1, which was subsequently refined in Models 2 to 4. The second and related contribution is to extrapolate the theories from the sequence literature to the domain of delays. The third contribution is to demonstrate that methodological issues such as the choice of parameters is important, as variations in parameters could invoke different psychological mechanisms in evaluations. The fourth contribution is to provide retailers with more insights into consumer behaviour towards sequences of delays, especially in situations where such delays are unavoidable.

7.2 Summary of thesis and key contributions

This section is organised as follows. First, this section will describe the initial motivation for this thesis and a précis of the key concepts that lead to the development of the first study (i.e. the Internet Delay Study or IDS) in chapter 3. Next, the section will summarise the key developments of the thesis in chronological fashion, to illustrate the developments that led to the key findings and contributions of this thesis.

7.2.0 Motivation

The initial motivation for this thesis was to explore issues related to consumers' experiences with internet delays. It was argued that research on the impact of such delays is important because of the growing importance of electronic commerce in everyday consumer experiences. There is evidence to suggest

that internet delays may significantly affect commerce in a detrimental way if they are not managed properly. One way to deal with this is to increase the processing capacity of websites to ensure that delays are minimised or eliminated altogether. However, this is not always possible with even the most up to date technology. This means that retailers need to understand how consumers react to delays, and to determine whether there are any other means to reduce the impact of delays should the delays become unavoidable.

The review of the general research literature on delays, reported in section 2.1 of chapter 2, showed that a large number of studies had investigated various means to moderate the negative impact of delays. These include methods such as providing information on the expected duration and the nature of the delays, which helps consumers to make adjustments to the level of service that they expect. However, it was argued that the use of these methods to moderate the negative impact of internet-generated delays was limited because the findings only apply to situations where delays were experienced in a single-period. In an internet experience, on the other hand, consumers experience multiple periods of delays in a sequence as they move from one webpage to another while surfing. This means that they face the likelihood of encountering delays at every change of page, as web pages require time to process the data and to fully load up. As internet delays are often experienced as multiple delays in a sequence, this made it necessary to understand the literature on how people evaluate sequences of delays, and sequences of other types of stimuli.

The research literature on consumers' evaluation of delay sequences, however, was relatively undeveloped and limited to a handful of studies at the time of the first study in this thesis (i.e. the Internet Delay Study) was completed. The theoretical approach adopted in these studies (e.g. Dellaert and Kahn, 1999; Carmon and Kahneman, 1993) had applied behavioural theories to model consumers' evaluations of delay sequences. Carmon and Kahneman (1993) showed that the end point of a sequence of queues has a disproportionate influence over consumers' evaluations, while Dellaert and Kahn (1999) demonstrated that the positioning of delays within an internet

sequence of events also affects evaluation. These two papers provided the starting point for the studies reported in this thesis, as well as its raison d'etre, which was to further explore the effects that delay sequences have on consumers' evaluations of their internet experiences. The broader literature on sequences, which was reviewed in chapter 2, provided a theoretical foundation on which to model consumer behaviour in sequences of internet-generated delays.

7.2.1 The development of a model to predict the evaluation of delay sequences

A model (Model 1) that predicted consumers' evaluation of a sequence of internet delays was developed in chapter 3, based on the findings in the general sequence literature that people attach end-loaded weights to a sequence of aversive events. A brief derivation and framework of the model is reproduced below. The main contribution of this framework to advancing the state of the literature is in bringing together various disparate concepts from both the sequence literature and the delay literature.

Let d=delay. If d1>d2 (i.e. one delay is longer than the other), then a comparison of the negative evaluation of the delays, represented by the utility function V(.), can be written as V(d1)>V(d2). In other words, the evaluation of d1 is more aversive than d2, because V(d1), which has a negative direction because the stimuli is aversive, is larger in magnitude than V(d2). This key assumption of how people would behave towards delays, which is derived from the delay literature reviewed in chapter 2, is then incorporated into the theory of sequences by using the weighted utility integration principle (UIP) postulated by Kahneman et al. (1997). The incorporation is demonstrated in the following equation.

$$U(x1, x2, x3,\ldots, xn) = w1.U(x1)+w2.U(x2)+w3.U(x3)+\ldots+wn.U(xn)$$

In this model, xn represented the individual components of an aversive stimulus, wn the weights, and U(xn) the negative utility from the stimulus. Given that many studies in the literature had reported that weights were end-

loaded (Fredrickson and Kahneman, 1993; Ariely, 1998), this meant that for a distribution of weights that is monotonically increasing (w1<w2<…wn), a comparison of sequence evaluation for an increasing versus a decreasing sequence would yield the following inequality.

w1.U(x1)+w2.U(x2)+ …+wn.U(xn) > w1.U(xn)+…+w(n-1).U(x2) +wn.U(x1) or

U(x1, x2, …, xn)>U(xn, …, x2, x1)

The inequality above implied that the evaluation of an increasing sequence of aversive components is more negative than a decreasing sequence. Model 1 was derived by extending this inequality to sequences of delay. Let (d1<d2<…<dn) be a sequence of increasing delay, and (dn, …, d2, d1) be a decreasing delay sequence, and assume that the distribution of weights are end-loaded, that is, (w1<w2<…wn). Therefore, replacing xn with dn would give:

V(d1, d2, …, dn)>V(dn, …, d2, d1) … [Model 1]

This model predicts that an increasing delay sequence would be evaluated more aversively than a decreasing delay sequence, which was the hypothesis tested in the IDS. This study required participants to evaluate their annoyance from the sequence of delays in a simulated internet retail website.

The results of the IDS, however, did not support the prediction of Model 1, as no differences between the evaluations of both sequence patterns were found. It was suggested that the design of the delayed tasks used in the study might have allowed other non-gestalt factors such as carry-on effects and expectations to inadvertently influence evaluation. No firm conclusions could be drawn from the results, however.

The unanticipated result from the IDS meant that the assumption of annoyance generated by the delayed tasks need to be re-examined. Further, there were also a number of important post-1999 developments in the literature, published

just after the completion of the IDS. This newer literature, which was reviewed in chapter 4, had highlighted some of the methodological inconsistencies in the earlier literature, as well as further developments on the factors that influence sequence evaluation. This review concluded that there is a wide range of factors, other than patterns, that could also influence sequence evaluation. Two specific points from this review are highlighted, as they formed the motivation and theoretical starting point for the studies that were developed in chapter 5 and 6.

7.2.2 Incorporating key themes from the post-1999 literature to refine the model of delay sequence evaluation developed

The first point relates to the finding that cognitive ability has a role to play in moderating pattern effects for some sequence domains. As argued in section 4.1.4 of chapter 4, patterns are a type of heuristic that people use to summarise their sequential experiences. The review in the aforementioned section found that the effects of pattern have more influence over sequence evaluation when people do not have sufficient ability and/or knowledge to evaluate a sequence based on normative principles, such as the utility integration principle. For a sequence of delays, this implied that the pattern of delays would only be a significant influence on evaluation if the patterns were used as a heuristic. For example, it was suggested that if the delays significantly obstructed a task, this could enhance the role of the delay pattern as a heuristic, because people have to switch to simpler strategies, such as evaluation based on patterns, to cope with the increased processing requirements of the task (see Gigerenzer and Todd, 1999; Ariely and Carmon, 2003). This suggestion was formalised into a hypothesis, and tested in the four studies reported in chapter 5.

The second point relates to the finding that for many sequence domains, preferences are driven by expectations, which are in turn derived from people's personal experiences with sequences. For example, people prefer a sequence where their health declines over their life span, as it is based on their observation of what usually happens (Chapman, 2000). In another example, Read and Powell (2002) showed that people prefer a sequence of evenly

spread monthly wage payments instead of receiving their wages in one lump sum annual payment, because this is the way that most organisations pay their wages. This point is important because when applied to the context of delays, it suggests that people may not always prefer a delay sequence with an improving pattern, if it is unrealistic for them to expect such a pattern. This implies that preferences for delay sequences are conditional on the domain in which the delay is experienced, because different domains would induce different expectations. These propositions were further developed into a model in the studies reported in chapter 6.

In addition to the two points discussed above, this review also presented a novel approach to analysing the methodologies used in the literature, in section 4.2. This analysis of methods represents one of the four key areas where this thesis had contributed to the literature. The contribution was to highlight key parameters and techniques used in the studies reported in the literature, such as considerations of statistical power in choosing an appropriate sample size, experimental methods adopted (e.g. laboratory or questionnaire based methods), and type of dependent variable chosen. The conclusions of this analysis were also used as a theoretical defence of the choice of parameters, independent and dependent variables, and experimental method chosen for the studies presented in this thesis.

To summarise, Model 1 was developed to predict the evaluation of internet delay sequences, but the prediction was not supported by the data. It was argued that a re-examination of some of the assumptions made in Model 1 was needed because of the significant development in the literature after the IDS was concluded. The next section describes the motivations and the key contributions of the studies reported in chapter 5.

7.2.3 Improving the model on the evaluation of delay sequences

The general objective of the studies in chapter 5 was to introduce a test of Model 1 that is different from the IDS, given that the results of the IDS were inconclusive. It was argued that the delays in the tasks of the IDS might not be obstructive enough to generate sufficient negative emotions to induce an

evaluation based on patterns, which lead to the development of an alternative method of manipulating task obstruction. In the first Memory Study, MS1, of chapter 5, the task manipulation of the delays was designed such that it prevented participants from fully achieving the memory tasks. In MS1, participants were required to commit objects to memory, but where delays made the tasks more difficult because it lengthened the time needed for the objects to remain in memory. This manipulation of the delayed task was unlike the one used in the IDS, where the tasks were always achievable regardless of the delay duration.

The results of MS1 was in the direction of the predictions made in Model 1, whereby the evaluation of a sequence of increasing delays was more aversive than a decreasing delay sequence. However, the results of MS1 were inconclusive, because the task loads in this study were confounded. It was only realised after the study was completed that the task load for the increasing delay sequence was higher compared to the decreasing delay sequence. This is because in the former sequence, more objects have to be kept in memory for a longer period than the latter sequence. This confounding meant that it was impossible to know whether the preference for the decreasing delay sequence was due to a pattern effect or was caused by the confounded task loads.

The objective of the second Memory Study or MS2 was to improve on the design of MS1 by correcting for the confounding by task load. This was achieved by changing the positioning of the memory tasks within the sequence such that it would not interfere with the delays. This has the effect of ensuring that for both delay sequences, task loads remain identical, because all objects have to be kept in memory for an equal amount of time. The results of MS2 provided support for the predictions made in Model 1, whereby a sequence of increasing delays was found to be more negative than a sequence based on decreasing delays.

In Model 1, as discussed earlier, it was assumed that the evaluation of a delay sequence based on its patterns was induced by the negative emotions

generated by the delays, because delays obstructed progression. In the previous MS2, however, it was shown that an evaluation based on patterns was induced when the delays made the tasks unachievable. This lead to the proposition that the negative emotions generated by the delays come from people being frustrated because they cannot fully achieve the tasks set, rather than from the perception that their progression had been obstructed, as first assumed in the IDS. The final two studies of chapter 5, the first and second Visual Search Studies or VS1 and VS2 respectively, were set up to test this proposition.

The purpose of VS1 was to develop a model to demonstrate that any manipulation that obstructs task achievement would induce an evaluation based on patterns. The model developed was based on modifications to Model 1. Let O=any factor that obstructs task achievement, and when one factor causes more obstruction to task achievement than another or O2>O1, then the lesser obstruction would be evaluated less aversively or V(O2)>V(O1). Extending this to Model 1 would give rise to the following inequality:

V(O1, O2, …, On)>V(On, …, O2, O1) … [Model 2]

Model 2 predicts that for a sequence of tasks, the increasing obstruction pattern is rated more aversely than a decreasing obstruction pattern. This model is a more general version of Model 1, because it implies that an evaluation based on patterns can be induced in a sequence of tasks using any manipulation that will obstruct a task, regardless of whether the obstruction is caused by delays or originating from other sources. Model 2 was tested by introducing an alternative manipulation of obstruction in the form of time constraints (TC) being imposed on the sequence of tasks. This manipulation was represented in Model 3, which predicts that a sequence of tasks that are increasingly time constrained is evaluated more aversively than a sequence of decreasing time constraints.

V(TC1, TC2, …, TCn)>V(TCn, …, TC2, TC1) … [Model 3]

The results of VS1 supported the predictions of Model 3, where a sequence of increasing time constraints that obstructed task achievement was rated more aversively than a decreasing time constraints sequence. At a more general level, the combined results from MS2 and VS1 also provide evidence to support Model 2, because these studies showed that patterns significantly influenced the evaluation of a sequence of tasks when task achievement was obstructed by the experimental manipulations (e.g. delays in MS2 and time constraints in VS1) introduced.

However, implicit in all three studies (MS1, MS2 and VS1) was the assumption that an evaluation based on patterns was only induced for a sequence of tasks when the tasks were unachievable. It cannot, however, be deduced from these studies whether such evaluations can still be induced if the tasks were simple. The final study of chapter 5, VS2, was therefore set up to determine the conditions required for evaluation based on patterns to be induced for a sequence of tasks. It was concluded from the results of VS2 that an evaluation based on patterns for a sequence of tasks would only be induced when the obstructions made the tasks unachievable.

To summarise, the broad objectives of the studies in chapter 5 was to further refine the methods used to test Model 1, which predicted how people would evaluate sequences of delays. Previous research on how gestalt theories could be used to model sequences of delays were limited to the documentation of end effects in a sequence of queues (i.e. Carmon and Kahneman, 1993), and have not developed more coherent models, presented in chapter 5, which links delays with obstruction and negative emotions, or acknowledged and addressed boundary issues relating to the application of gestalt theories to model the evaluation of delay sequences. The key finding was to demonstrate that axioms such as the preference for an improving sequence could be used to predict the evaluation of a sequence of delayed tasks, albeit only when certain conditions are met. It was shown that delays generate sufficient negative emotions to induce an evaluation based on patterns only when it prevented full task achievement.

The key contributions made up to this point to the sequence literature are as follows. The first was in the development of a model of how people evaluate sequences of delays (i.e., Model 1), and its subsequent refinements (Models 2 to 4). In the process of refining the model, it has been demonstrated that pattern effects are only present under certain conditions, thus providing further evidence, in the domain of delay sequences, to support the theories of moderation emphasised in the post-1999 literature review in chapter 4. A second contribution is in providing a more comprehensive analysis of methods used in the sequence literature, compared to the more recent attempt (i.e., Guyse et al., 2002, which only covers less than half of the studies reported in Table 4.1 of chapter 4).

In applying these findings to the context of sequences of internet delays, online retailers should note that an evaluation based on patterns would only be induced if the delays significantly obstructed consumers' ability to navigate a website successfully. Therefore, while internet retailers should be aware that the effects of delays on consumers can be minimised by manipulating the sequence order of the delays to create a decreasing delay pattern, it is, however, more important for retailers to ensure that delays are not excessive to the point that it significantly obstructs or prevents task achievement. This is because delays from unachievable tasks generate more negative emotions than delays from tasks that are achievable, at least within the parameters and context of the studies reported in chapter 5. This application represents another contribution that this thesis makes, which is provide insights into consumer behaviour towards retailing experiences where sequences of delays cannot be avoided.

So far, however, the chapter 5 studies had only focus on short delay sequences, defined as sequences with duration less than 3 minutes. While this duration may be a more realistic parameter for internet delay sequences, it is suggested that the duration for other delay domains is considerably longer than 3 minutes. For this reason, the studies in chapter 6 proposed that tests of Model 1 are extended to sequences with longer durations.

7.2.4 Extending the models developed to longer sequences of delay

The broad objective of the three studies in chapter 6 was to explore the factors that influence sequences of longer delays. A distinction was made between different delay durations because it was conjectured that the durations utilised in chapter 5 was unable to generate sufficiently large negative emotions to induce evaluation based on patterns. Instead, such evaluations were only induced when the delays prevented full task achievement, which lead to the proposition that there are two sources from which delays can generate negative emotions. The first is from slowing progression, and the second is from obstructing task achievement.

Therefore, the purpose of the first study of chapter 6, the Transport Delay Study or TDS, was to test the predictions of Model 1 for longer sequences of delay. It was assumed that the longer duration of 2 hours used would generate sufficient negative emotions to induce an evaluation based on patterns. The results of the study showed that people preferred the decreasing to the increasing delay sequence, which supported the predictions of Model 1. This result also supported the proposition that there are two sources of negative emotions that are generated from sequences of delays. Since the scenarios used in the TDS does not involve any unachievable tasks (e.g. preventing people from arriving at their destination rather than merely slowing their journey), it is therefore concluded that an evaluation based on patterns was induced because the negative emotions generated from the delays was large enough to be a material consideration in evaluation. In terms of the significance of the findings in the TDS for the literature, this study has shown that a Utility Integration Principle or UIP-derived model, Model 1, which was based on Kahneman's (1997) postulates, could also be applied to predict behaviour in sequences of longer delays. This provides further empirical evidence to support the models developed earlier in the thesis.

The purpose of the second study (Domain Study or DS) in chapter 6 was to explore the effects of domain on preferences for delay sequences. Domain here is defined as the environment or context in which a delay is experienced.

For example, a delay of 6 hours waiting at an airport due to snow storms is considered as a different domain from a delay of 6 hours when queuing to buy tickets for Wimbledon. It was conjectured that preferences for delay sequences were conditional on domain, based on evidence from the (non-sequence) delay literature reported in chapter 2. The manipulation of domain in this study was operationalised by creating a scenario where each participant experienced four different delayed events. To ensure comparability between the TDS and the Domain Study, an experimental set-up (apart from the scenario of the delayed events) that was based on the TDS was used. The results showed that people preferred the increasing delay sequence in the DS, which was opposite to the TDS result, where preference was for the decreasing delay sequence. This result demonstrated that domain had a significant influence over preferences for delay sequences. To provide an explanation for the observation that the delay domain influences preferences for delay sequences, a model based on expectations theory (Chapman, 2000) was developed. Expectations theory was used because the literature (e.g. Chapman, 2000; Read and Powell, 2002) had shown that people have different expectations for different (non-delay) sequence domains, and that preferences are influenced by expectations.

Model 4, which is described below, examined the relationship between preferences and expectations for delay sequences. Chapman (2000) explained that the mechanism in which expectations influence preferences for health sequences is through the reference point effect, whereby decision-makers compare sequence outcomes to a reference sequence that is based on expectations of how such outcomes usually unfold over time. Chapman's (2000) studies showed that decision-makers prefer sequences with expected patterns to sequences with unexpected patterns, because unexpected sequences generate net negative utilities. The theoretical concept discussed in Chapman (2000) was extended to delay sequences in Model 4, where it was predicted that preferences for delay sequences could be explained by the expectations for such sequences. Expectations were incorporated into Model 1 through the relative weighting of individual delays. The derivation of Model 4

is as follows. It was assumed that the magnitude of the individual weight (||wn||) is a function of the expectations, f(expect), of the duration for the event.

||wn|| = f(expect) … proposition 1

This means that the distribution of weights (w1, w2, …, wn) may take different forms, depending on the expectations for the individual events within the sequence. The evaluation of a sequence of delays, V(d1, d2, …, dn), therefore, is conditional on expectations.

V(d1, d2, …, dn) = w1.V(d1) + w2.V(d2) + … + wn.V(dn)

Combining proposition 1 with the weighted utility model above gives:

V(d1, d2, …, dn) = f(expect1).V(d1) + f(expect2).V(d2) + … + f(expectn).V(dn) … Model 4

Model 4 predicted that the evaluation of a sequence of delays is a function of the individual expectations for each event, in addition to the negative utility from the individual delays. The test of Model 4 was operationalised in the Expectations Study or ES, which is the final study of chapter 6. In the ES, participants were required to indicate their preferences and expectations for different sets of delay sequences, where the event and duration of delay was systematically varied. The results showed that people generally prefer the sequence patterns that they expect. In addition, the results also demonstrated that preferences for delay sequences varied with the different combinations of event and delay duration. This supported the assertion of Model 4 that the preference for a delay sequence is a function of the individual expectations for each delayed event. This finding implied that a preference for delay sequences that is made up of different domains is difficult to predict, because expectations for each of the domain can vary significantly.

To summarise, the broad objectives of the discussions in chapter 6 was to extend Model 1 to sequences of longer delays. This route of developing Model

1 was chosen because the results of the studies in chapter 5 only supported the prediction made in Model 1 when delays obstructed task achievement. This suggests that delays may not always generate sufficient negative emotions to induce an evaluation based on patterns. This finding, however, was in contrast with the findings of the (non-sequence) delay literature, on which Model 1 was developed. This literature showed that people are frustrated with delays because delays obstruct progression, even if full task achievement is possible. For example, for the first type of delay, people are frustrated from waiting for public transport, even if the public transport does eventually arrive, and take them safely to their destination. For the second type of delay, people are frustrated because they have missed their public transport connection as a result of the delays, which may lead to a cascading effect of missed connections. The arguments presented in chapter 5 and 6 distinguish between the first type of delays from the second type of delays. In the first type of delay, people become frustrated because of the length of the delays. In the second type of delay, people become frustrated when they are unable to achieve the task(s) set out (i.e. arrive at their destination) due to obstruction caused by the delays. It can also be implied from the results of these studies that the delay duration does not have to be long in order to evoke sufficient negative emotions to induce an evaluation based on patterns. All that is required is for the delays to be of sufficient duration to prevent full task achievement.

The key contribution of the three studies reported in chapter 6 is to further refine our understanding of the psychological mechanisms underpinning the evaluation of delay sequences. It is argued that the objectives in the studies of chapter 6 were chosen to increase the robustness of the theory developed in Model 1. Previous models of sequence evaluation in the literature were often only based on a narrow set of methods and parameters used.

The next section (7.3) discusses the limitations of the studies reported in this thesis, and suggests a number of ways in which these studies can be improved. It also provides a brief discussion on themes for future research. The thesis is then concluded in section 7.5.

7.3 Limitations and future

The objective of this section is to discuss the limitations of the research presented in this thesis and to suggest future studies that improve on these limitations. While it may not be possible to address all criticisms within the constraints of a thesis, it is argued that there are four aspects of the studies in this thesis that warrant further improvement.

The first is the range of patterns that were compared – all of the studies in this thesis have only compared the evaluations between two sequence patterns, an increasing trend versus a decreasing trend. These patterns were chosen because the primary objective of the studies was to demonstrate that, for aversive stimuli, patterns that end in improvement would receive less negative evaluations compared with other patterns. Given that improvement can be defined in a number of ways (i.e. higher velocity, trend, or peak-end), the sequence literature had explored the effects of other sequence patterns on evaluation, to determine the characteristics of pattern that induces the greatest contrast in evaluation. For example, as reported in the earlier literature reviews, Ariely (1998) found that the end of a sequence, and the trend towards the end (rather than the trend for the entire sequence), had the most influence over evaluation. Chapman (2000) and Read and Powell (2002) showed that while people often prefer a sequence with an improving trend, there are many occasions where other sequence patterns were preferred, either because they were the most realistic patterns given the scenario (e.g. declining sequences of long-term health), or they were more appropriate given the personal circumstances of the decision-maker (e.g. a constant sequence of monthly pay-cheques). Future studies on delay sequences should investigate how different patterns (e.g. constant or flat trend, constant trend with a spiked ending, increasing followed by decreasing trend etc.) affect the evaluation of delay sequences, because this can provide information to discriminate the patterns that have the most influence over evaluation. This can help answer questions such as whether the less aversive evaluation of a decreasing compared to an increasing delay sequence is caused exclusively by an end

effect, trend effect, or trend towards the end effect. Ariely and his colleagues (Ariely, 1998; Ariely and Zauberman, 2000; Ariely and Carmon, 2000) have attempted to address these questions for sequences of painful stimuli, and therefore it would be logical to extend these research objectives to delay sequences as well.

A second limitation of the research here is on issues of comparability, given the range of dependent variables used. For example, the memory studies in chapter 5 used frustration as the term to describe the negative emotions, while the visual search studies used annoyance. The appropriateness of the terms to describe the negative emotions generated by the stimuli in the studies was developed based on pilot studies, which is the reason for the differences in terminology used. While it could be argued that both of these terms are representative of negative emotions and therefore allows greater comparability between studies, there is perhaps also a need for further research to study the impact of the choice of dependent variables for delay sequences. Such a study can answer the question of whether people's evaluations change when the dependent variable used to capture evaluation is varied.

The third limitation is on the realism of the scenarios used in the prospective studies in chapter 6. It was previously argued that one advantage of using a prospective paradigm of testing hypotheses in sequence research is that it allows for the greater manipulation of variables and parameters. The disadvantage, however, is that questions can often be raised on the realism of the scenarios used. For example, while the events for the delays were selected on the assumption that it represented events where it was not uncommon to expect delays, it could be argued that the sequence of scenarios used might be hard to imagine, as the events are fairly diverse. It is proposed that an empirical approach is perhaps one of the best ways of addressing such criticisms. For example, the Expectations Study could be replicated using a retrospective evaluation paradigm, meaning that the actual delay scenarios are constructed for people to experience and evaluate afterwards. As argued in the previous chapter, this seems to be the approach taken in the literature,

where theories developed in prospective studies were shown to be replicable in retrospective studies.

A fourth limitation of the studies reported in this thesis is that they have only investigated delay sequences for two durations, one less than 3 minutes, and the other, for 2 hours, which places a constraint on the extent to which the results presented could be generalised to durations of other lengths. Future studies should systematically investigate the role of sequence duration on evaluation, so that the theories developed in the studies reported in this thesis can be extended to other durations. The discussions so far have focussed on the limitations of the present studies in this thesis and how these studies can be improved. The discussion in the next paragraph will focus on a theme for future research that has not yet been addressed in any of the literature.

One of the themes that emerged from the discussion of the Expectations Study was the need to explore preferences for delay sequences across different cultures. There is a need to better understand the development of expectations and the impact of culture on expectations. There is also need to know the extent to which preferences for delay sequences are driven by expectations across different cultures. The importance of culture is underlined here because the bulk of research on sequences are conducted in the environment of western cultures, which suggests that there is scope for further research to determine whether these theories readily extend to people from other cultures.

7.4 Conclusions

The original motivation of research carried out in this thesis was to predict the evaluation of internet delay sequences using a model developed from the sequence literature. This model was tested in the Internet Delay Study (IDS) of chapter 3. The results of that study, however, did not support the model, which lead to a re-examination of some of the basic concepts used in the model's development. The re-examination lead to the further refinement of the model used to predict evaluations of delay sequences. The main contributions

to the literature of the research presented in this thesis demonstrated that axioms such as the preference for an improving sequence of aversive stimuli could also be used to model preferences for delay sequences, provided that certain conditions were met. These conditions exist because delay sequences do not always evoke sufficient negative emotions to induce an evaluation based on patterns. This is because the negative emotions that delays evoke are conditional on the context in which it is experienced. The studies have shown that parameters such as the delay duration and the achievability of the delayed tasks, and factors such as expectations, are important influences on preferences for delay sequences. The next step forward in the research on delay sequences is to develop models that can explain the factors influencing the formation of expectations and preferences in delay sequences. This research will provide us with a better understanding of our own preferences for delay sequences.

References

Ariely D (1998), "Combining experiences over time: the effects of duration, intensity changes and on-line measurements on retrospective pain evaluations," Journal of Behavioral Decision Making 11, 19-45.

Ariely D (2001), "Seeing sets: Representation by statistical properties," Psychological Science 12(2), 157-162.

Ariely D and Carmon Z (2000), "Gestalt characteristics of experiences: the defining features of summarized events," Journal of Behavioral Decision Making 13(2), 191-201.

Ariely D and Carmon Z (2003), "Summary assessment of experiences: The whole is different from the sum of its parts," in chapter 11, Loewenstein G, Read D and Baumeister RF (eds.), Time and decision: Economic and psychological perspectives on intertemporal choice, 323-349, Russell Sage, NY.

Ariely D and Loewenstein G (2000), "When does duration matter in judgment and decision making?" Journal of Experimental Psychology: General 129(4), 508-523.

Ariely D and Zauberman G (2000), "On the making of an experience: The effects of breaking and combining experiences on their overall evaluation," Journal of Behavioral Decision Making 13(2), 219-232.

Ariely D and Zauberman G (2003), "Differential partitioning of extended experiences," Organizational Behavior and Human Decision Processes 91(2), 128-139.

Ariely D, Kahneman D and Loewenstein G (2000), "Joint comment on 'When does duration matter in judgment and decision making?'" Journal of Experimental Psychology: General 129(4), 524-529.

Babad E and Katz Y (1991), "Wishful thinking – against all odds," Journal of Applied Social Psychology 21, 1921-1938.

Baumgartner H, Sujan M and Padgett D (1997) "Patterns of affective reactions to advertisements: the integration of moment-to-moment responses into overall judgments," Journal of Marketing Research 34, 219-32.

Benson L and Beach LR (1996), "The effects of time constraints on the pre-choice screening decision options," Organizational Behavior and Human Decision Processes 67(2), 222-228.

Ben-Zur H and Breznitz SJ (1981), "The effects of time pressure on risky choice behavior," Acta Psychologica 47, 89-104.

Brunner JS and Minturn AL (1955), "Perceptual identification and perceptual organization," Journal of General Psychology, 53, 21-28.

Carmon Z and Kahneman D (1993) "The experienced utility of queuing: Real time affect and retrospective evaluations of simulated queues," Working paper, INSEAD, Paris.

Carmon Z, Shanthikumar JG and Carmon TF (1995) "A psychological perspective on service segmentation models: The significance of accounting for consumers' perceptions of waiting and service," Management Science 41(11), 1806-1815.

Chapman GB (1996a), "Expectations and preferences for sequences of health and money," Organizational Behavior and Human Decision Processes 67(1), 59-75.

Chapman GB (1996b), "Temporal discounting and utility for health and money," Journal of Experimental Psychology-Learning Memory and Cognition, 22(3), 771-791.

Chapman GB (2000), "Preferences for improving and declining sequences of health outcomes," Journal of Behavioral Decision Making 13(2), 203-217.

Cohen J (1969), Statistical power analysis for the behavioral sciences, Academic Press, London and New York.

Dawson NV and Arkes HR (1987), "Systematic errors in medical decision making: Judgment limitations," Journal of General Internal Medicine 2, 682-683.

Dellaert BGC and Kahn BE (1999) "How tolerable is delay? Consumers' evaluations of internet web sites after waiting," Journal of Interactive Marketing 13(1), 41-54

Denzin NK (ed.) (1970), Sociological methods: A sourcebook, Aldine, Chicago.

Diener E, Wirtz D and Oishi S (2001), "End effects of rated life quality: The James Dean effect," Psychological Science 12(2), 124-128.

Donovan RJ and Rossiter JR (1982), "Store atmosphere: An envrionmental psychology approach," Journal of Retailing 58, 34-57.

Dube L, Leclerc F and Schmitt BH (1991), "Consumers' duration estimate of delays at different phases of a service delivery process," paper presented at the 6th John-Labatt Marketing Research Seminar, University of Quebec, Montreal, Canada.

Elster J and Loewenstein G (1992), "Utility from memory and anticipation," in Loewenstein G and Elster J (1992), Choice over time, 213-234, Russell Sage, NY.

Fischer GW and Hawkins SA (1993), "Strategy compatibility, scale compatibility, and the prominence effect," Journal of Experimental Psychology: Human Perception and Performance 19, 252-266.

Fischer GW, Carmon Z, Ariely D and Zauberman G (1999), "Goal-based construction of preferences: Task goals and the prominence effect," Management Science 45(8), 1057-1075.

Folkman S (1984), "Personal control and stress and coping processes: A theoretical analysis," Journal of Personality and Social Psychology 46(4), 839-852.

Fredrickson BL (2000), "Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions," Cognition and Emotion, 14(4), 577-606.

Fredrickson BL and Kahneman D (1993), "Duration neglect in retrospective evaluations of affective episodes," Journal of Personality and Social Psychology 65, 45-55.

Gigerenzer G and Todd PM (1999), Simple heuristics that make us smart, OUP, NY.

Gilovich T, Griffin D and Kahneman D (eds.) (2002), Heuristics and biases: The psychology of intuitive judgement, Cambridge University Press, Cambridge.

Guyse JL, Keller LR and Eppel T (2002), "Valuing environmental outcomes: Preference for constant or improving sequences," Organizational Behavior and Human Decision Processes 87(2), 253-277.

Higbee KL and Wells MG (1972), "Some research trends in social psychology during the 1960s," American Psychologist, 27, 963-966.

Hockey GRJ (1997), "Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework," Biological Psychology 45, 73-93.

Hogarth RM (1987), Judgement and choice, 2nd edition, John Wiley, UK.

Holbrook MB and Gardner MP (1993), "An approach to investigating the emotional determinants of consumption durations: why do people consume what they consume for as long as they consume it?" Journal of Consumer Psychology 2(2), 123-142.

Hornik J (1984), "Subjective versus objective time measures: A note on the perception of time in consumer behavior," Journal of Consumer Research 11, 615-618.

Hsee CK and Abelson RP (1991), "Velocity relation: satisfaction as a function of the first derivative of outcome over time," Journal of Personality and Social Psychology 60(3), 341-347.

Hsee CK, Abelson RP and Salovey P (1991), "The relative weighting of position and velocity in satisfaction," Psychological Science 2, 263-6.

Hsee CK, Loewenstein G, Blount S and Bazerman MH, "Preference reversals between joint and separate evaluations of options: A review and theoretical analysis," Psychological Bulletin 125(5), 576-590.

Hsee CK, Salovey P and Abelson RP (1994), "The quasi-acceleration relation: Satisfaction as a function of the change of velocity of outcome over time," Journal of Experimental Social Psychology 30, 96-111.

Huber J, Lynch J, Corfman J, Holbrook MB, Lehmann D, Munier B, Schkade D and Simonson I (1997), "Thinking about values in prospect and retrospect: Maximizing experienced utility," Marketing Letters 8(3), 323-334.

Hui MK and Tse DK (1996), "What to tell consumers in waits of different lengths: An integrative model of service evaluation," Journal of Marketing 60(2), 81-90.

Hui MK and Zhou LX (1996), "How does waiting duration information influence customers' reactions to waiting for services?" Journal of Applied Social Psychology 26(19), 1702-1717.

Hui MK, Dube L and Chebat J-C (1997), "The impact of music on consumers' reactions to waiting for services," Journal of Retailing 73(1), 87-104.

Jones EE, Rock L,Shaver KG, Goethals GR and Ward LM (1968), "Pattern of performance and ability attribution: An unexpected primacy effect," Journal of Personality and Social Psychology 10(4), 317-340.

Jung J (1969), "Current practices and problems in the use of college students for psychological research," Canadian Psychologist, 10, 280-290.

Kahneman D (2003), “Maps of bounded rationality: Psychology for behavioral economics,” American Economic Review, 93(5), 1449-1475.
Kahneman D and Frederick S (2002), “Representativeness revisited: Attribute substitution in intuitive judgment,” in chapter 2, Gilovich T, Griffin D and Kahneman D (eds.), Heuristics and biases: The psychology of intuitive judgment, 19-81, Cambridge University Press, Cambridge.
Kahneman D and Miller DT (1986), “Norm theory: Comparing reality to its alternatives,” Psychological Review 93(2), 136-153.
Kahneman D and Ritov I (1994), “Determinants of stated willingness to pay for public goods: A study of the headline method,” Journal of Risk and Uncertainty 9, 5-38.
Kahneman D and Tversky A (1979), “Prospect theory: An analysis of decision under risk,” Econometrica 47(2), 263-92.
Kahneman D, Diener E and Schwartz N (eds.) (1999), Well-being: Foundations of hedonic psychology, Russell Sage, NY.
Kahneman D, Fredrickson BL, Schreiber CA and Redelmeier DA (1993), “When more pain is preferred to less: Adding a better end,” Psychological Science 4, 401-405.
Kahneman D, Knetsch JL and Thaler RH (1990), “Experimental tests of the endowment effect and the coase theorem,” Journal of Political Economy, 98, 25-48.
Kahneman D, Slovic P and Tversky A (eds.) (1982), Judgment under uncertainty: Heuristics and biases. Cambridge University Press, Cambridge.
Kahneman D, Wakker P and Sarin R (1997), “Back to Bentham? Explorations of experienced utility,” Quarterly Journal of Economics 112(2), 375-406.
Keller LR and Sarin RK (1995), “Fair processes for societal decision involving distributional inequalities,” Risk Analaysis 15, 49-59.
Kent G (1985), “Memory of dental pain,” Pain 21, 187-194.
Langer T, Sarin R and Weber M (2005), “The retrospective evaluation of payment sequences: Duration neglect and peak-and-end effects,” Journal of Economic Behavior and Organization 58, 157-175.
Larson, RC (1987), “Perspectives on queues: social justice and the psychology of queuing,” Operations Research 35, 895-904.
Lawson R (1965), Frustration, MacMillan, NY.
List JA (2003), “Does market experience eliminate market anomalies?” Quarterly Journal of Economics, 118, 41-71.
List JA (2004), “Neoclassical theory versus prospect theory: Evidence from the marketplace,” Econometrica 72(2), 615-625.
Liyanarachchi GA and Milne MJ (2005), “Comparing the investment decisions of accounting practitioners and studnets: an empirical study on the adequacy of student surrogates,” Accounting Forum 29(2), 121-135.
Loewenstein G (1987), “Anticipation and the valuation of delayed consumption,” Economic Journal 97 (386), 666-684.
Loewenstein G (1999a), “Because it’s there: The challenge of mountaineering … for utility theory,” Kyklos 52, 315-344.
Loewenstein G (1999b), “Experimental economics from the vantage-point of behavioural economics,” Economic Journal 109, 25-34.
Loewenstein G and Elster J (1992), Choice over time, Russell Sage, NY.
Loewenstein G and Prelec D (1991), “Negative time preference,” American Economic Review: Papers and Proceedings 82, 347-352.
Loewenstein G and Prelec D (1993), “Preferences for sequences of outcomes,” Psychological Review 100(1), 91-108.
Loewenstein G and Schkade D (1999), “Would it be nice? Predicting future feelings,” in Kahneman D, Diner E and Schwartz N (eds.), Well-being: Foundations of hedonic psychology, 85-105, Russell Sage, NY.
Loewenstein G and Sicherman N (1991), “Do workers prefer increasing wage profiles?” Journal of Labor Economics 9, 67-84.
Loewenstein G, Read D and Baumeister RF (eds.), Time and decision: Economic and psychological perspectives on intertemporal choice, Russell Sage, NY.
Maister DH (1985), “The psychology of waiting lines,” in Czepiel JA, Solomon MR and Surprenant CF (eds.), The Service Encounter: Managing Employee/Customer Interaction in Service Businesses, 113-123, Lexington Books, MA.
Matsumoto D, Peecher ME and Rich JS (2000), “Evaluations of outcome sequences,”

Organizational Behavior and Human Decision Processes 82(2), 331-352.
Maule AJ, Edland AC (1997), "The effects of time pressure on human judgment and decision making," in Ranyard R, Crozier WR and Svenson O (eds.), Decision making: Cognitive models and explanations, 189-204, Routledge, London.
Maule AJ, Hockey GRJ and Bdzola L (2000), "Effects of time pressure on decision making under uncertainty: Changes inaffectivestate and information processing strategy," Acta Psychologica 104(3), 283-301.
Mehrabian A and Russell JA (1974), An approach to environmental psychology, MIT Press, Cambridge, MA.
Menon S and Kahn B (2002), "Cross-category effects of induced arousal and pleasure on the Internet shopping experience," Journal of Retailing 78(1), 31-40.
Novemsky N and Ratner RK (2003), "The time course and impact of consumers' erroneous beliefs about hedonic contrast effects," Journal of Consumer Research 29, 507-516.
O'Donoghue T and Rabin M (2000), "The economics of immediate gratification," Journal of Behavioral Decision Making 13, 233-250.
Osuna EE (1985), "The psychological cost of waiting," Journal of Mathematical Psychology 29(1), 82-105.
Payne JW, Bettman JR and Johnson EJ (1988), "Adaptive strategy selection in decision making," Journal of Experimental Psychology: Learning, Memory, and Cognition 14, 534-552.
Payne JW, Bettman JR and Johnson EJ (1993), The adaptive decision maker, Cambridge University Press, Cambridge.
Poynter WD and Holma D (1985), "Duration judgement and the experience of change," Perception and Psychophysics 33, 548-560.
Rachman S and Eyrl K (1989), "Predicting and remembering recurrent pain," Behaviour Research and Therapy 27, 621-635.
Raghunathan R and Irwin JR (2001), "Walking the hedonic product treadmill: Default contrast and mood-based assimilation in judgements of predicted happiness with a target product," Journal of Consumer Research 28(3), 355-368.
Rastegar H and Landy FJ (1993), "The interactions among time urgency, uncertainty and time pressure," in Svenson O and Maule AJ (eds.), Time pressure and stress in human judgment and decision making, 217-239, Plenum, NY.
Read D and Powell M (2002), "Reasons for sequence preferences," Journal of Behavioral Decision Making 15(5), 433-460.
Redelmeier DA and Kahneman D (1996), "Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures," Pain 66, 3-8.
Rosenthal R and Rosnow RL (1991), Essentials of behavioral research: Methods and data analysis, 2nd edition, McGraw-Hill, NY.
Ross WT and Simonson I (1991), "Evaluations of pairs of experiences: a preference for happy endings," Journal of Behavioral Decision Making 4, 273-82.
Sawrey JM and Telford CW (1971), Psychology of Adjustment, 3rd edition, Allyn and Bacon Inc., Boston.
Schkade D and Kahneman D (1998), "Does living in California make people happy? A focusing illusion in judgments of life satisfaction," Psychological Science 9(5), 340-346.
Schmitt DR and Kemper TD (1996), "Preference for different sequences of increasing or decreasing rewards," Organizational Behavior and Human Decision Processes 66(1), 89-101.
Schreiber CA and Kahneman D (2000), "Determinants of the remembered utility of aversive sounds," Journal of Experimental Psychology: General, 129(1), 27-42.
Sedlmeier P and Gigerenzer G (1989), "Do studies of statistical power have an effect on the power of studies?" Psychological Bulletin, 105, 309-316.
Simon HA (1959), "Theories of decision making in economics and behavioral science," American Economic Review 49, 253-283.
Simon HA (1981), The Sciences of the Artificial, 2nd edition, MIT press, MA.
Soman D (2003), "Prospective and retrospective evaluations of experiences: How you evaluate an experience depends on when you evaluate it," Journal of Behavioral Decision Making 16(1), 35-52.
Soman D and Shi M (2001), "Modelling and measuring the judgments of services over time: Virtual progress and choice," in Heil O and Schunk H (eds.), Proceedings of the

Marketing Science 2001 Conference, 140-141, Johannes Gutenberg Press, University of Mainz, Mainz, Germany.
Spiegel D (1998), “Getting there is half the fun: Relating happiness to health,” Psychological Inquiry 9, 66-68.
Stroop JR (1935), “Studies of interference in serial verbal reactions,” Journal of Experimental Psychology 18, 643-662.
Svenson O and Edland A (1987), “Changes of preference under time pressure: Choices and judgments,” Scandinavian Journal of Psychology 28, 322-330.
Svenson O and Maule AJ (eds.), Time pressure and stress in human judgment and decision making, 217-239, Plenum, NY.
Taylor S (1994), “Waiting for service: The relationship between delays and evaluations of service,” Journal of Marketing 58(2), 56-69.
Thaler RH (1980), “Towards a positive theory of consumer choice,” Journal of Economic Behavior and Organization 1(1), 39-60.
Thomas DL and Diener E (1990), “Memory accuracy in the recall of emotions,” Journal of Personality and Social Psychology 59, 291-297.
Treisman A (1982), “Perceptual grouping and attention in visual search for features and for objects,” Journal of Experimental Psychology: Human Perception and Performance 8(2), 194-214.
Tversky A and Kahneman D (1981), “The framing of decisions and the psychology of choice,” Science 211, 453-458.
Tversky A and Kahneman D (1986), “Rational choice and the framing of decisions,” Journal of Business 59(4), S251-78.
Tversky A, Sattath S and Slovic P (1988), “Contingent weighting in judgment and choice,” Psychological Review 95, 371-384.
Varey C and Kahneman D (1992), “Experiences extended across time: evaluation of moments and episodes,” Journal of Behavioral Decision Making 5, 169-85.
Wallsten TS and Barton C (1982), “Processing probabilistic multidimensional information for decisions,” Journal of Experimental Psychology: Learning, Memory, and Cognition 8, 361-384.
Wright PL (1974), “The harrased decision maker: Time pressures, distractions, and the use of evidence,” Journal of Applied Psychology 59, 555-561.
Zauberman G and Ariely D (2004), “Hedonic and informational evaluations of experiences that extend over time: the moderating role of evaluation type on sequential order effects,” Working paper, Kenan-Flager Business School, University of North Carolina at Chapel Hill.

Appendix

Chapter 4:

Appendix 4.1: Experimental design of the studies cited in Table 2.1 and 4.1

Study	Design
1	Study 1: 202 Q.; Study 2: 135 Q (x2); Study 3: 40 E (x2)
2	Study 1a: 40 Q; Study 1b: 40 Q
3	Study 1a: 24 Q; Study 1b: 24 Q; Study 2: 36 E (x12);
4	Study 1: 96 Q
5	Study 1: 48 Q; Study 2: 46 E
6	Example (E.g.) 1: 95 Q; E.g. 2: 48 Q; E.g. 3: 27 Q; E.g. 4: 51 Q; E.g. 5: 101 Q; Study 1: 52 E; Study 2: 57 E
7	Study 1: 32 E (x2); Study 2: 96 E (x2)
8	Study 1: 32 E
9	Study 1: 62 E (x2); Study 2: 163 E (x4)
10	Study 1: 20 E (x2); Study 2: 87 Q (42 v 45 split)
11	Study 1: 53 RL (x2)
12	Study 1, 2 & 3: 40, 50, 79 Q
13	Study 1: 28 E; Study 2: 34 E (x2); Study 3: 94 E (x4)
14	Study 1: 682 RL (x2)
15	Study 1: 20 E (x2), Study 2: 40 E (x2)
16	Study 1: 45 Q; Study 2: 376 Q (x4); Study 3: 50 Q
17	Study 1 to 4: 45 E per study
18	Study 1: 37 RL
19	Study 1: 54 E (x2); Study 2: 120 E (x4)
20	Study 1a & 1b: 20 & 24 E (x4); Study 2, 3, 4: 37, 36, 28
21	Study 1: 100 Q (x2); Study 2: 90 Q (x2); Study 3: 115 Q (x8); Study 4: 32 comparison of Study 1 to 3
22	Study 1: 115 Q (x4); Study 2: 55 Q (x4); Study 3: 34 Q (x2)
23	Study 1: 134 E (x8); Study 2: 225 E (x8); Study 3: 220 E (x8)
24	Study 1: 48 E (x4)
25	Study 1: 40 E and 25 E
26	Study 1: 136 E (x6); Study 2: 77 E (x4); Study 3: 91 E (x8); Study 4: 182 Q (x4)
27	Study 1: 45 E (x3); Study 2: 66 E (x3); Study 3: 60 E (x3); Study 4: 82 E (x2)
28	Study 1: 40 E (x2); Study 2: 108 E (x4); Study 3: 45 E (x3)
29	Study 1: 36 E; Study 2: 168 E

How to read this table:
The numbers on the left hand column corresponds with the studies cited in Table 2.1 and 4.1. For example, in the format: Study 1: 96 E (x4) denotes:-
The first number denotes the number of participants. The subsequent alphabet denotes the type of study, where Q = questionnaire-based study; E = laboratory-based study; RL = study based on real-life experiences. The final number in the brackets that proceeds with an 'x' indicate the number of between-participants factors. To obtain the number of participants employed for each between-participant cell condition, divide the total number of participants by the number of between-participants factors If there are no brackets, it means that the entire study was a within-participants design.

Chapter 5:

Appendix 5.1: Instructions for the memory (I) study

Your objective is to remember as many objects as possible that will appear in a sequence of 6 different pages.

These objects are in the form of colours and letters. You have to remember both the letter and its colour.

You will only be given a limited time to remember the objects before the page changes.

At the end of the sequence, you will be asked to recall the objects remembered. The answers will be in the 'multiple choice format' in your answer booklet.

Click here to START when you are ready.

Appendix 5.2: Instructions for the memory (II) study

This experiment has two consecutive tasks, A and B. B will appear immediately after A is completed.

Task A
In this task, you are required to memorise a number of coloured letters, and there is a time limit for this. You have to remember both the letter and the colour, but not the order of appearance or where it is located on the screen. For example, blue H, red X, green W, etc.

Task B
In this task, there will be a sequence of 6 pages that will appear consecutively. Each page will contain the name of a colour. In addition, the letters will also be coloured.

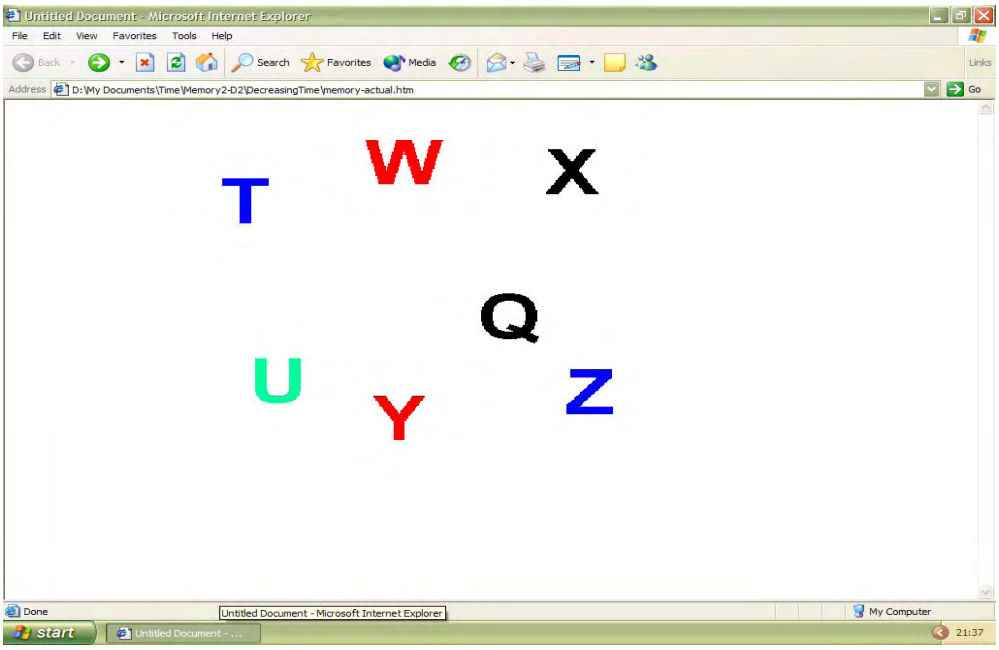
- Sometimes the colour spelled out will match the actual colours: e.g. BLUE (colour blue), RED (colour red), GREEN (colour green).
- At other times however the colour spelled will not match the actual colour: e.g. BLUE (colour red), RED (colour green), GREEN (colour blue).

You have to count the total number of matching name and colour, but ignore the colours that do not match the names: For example, the series BLUE (colour blue), BLUE (colour red), GREEN (colour blue), RED (colour green) has only one (1) correct match.

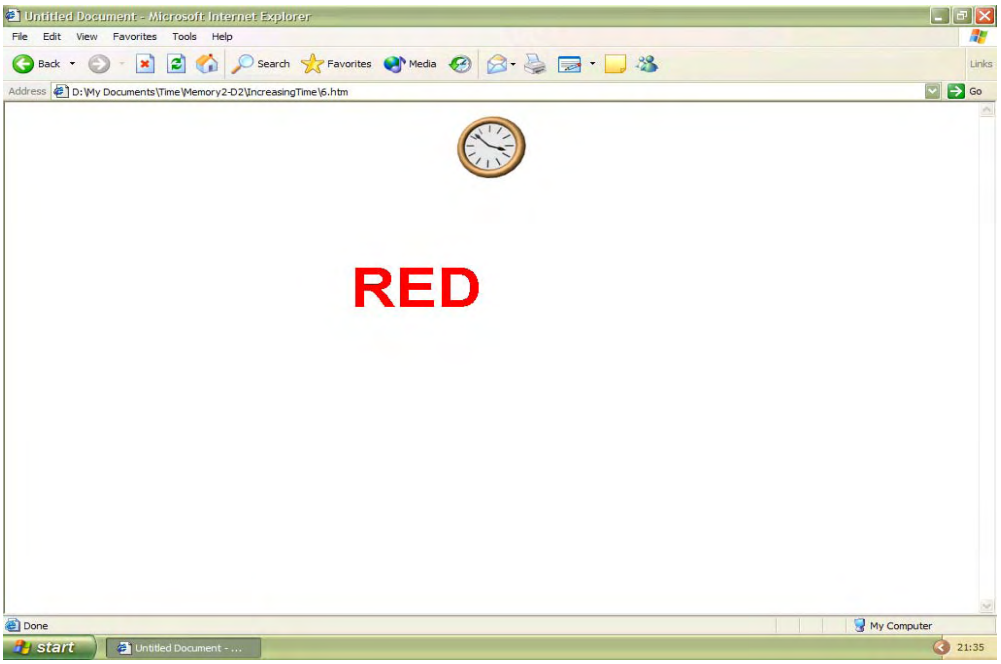
Each page will only appear for a few moments (e.g. 12 seconds) before the next page automatically replaces it. There will be a clock ticking on the screen to give you an indication of time passing. You are however not required to count time.

What you have to do when task A appears immediately after clicking 'start', try to memorise as many letters and colours before the time runs out. When the time runs out, task B starts, where you have to determine whether the letter and colour matches or not in the sequence of 6 pages. Task A and B will not be explicitly indicated. Instead, it will become obvious which is task A and B once you have completed the practice session a few times. Once you click start, the web pages will change automatically.

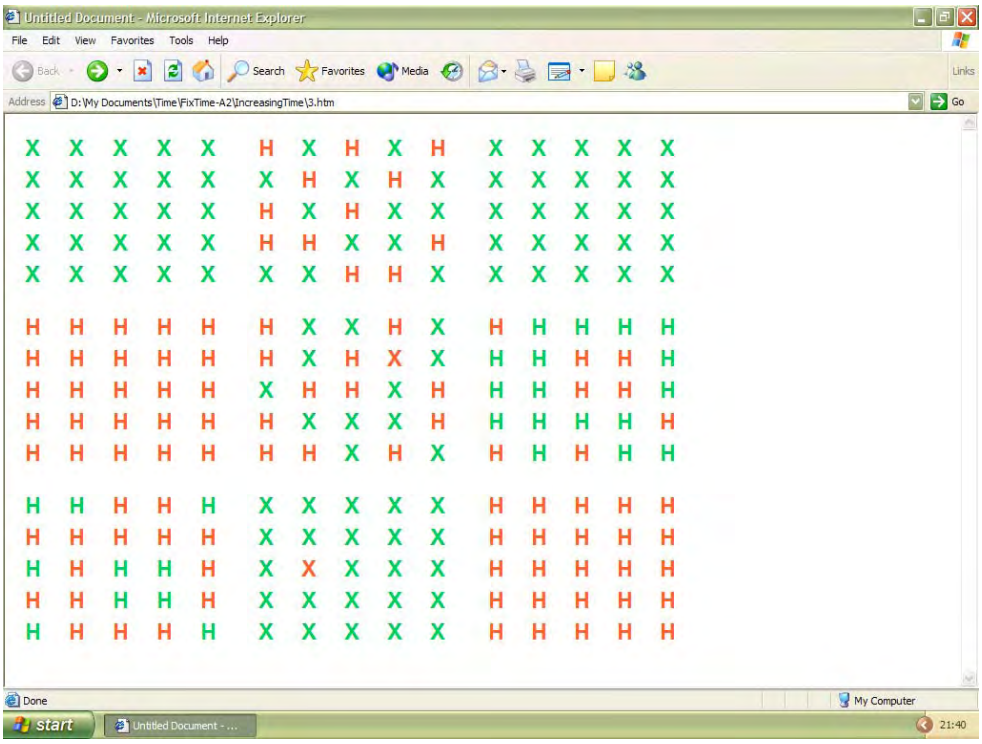
Appendix 5.3a: Computer screen illustration of the memory task in memory study (II)



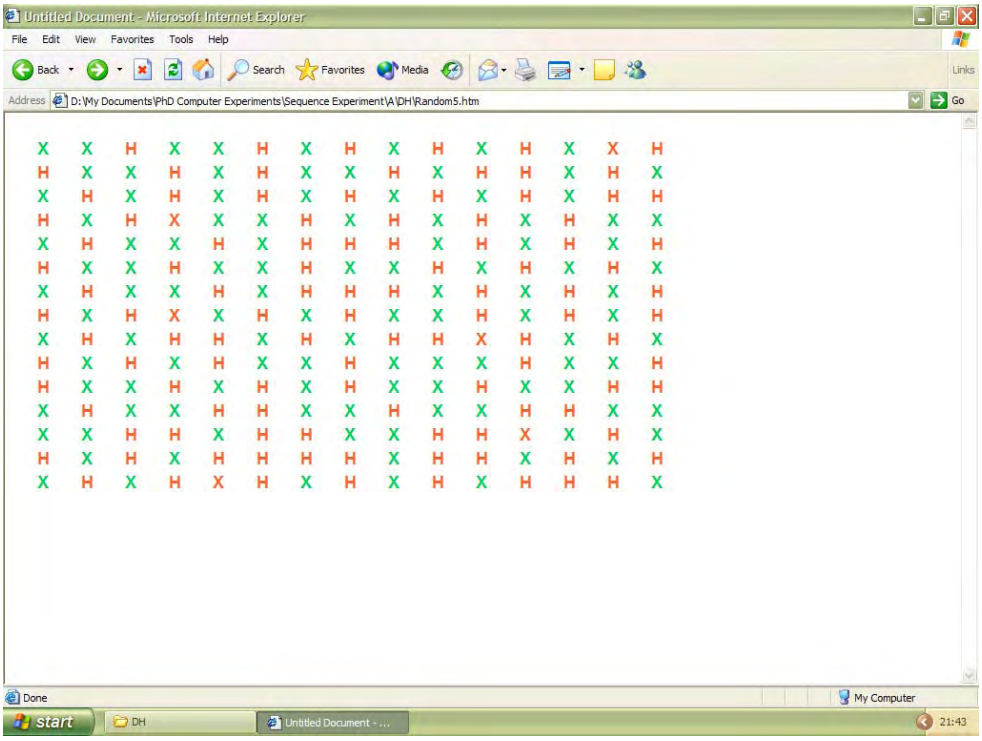
Appendix 5.3b: Computer screen illustration of the stroop-matching task in memory study (II)



Appendix 5.4: Computer screen illustration of one of the five tasks in visual search (I)



Computer screen illustration of one of the six tasks in visual search (II)



Appendix 5.5: Instructions for the visual search (I) and (II) experiment

Task:

Your task is to estimate the total number of the entire red X on each page.

For example, if you think the first page has two red X, the second page has three red X,, and the third page has five red X , then the estimated total number of red X is ten (2 + 3 + 5 = 10).

There is a time limit on each page, so try to complete the task as quickly as you can.

Write your estimate of the total number at the very end of the experiment.

Click here to start

Chapter 6:

Appendix 6.1: The questionnaires used for the Transport Delay Study (Results are included in brackets for the readers' convenience)

Both 'original' and 'replication' questionnaires:

Imagine that you have to get from A to B by public transport. You have to change transport four times while going from A to B. At each of the four changes, you experience time delays because you have to wait for the transport to arrive. This means that your travel time has increased.

Scenario X:
In this scenario, you experience a sequence of **decreasing** delays. There is a **60** minute delay for the first change, a **40** minute delay for the second, a **15** minute delay for the third and a **5** minute delay for the fourth. This means that the total delays you experienced is 120 minutes.

Scenario Y:
In this scenario, you experience a sequence of **increasing** delays. There is a **5** minute delay for the first change, a **15** minute delay for the second, a **40** minute delay for the third and a **60** minute delay for the fourth. This means that the total delays you experienced is 120 minutes.

'Original' questionnaire only

Question:
When you finally arrive at your destination B, which scenario would make you feel **more** frustrated? **Circle** one answer. [N=50]

Scenario X [14.0%] Scenario Y [78.0%] Both about the same [8.0%]

'Replication' questionnaire only

Question:
If you had to choose one of the above scenarios, which would you choose? **Circle** one answer. [N=60]

Scenario X [78.3%] Scenario Y [21.7%]

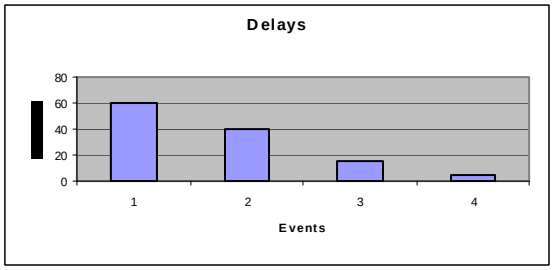
Appendix 6.2: Questionnaire used for the Domain Study

Imagine that you are now about to leave a hotel to catch a train. Before you can do this, however, you have to do the following:

Check out of the hotel
Call and wait for a taxi
The taxi arrives and takes you to the train station
Queue at the train station to buy a ticket

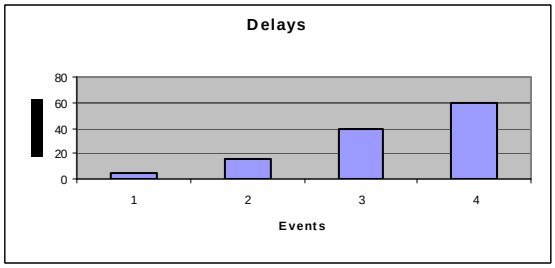
Scenario X:
In this scenario, you are rushing for the train, and you face delays. You just managed to board the train before it departs (i.e. you nearly missed the train). The total delay that you had experienced is 2 hours (120 minutes). The time delays for each event are as follows:

Event:	Delays:
1. Check out of the hotel	60 minutes
2. Call and wait for a taxi	40 minutes
3. The taxi arrives and takes you to the train station	15 minutes
4. Queue at the train station to buy a ticket	5 minutes
Total delays	120 minutes



Scenario Y:
In this scenario, you are rushing for the train, and you face delays. You just managed to board the train before it departs (i.e. you nearly missed the train). The total delay that you had experienced is 2 hours (120 minutes). The time delays for each event are as follows:

Event:	Delays:
1. Check out of the hotel	5 minutes
2. Call and wait for a taxi	15 minutes
3. The taxi arrives and takes you to the train station	40 minutes
4. Queue at the train station to buy a ticket	60 minutes
Total delays	120 minutes



Which scenario would make you feel more **FRUSTRATED**? **Circle** one answer. [N=69]

Scenario X [56.5%] Scenario Y [29.0%] Both about the same [14.5%]

Appendix 6.3: Stimuli used in the Expectations Study

Instructions:
There are four questionnaires in this booklet. Please answer all four of them, by circling your response. There is no right/wrong or better/worse answer for all of the four questionnaires.

Imagine yourself in a situation where you have to face four consecutive delays, which occur in the sequence order shown below. The delay duration is shown in the right column.

Sequence X:

'Old' events	'New' events	Duration (minutes)
Check out of a hotel	Queue at the post office	5
Wait for a taxi	Queue to try on clothes at a fashion store	15
Travel to a station	Queue in a waiting room for major dental treatment	40
Queue for a ticket	Queue to buy a ticket for a concert of your favourite musicians	60

Sequence Y:

'Old' events	'New' events	Duration (minutes)
Check out of a hotel	Queue at the post office	60
Wait for a taxi	Queue to try on clothes at a fashion store	40
Travel to a station	Queue in a waiting room for major dental treatment	15
Queue for a ticket	Queue to buy a ticket for a concert of your favourite musicians	5

Dependent variable 'preference':
Which sequence would make you feel more FRUSTRATED? Circle one answer.

Sequence X Sequence Y

Dependent variable 'expectations': Which sequence is more realistic? Circle one answer.

1	2	3	4	5	6	7
X is much more realistic			X and Y equally realistic (or unrealistic)			Y is much more realistic

-- End --