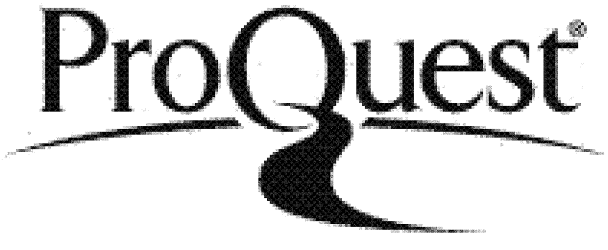
Strategies for time allocation across multiple on-line texts

Thesis submitted to Cardiff University
for the degree of
Doctor of Philosophy
by
Susan Claire Wilkinson

School of Psychology
Cardiff University
August 2007



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

TABLE OF CONTENTS

List of Figures

List of tables

List of Appendices

Acknowledgements

I would like to thank my supervisors, Steve Payne, Pat Wright, and Rob Honey, for their help and guidance throughout the course of this research.

I would also like to say a special thank you to all my family and friends for their endless love and support, especially to my Mum for her continuous emotional support throughout my PhD.

This thesis is dedicated to my family; I love you all dearly.

ABSTRACT

With growing information sources on-line, the issue is no longer the availability of information, but how people manage the task of allocating their time effectively to the best information sources for their needs. This thesis investigates the strategies people use to allocate time across multiple written texts when under time pressure.

Experiment 1 examined the effect that study time and number of resources had on the time allocation behaviour of participants presented with the task of learning for an unseen test. The browsing behaviour was recorded and indicated that people were using the satisficing strategy (where judgement and learning are integrated, as opposed to a sampling strategy, where judgement and learning are separated).

Experiments 2 and 3 investigated how adaptive participants were at allocating their time to appropriate information sources by asking them to learn for a specific test and providing them with four texts which varied in relevance to the test questions. Data showed that participants used the satisficing strategy to allocate more time to relevant texts.

Assuming that people therefore use the satisficing style of browsing behaviour when allocating time across texts under these conditions, Experiments 4 and 5 used eye-tracking methods to determine *how* the satisficing strategy is played out, and what level of text is generally used to make judgement decisions (i.e., *how* do people integrate judgement and learning and *when* do they make their judgement decisions). The data suggested that readers were using small units of text (e.g., paragraphs, or quarter sections of a page) to make their judgements, and that each time they rejected a unit of text, they would jump to the next unit. The skimming nature of their behaviour led to the development from the simple satisficing model to the satisficing then skimming model. The data regarding time spent on sections of pages also necessitated the development of the model into one which involved two rather than one level of text unit.

Experiment 6 investigated whether this satisficing strategy was restricted to the experimental conditions used in Experiments 1 to 5, or because it is a widely used strategy. Participants were either aware (informed condition) or unaware (un-informed condition) of the different qualities of the available texts. Those who were aware sampled more, but half still satisficed, thus suggesting that satisficing was not simply a product of the experimental conditions used in previous experiments.

This thesis adds to research in the area of time allocation, and proposes that people use a strategy based on satisficing in order to adaptively allocate their time across information sources. This conclusion has implications for the design of on-line texts, as knowing how people allocate time across a text can inform about how a text can be designed in order to support this kind of time allocation strategy.

CHAPTER 1

1 General Introduction

The world consists of information societies, that is, societies where modern day living centres around the creation, processing, communication, use, and evaluation of information (Rice, McCreadie & Chang, 2001). Information seeking behaviour is a topic which is much debated and researched (e.g., Pirolli & Card, 1999; Sandstrom 1994; O'Connor, 1993), and efforts indicate a shift in emphasis onto the user in order to inform the design of information systems.

This thesis addresses a widespread problem concerning the increase in availability of online resources, that is, the problem of how people allocate their time across these many information sources. In reality, people are under at least some degree of time constraint when gathering information they require from online sources, and they are forced to navigate their way through this information in order to allocate more time on the sources most appropriate to their information needs. Therefore, considering the number of information sources in relation to the amount of time people are likely to have in which to search, the need to allocate time adaptively becomes of paramount importance. Indeed, it has been a social aim for some time to improve people's ability to access and make sense of available information and this aim has also been implicated in efforts to improve modern day productivity (see Pirolli & Card, 1999).

The issue of time allocation to information sources has been argued by Pirolli and Card (1999) to be analogous to the problem of how animals forage for food. Other research has shown that time allocation across tasks varies with factors such as difficulty and time (Son & Metcalfe, 2000; Theide & Dunlosky, 1999). The experiments in this thesis present the readers with texts that are all *potentially* relevant to the task. This situation is similar to what a person might encounter when conducting a search using an internet search engine, where all the results are likely to be potentially relevant, which it can be argued reduces the effect that some of the factors which influence text choice may have on the reader (e.g., the interest the reader has in the topic).

The research in this thesis will attempt to shed light on the problem of allocating limited time to the most appropriate information source when people are faced with a potential overload of information. Thus, experiments in this thesis will provide an analysis of the strategies which readers use when faced with a situation in which they have more texts to consult than the time that they have available. This chapter will also present the research questions which the thesis aims to address, and how answering these questions will advance research about how people choose what information source to exploit, and how they go about making this choice.

This chapter will examine what is already known about how readers allocate their time across multiple texts under time pressure when studying for a test, and discuss relevant research areas, highlighting how they have shaped the current research. In terms of adjacent research in the literature, this chapter will firstly include a brief discussion about browsing, and will then go on to address the topic of information foraging, which will include examining optimal foraging theory and rules of thumb, specifically examining the work of Pirolli and Card, and Herbert Simon. There will then be a section to discuss the area of metacognition, specifically study time allocation. After relating these three areas of research to the research in this thesis, there will be a discussion about previous research carried out in the area of time allocation across multiple texts (Reader & Payne, 2007) and finally a description of different time allocation strategies.

1.1 Applied Browsing Research

There has been some research into the importance of accessing and browsing information in general, and this research is reviewed by Rice, McCreadie and Chang (2001). They claim that the traditional approach to studying information seeking as an intentional process, which involves a user with a known problem using an information resource to find specific answers that would solve their problem, is somewhat narrow and fails to consider the interdependency of tasks and the issue of tasks having blurred boundaries.

'Browsing, in essence, is an examination of unknown items of potential interest by scanning or moving through an information space in order to judge the utility of the items, to learn about something of interest in the item, or to satisfy curiosity about something' (Rice, McCreadie, & Chang, 2001, p. 302). Since much of

the research in this thesis centres round the way in which people adaptively allocate time across multiple texts, it is appropriate to briefly review the research carried out in the area of applied browsing.

Rice et al. (2001) propose that in order to conceptualise information seeking behaviour it must be seen as a broad and pervasive aspect of human behaviour, and therefore that a user-centered approach is necessary for it to be fully understood. It is argued that faced with information overload and limited time it is not easy for seekers to achieve optimality (see Katzer, 1987; Roberts, 1982). This idea of bounded rationality corresponds to Simon's (1976) satisficing theory in that people will search for something that is satisfactory rather than optimal (as discussed in Section 1.1).

Rice et al. (2001) claim that browsing is more than the mere retrieval of information, more than just the 'foraging' for information, that it is still a process that involves scanning through an information space in order to find required information. In this sense, it is similar to what Pirolli and Card (1999) describe as the process of foraging within an information patch for required information. So despite the different focus points of foraging theory and the more general concept of browsing behaviour research, these two approaches are linked by this major thread.

It has also been argued that the strategies employed when seeking information can be either analytical or browsing (Marchionini, 1995). Analytical strategies are characterised by careful planning, whereas browsing strategies are heuristic and opportunistic, and they depend on the recognition of relevant information. Based on these distinctions, the strategies identified in this thesis for the way in which readers allocate their time across texts when learning will be classed as browsing strategies, since their seeking will not be based on plans, but on the immediate information gain supplied by each information source. The time allocation behaviour of interest in this thesis is defined as a directed and purposeful activity, as there is conscious activity on the part of the reader to move towards a pre-defined goal.

Rice et al. (2001) suggest that browsing is one form of information seeking, and that it is seen as a way of dealing with information overload. However, there are many different literatures in which browsing has been examined (e.g., library studies, consumer research, media studies, information science, etc.), and it is considered differently in all of them. The term browsing has been used by different researchers without clear definition, and often with specific but different meanings assigned to it (Rice et al., 2001); e.g., it is construed as a search strategy in library studies, a

screening technique in organisational literature, and an entertaining activity in consumer research. It is this inability to agree on a common definition for browsing that leads to the focus of this thesis being based on aspects with more theoretical basis that are clearly defined and better understood. That is, this thesis will use theoretical aspects of foraging theory and study time allocation research to help identify the strategies being used by people faced with information overload, limited time, and a specific learning goal.

The important question that remains is not what constitutes browsing (cf. accessability, interactivity; Rice et al., 2001), but how readers go about exercising this behaviour in order to find the information they desire. It would be beneficial to step away from the descriptive analysis of *what* browsing is, and instead focus on the strategies people use to perform this browsing behaviour, thus ultimately achieving their goal. This thesis integrates *how* users find the information they require by looking at the skimming processes they employ when allocating time across multiple texts. Tracking eye movements is one way of addressing this issue that will be employed in this thesis.

Eye-tracking research has helped us to gain an insight into the mechanics of information seeking behaviour, which in turn could inform user interface and website designers. It has also deepened our understanding of the reading process, and the different kinds of reading activity which people practice (Rayner, 1998). Eye-tracking is relevant to the research in this thesis to the extent that eye-tracking techniques are used to identify at what level of text (e.g., a paragraph, a sentence, etc.) readers are making the stay/leave decision during their reading (Experiments 4 and 5). It therefore helps to identify *how* readers are selecting sections of text to read or reject. This is important in determining how readers go about using the time allocation strategy they do in order to choose, and therefore allocate most of their time to the text(s) which are most useful to them. Eye tracking helps to illustrate *how* people adaptively allocate time across multiple texts at a micro level; in contrast, broader measures (such as time on pages) indicate *which* strategies they use.

1.2 Information Foraging

Miller (1983) described humans as 'informavores', based on the observation that we actively seek, compile, digest and share information to a spectacular degree.

As humans surrounded with information richness, we are constantly forced to choose between multiple information sources according to our goals and aims. To deal with this overload, we have developed sophisticated strategies that enable us to gather information, interpret this information, and use it to make decisions or solve problems (Dennett, 1991). Pirolli and Card (1999, p. 642) view these information seeking strategies from an evolutionary ecological perspective. Their theory of information foraging is based on the fact that, 'when feasible, natural information systems evolve toward stable states that maximize gains of valuable information per unit cost'. People will adapt (e.g., their strategies) in order to maximise the gain they are able to achieve from a certain resource. One example of a study investigating adaptive reading is by Pirolli and Card (1999) where they observed the behaviour of MBA students who were required to collect materials for an assignment. To forage for the information they needed, the students first used an electronic system to search for citations. The 300 citations that were returned as a result of the search were scanned by the students while still on the screen, and 51 of these were selected and printed out, over a span of 12 minutes. The students then categorised these citations into important topics, which left them with 27 articles to read. Students then searched through these articles further rejecting 9 which left them with 18 information sources. Pirolli and Card (1999) argue that this illustrates how the students acted to maximise information gain per unit time by using a set of filters to aid their future foraging within patches of information. These students showed signs of using schematic representations to judge the utility or relevance of the information sources.

In the real world, where maximising is not practical due to time and resource constraints, rational choice may lead people to satisfice (stop searching after meeting some minimal requirement or when they have found something which is 'good enough') when making decisions rather than continuing to search for the optimal result. The term 'satisficing' is derived from the words 'suffice' and 'satisfy', and refers to a theory of choice which focuses on the process by which alternatives are examined and assessed. In this way, satisficing contrasts with optimization theory. This contrast is explained by Simon, who explains optimizing as being 'looking for the sharpest needle in the haystack' and satisficing as being 'looking for a needle sharp enough to sew with' (as cited in Winter, 2000). The concept of satisficing was introduced by Simon (1956) as a decision that meets some minimum requirement or that is 'good enough', even though it may not be 'optimal'. Simon describes a

satisficing search as a search for an optimal solution that terminates when a solution is found such that the cost of further searching exceeds the expected benefits of doing so (1956). Goodrich, Stirling and Frost (1998) point out that a satisficing solution is different to a suboptimal solution in the sense that the latter presupposes that an optimal solution must exist and is usually obtained by simplifying the original problem to gain a solution. Satisficing is still possible when a best solution either does not exist or is not attainable, since the notion of a minimum standard is still likely to apply. Goodrich et al. (1998) also point out that because satisficing must meet a minimum requirement this makes it different to ad hoc solutions which tend to be based on vague notions of desirability under specific circumstances.

Simon (1979) views satisficing as one of four possible sub-goal termination strategies. He explains that when using a satisficing strategy the person will end their search and move on when a sub-goal has been achieved 'well enough' (i.e., when their search has yielded at least *n* pieces of required information). Other sub-goal termination strategies he discusses include aspiration achievement, whereby the sub-goal may be to achieve a certain level of gain; impatience, in which the search is terminated when a certain time has passed and the person selects the 'best so far'; and also discouragement, when the search may be terminated if a certain set of processes to achieve a sub-goal have been tried and have failed (Simon, 1979). The satisficing discussed in this thesis does not relate directly to Simon's concept of satisficing as a sub-goal termination strategy but rather it relates more to his concept of 'aspiration achievement' as a sub-goal termination strategy (as this relates to someone moving on when information drops below a certain aspiration level). Although the focus in this research is on that of people abandoning their a text when information gain drops below a pre-set threshold (called satisficing but more like Simon's 'aspiration achievement'), it must be taken into account that other plausible accounts do exist for why a person may abandon a resource in favour of searching elsewhere.

Indeed, Simon (1979) points out that although organisms are adaptive in learning and choice situations this adaptiveness still fails to result in maximising behaviour. He suggests that organisms adapt well enough by satisficing, so they do not tend to optimise. One of the primary problems for information foraging theory is to understand the mechanisms behind the way in which foragers allocate their time and attention to the most useful of the information sources available to them.

There exist time, resource, and opportunity costs of different information foraging strategies, and these are influenced by the structure of the interface between people and information sources (i.e., the physical layout of the information, functionality, and accessibility of the information). In applying the theory of information foraging, Pirolli and Card (1999) make the assumption that in order to maximize their rate of valuable information gain, people will adapt their strategies (e.g., they may do this through use of a satisficing or an optimising strategy), or if possible, adapt the structure of the interface between them and the information repository (e.g., make the information more accessible by bringing it closer, or organising it more efficiently). They claim that one cognitive strategy will be used in preference to another if it produces more useful information per unit cost. In addition to expecting that information systems (such as internet search engines) will inevitably develop with the aim of improving the returns on information foraging, these authors also suggest that one would expect that, faced with these foraging tasks, cognitive structures and strategies will also gradually develop through learning and practice to maximize the gain per unit cost. It is also possible, however, that people switch between different strategies when foraging for information. In some cases, for example in completely new environments or when searching information for new and specific tasks which the seeker has not encountered before, new foraging strategies may even need to be explicitly taught in order for people to be able to use them.

A primary aim of Pirolli and Card (1999), in their exploration of information foraging, is to better understand the extent to which information foraging behaviour is adaptive with regards the environmental context in which it arises. They term this 'adaptation analysis' (Pirolli & Card, 1999, p. 644) in the sense that in biology 'adaptation is demonstrated by observed conformity to a priori design specifications', (Williams, 1992, p. 40). They explain that adaptation analysis comprises three components: decision, currency, and constraint. In terms of information foraging, the first assumption refers to deciding how much time to spend on a particular source of information or whether or not to even consult certain sources; the second refers to how choices are evaluated in terms of maximising the currency (information gain); and the last refers to constraints (costs) that may arise. There are two types of constraint/cost involved in information foraging. Resource costs, which are the costs spent carrying out the activity, and opportunity costs, which are the costs lost by engaging in one activity over another in which there may have been benefits gained.

These will be explored in the following experiments by manipulating time and resource availability. It would be difficult for the forager to assess these costs without prior knowledge about each activity (information source) and the benefits that would accrue from it. The assessments that are made regarding these costs are therefore likely to be very uncertain. For example, a forager may decide to switch texts due to a lack of information gain, but this decision would not be based on the properties of the text they decide to switch to, since these are not known beforehand. The following experiments are set up in such a way that participants are unable to assess costs prior to opening texts, and therefore the experiments investigate the strategies people employ to overcome such problems of not being able to assess costs and benefits before selecting.

In terms of seeking information from written texts, the resource cost would be the time taken to read a certain text, and the opportunity cost would be the cost of spending time in one text rather than another which may contain more useful information. An information forager searching for task relevant information faces the problem of different information sources being more or less profitable with regards to the valuable information returned per unit cost, as well as being more or less accessible or effortful to process. Therefore, the best information forager is one who can maximize the rate of information gain under the constraints of the environment (profitability, access costs, etc).

Much of the information in a given environment occurs in patches. A patch of information could be a website, different papers, pages within a book, etc. When readers are allocating time across information sources, the information patch is assumed to be the section of text/information source that a reader will abandon when information gain has dropped below a certain threshold. If a patch was a website, then when the user had extracted all the information they felt useful from that site, they would abandon it and jump to the next website. If the patch was a page of a book, then the reader would leave the page when information gain began to drop below a threshold, and they would jump to the next page of that book. It is the job of the forager to direct themselves from one of these patches to another. The forager must decide how to allocate their time to between-patch and within-patch foraging. One way of doing this would be through *enrichment*, which is when information foragers mould the task environment in order to suit the strategies available to them. Examples of environmental enrichment include: reducing the average cost of

switching information patches (i.e., minimising between-patch foraging costs); and making information patches that yield better returns (i.e., changing the environment in order to improve within-patch foraging outcomes) by for example refining keyword queries for a search engine so the list it returns is potentially more relevant. The environment can be modified to minimize between-patch costs and maximize with-in patch results. However, this is not pursued in this thesis.

In order to understand Pirolli and Card's approach to information foraging, it is important to explore the basics of optimal foraging theory (Stephens & Krebs, 1986) that deal with adaptations of organisms to the environmental constraints of searching for food. Optimal foraging theory is based on the supposition that animals have been designed by natural selection to behave in ways which will maximise their fitness by maximising the net rate of energy intake while feeding. It follows, therefore, that given the constraints of any situation, it can be predicted how the animal is expected to behave while foraging. As Charnov (1976) explains, the prey of most predators is scattered across 'patches' and the animal must decide which patches to visit and when to move from one patch to the next. The problem the animal faces, if patch quality cannot be judged beforehand, is how long to stay in each patch. Charnov proposes that the optimal solution is for the predator to move from a patch when its capture rate within a patch drops below a threshold, which he calls the marginal capture rate, which is equal to the average capture rate for the environment as a whole.

All animals must solve their foraging problems using some kind of mechanism (or rule of thumb), and most research in this area deals with patch-leaving decisions (e.g., Green, 1984; McNair, 1982; Iwasa et al., 1981). All of these involve one or more of the following possible rules: a number rule, leave after catching n prey; a time rule, leave after t seconds; a giving-up time rule, leave after g seconds of unsuccessful search; and a rate rule, leave when intake rate drops to a critical value r (Stephens & Krebs, 1986). McNair (1982) examined performance of number, time, and giving-up time rules for animals foraging in a single patch environment and when their encounters with prey within the patch were stochastic. He claims that the nature of the gain function influences which rule is best. When comparing rules when there are two patches, McNair suggested that if the forager knows the quality of the patch beforehand, then they should be more persistent (i.e., have a longer giving-up time) in a good patch after a run of bad luck than in a bad patch (Stephen & Krebs, 1986).

The problem becomes more complicated when the patches vary in quality and cannot be judged before foraging. This is relevant to the experiments in this thesis because they involve situations whereby participants are given texts which vary in quality (Experiments 2-6) and they are not able to judge text quality before the reading phase. Green (1984) and Iwasa (1981) analyse this problem by studying the performance of the three rules of number, time, and giving-up time, when the forager is faced with patches which vary in the distribution of prey. Iwasa generalised that if each patch contains the same number of prey (but encounters with prey are stochastic) a number rule works best, when the number of prey in each patch varies a lot then a giving-up time rule works best, and when the number of prey in each patch follows a Poisson distribution then a time rule works best. As with Iwasa's (1981) model, Green's assessment model considers how the forager should make his stay/leave decisions when the patches vary in quality and when this quality cannot be recognised before searching. Green's assessment rule thus states that the forager should leave the patch if there is less than a certain level of gain after a certain number of looks. The increments are additive so that the rule responds to the average rate and not the local rate. This means that an initial burst of gain early on would have a cumulative effect in increasing the overall residence time in that patch. This makes it better than a simple giving-up time rule because it reduces the likelihood that the forager will leave the patch too early because of a run of bad luck initially. In relation to the experiments in this thesis, if a text had a lot of required information early on, say on the first page, then this may encourage the reader to stay longer in that text later on when they are not getting as much information, according to Green's assessment model.

However, Ward and Blaustein (1992) believe that the concept of satisficing has been misrepresented in the ecological literature. One aspect of satisficing theory in an ecological context is that the decision maker is satisfied after meeting some minimum requirement (e.g., additional increases in rate of energy intake over and above a certain threshold do not provide an increase in fitness). However, as Ward and Blaustein point out, the concept of a minimum requirement is not part of the satisficing theory formulated by Simon (1955) because there is no information processing constraint on the forager. Some researchers (e.g., Stephen & Krebs, 1986; Herbers, 1981) believe that merely accepting that animals are satisfying some minimum requirement too readily fit the facts because, as Herbers (1981) points out,

animals spend a lot of their time doing nothing and this supports satisficing because they stop feeding when they are satisfied by a certain intake criterion. This aspect of satisficing too readily fits the facts because any animal that does not feed continuously could be seen to be satisficing.

Constraints such as incomplete information or insufficient time are likely to limit the ability of the decision-maker to make an optimal decision. It is unlikely a decision-maker will remember or know all possible combinations of choices that would lead to an optimal strategy (Stephen & Krebs, 1986), and so the decision-maker is constrained to satisfy some minimum requirement instead. Reaching an optimal decision may also be constrained by time. To illustrate how information processing constraints prevent decision-makers from making an optimal choice, Simon (1955) uses the example of a chess player having to decide which move to make. The optimal choice for a move in a chess game could involve processing 10 to the power of 120 combinations and therefore the player must make a move that is satisfactory, thus applying a satisfying strategy.

Optimal foraging theory was devised with the purpose of making the connection between evolutionary biological theories and behavioural science theories. An optimal foraging model is developed by characterising the foraging situation and identifying the currency to be maximised, as well as specifying any relevant constraints which the forager may experience (DiClemente & Hantula, 2003). In the case of foraging, the currency is energy intake per unit of time spent foraging for the energy source. Whatever behaviour maximises this currency can then be identified, which can lead to experiments being designed to ascertain whether the observed behaviour shares these properties. DiClemente and Hantula point out that optimal foraging theory can be directly applied within many contexts such as time and availability of resources, both of which are addressed in this thesis. However, as mentioned earlier, Simon (1956) claims that humans do not maximise, or optimise, which suggests an apparent conflict between optimal foraging in animals and human information foraging.

1.3 Metacognition in study time allocation

Research on metacognition in study time allocation deals with comprehension and the allocation of attention when learning. The question that arises here is whether

information foraging requires metacognition (i.e., does information foraging require a level of thinking that involves active control over the process of thinking?). Planning how one should approach a learning task, monitoring how much is being understood, and evaluating the progress towards the completion of a task are all skills that are metacognitive in nature. In terms of animals foraging for food, it seems that the animal's main goal is finding food and they are driven by this desire to find available food. If information foraging in humans is analogous to food foraging in animals, then this would suggest that metacognitive factors are not present and that people are driven purely by their goal to find information. It seems, however, due to time and resource limitations, that information foraging needs to be monitored in a strategic way, and therefore involves the consideration of these issues when proposing a theory for human behaviour.

It must be pointed out that the paradigm used in the literature on metacognition in study time allocation is different to that used in this thesis. Most studies in this area involve giving the participant ample time before the study phase in which to evaluate the study materials and make judgements about which ones they wish to use or learn. The paradigm used in the study time allocation literature does not offer choice nor does it consist of time constraints, and therefore is very different to the experimental set-up used in this thesis.

In order for time allocation to be adaptive, readers must be sensitive to differences in the texts in terms of how much learning each is likely to support. It may be difficult for readers to judge this and time spent appraising the texts may be better spent studying. One adaptive approach would be for readers to focus their time and resources on texts which they judge to be within the *zone of learnablity* (term used by Wolfe, 1998). Metcalfe (2002) introduced the term *region of proximal learning* (following other researchers, such as Vygotsky's, 1987 *zone of proximal development* and Atkinson's (1972) *transitional state*) in order to explain the region of concepts residing just beyond the learners' grasp where the learner is most amenable to learning. This model suggests that materials that are not too difficult and not too easy will be most beneficial to the learner.

Dunlosky and Hertzog (1998), in their investigation of the mechanisms of metacognitive control of study time, put forward the discrepancy-reduction model. In this model, learning is monitored in terms of the difference between the learning goal and the level of information gain potential from the information source. This

judgement about the information source is made prior to the study phase. This model proposes that people allocate time on the basis of the amount of discrepancy which exists between the information source or item and their desired state of knowledge after learning (assuming the information source contains relevant information the reader wants to learn). Once they have calculated these discrepancies, the model suggests that people will devote most study time to the materials with the largest discrepancies (hence, the most difficult). If this model was used in information foraging it would require a time allocation strategy which separates the judgement phase from the learning phase (e.g., sampling) for the forager to identify which items have the largest discrepancies.

Most extant models of study time allocation suggest that people allocate more time to items that are judged to be difficult than easy (e.g., discrepancy reduction model, Dunlosky & Hertzog, 1998; Nelson & Narens, 1994). One reason for this phenomenon is that it would take longer for people to master materials which are more difficult (or judged to be more difficult) than it would for them to master the easier materials. Therefore, assuming that the reading goal is to learn the materials, people need to, and do spend more time studying the difficult materials (Son & Metcalfe, 2000), see also, Thiede and Dunlosky (1999).

Further support for this view (that people will devote more time to items with the largest discrepancies between the information source and the desired state of knowledge) comes from a review of study time allocation experiments conducted by Son and Metcalfe (2000), which revealed that out of the 46 different treatment conditions (of 19 published studies) examined, 35 showed that people chose to study the materials which were perceived as more difficult. None of the studies showed that people allocated more time to easier items. However, the tasks mainly involved the use of cue-target pairs, and none included the use of longer pieces of text, so it is difficult to directly compare the two. It could also be the case that the participants who were tested already knew the information contained in the easier items, and therefore did not wish to spend any time studying them. The results of this review nevertheless indicate that the way which people behave is highly strategic (also suggested by Nelson & Narens, 1994), and in keeping with what is proposed by the discrepancy-reduction model (Dunlosky & Hertzog, 1998; Thiede & Dunlosky, 1999).

However, Son and Metcalfe (2000) conducted a study whereby participants were required to study biographies under time pressure (30 minutes), and their results showed that participants devoted their time to the materials they judged as easy rather than difficult. They suggest two reasons that might explain these results: (1) there were far more materials to learn than there was time to learn them; and (2) the materials used were more complex and demanding than the simpler materials used in previous experiments. This research casts doubt on the generality of the discrepancy-reduction model (Son and Metcalfe, 2000). The researchers point out that in nearly all of the reviewed studies, the learning goal was verbatim recall, and participants were given plenty of time to study the information until they were content that they had learned the materials. It is possible, therefore, that the discrepancy-reduction model stands only under these specific experimental conditions, and that the model may fail to account for the way in which people allocate time to study materials under different conditions.

Son and Metcalfe's (2000) study discussed above revealed that people do sometimes allocate more time to the easier items, depending on the task and conditions. They found that when the information to be learned consisted of long written passages, and when participants were put under extreme time pressure, a preference for spending more time on the easier items was apparent. They also showed that when studying for a test, participants preferred to devote more time to the difficult items, and when free-reading they were more likely to devote more time to the easier items. The reading goal thus appears to have an effect on the allocation of time to study materials. It cannot be predicted from the research by Son and Metcalfe (2000) whether participants in the experiments in this thesis will spend more time on the easy or the difficult texts, since both time limit and specific learning goal are used. It is recognised that these factors may cancel each other out. That is, a time limit results in more time being spent on the easy texts, and a specific test results in more time being spent on the difficult texts, or that one influence may be stronger than the other.

According to the region of proximal learning hypothesis, there are many factors that contribute to the decision about to what materials the learner should allocate the most study time. These factors include the difficulty of the material, the expertise of the participant, and the time the participant spends studying (Metcalfe, 2002). Research suggests (Son & Metcalfe, 2000) that although people do use their

metacognitions to allocate time to study materials as stated in previous theories, on the whole they are much more situation sensitive and strategic than was previously thought. The results of studies by Metcalfe (2002) all point towards the fact that people appear to be allocating study time according to their own region of proximal learning, which changes depending on their expertise in the area and the time they have to study. This raises the point that people are generally adaptive in their allocation of time to learning materials, and the tendency for them to allocate more time to materials (e.g., texts) from which they will most benefit, given the time constraints placed on them. This ties in with the discrepancy-reduction model to the extent that allocating time to the materials/texts which have the largest discrepancies (i.e., which are most different from the learner's desired state of knowledge) could be seen as adaptive behaviour.

1.4 Relating these research areas to this thesis

1.4.1 Information Foraging

In the context of the problem with which this thesis is concerned, the texts in the experiments could be considered the individual patches of information, and readers would need to negotiate how much time to spend foraging within the patch (e.g., within pages of the text) and how much time to spend foraging between the patches (e.g., between different texts); switching patches in the experiments in this thesis is very fast, due to the set-up of the experiment. Foraging theory states that a forager will continue searching within a patch (e.g., a text or page) as long as the marginal benefit of the information they are gathering is greater than the marginal cost involved of moving to a new patch. If, as in the experiments in this thesis, there is little or no time cost involved in switching between texts or pages, then the forager is assumed to continue searching in a text or page (depending on what their patch size is) as long as the information they are gaining from that text or page meets or exceeds a certain level of desired information gain, and when it drops below this desired level then the forager will jump to the next patch.

However, there are elements of foraging for information from on-line texts which are different from animals foraging in the wild for food. One of these factors is the proximity of each available patch. Between-patch foraging is relevant to animal

foraging theory to the extent that one patch of food may be a fair distance from another patch of food, and so the costs of switching patches and spending time moving from one patch to the next are important and need to be considered as constraints (however, it could also be the case that the patches of food are very close to each other). In terms of people foraging for information from on-line texts, it is possible that movement from one text to another is relatively easy and so considering the costs of switching from one text to the next may not be an issue, but it is also possible that switching texts is difficult due to time it takes to find the new texts, when the cost of switching would need to be considered when deciding whether to remain in the current patch or switch to another. In the experiments in this thesis, however, all the texts are one button away, and so the cost of switching texts is not an issue. Information foraging theory may suggest that in this case, foragers would not need to consider between-patch foraging costs, and would instead concentrate on within-patch foraging costs.

Another issue that must be considered is the role of individual differences, which are likely to play a role in the allocation of time across texts. Information foraging theory does not make predictions about how people forage for information based on such factors. Age and spatial reasoning have been found to be important factors in information seeking tasks (Egan, 1988), and cognitive, physical, and emotional differences between and within individual information seekers is also likely to have an effect on their behaviour (Marchionini, 1995). These issues are not addressed in foraging theory, and therefore potentially challenge the robustness and generality of the theory. Information foraging theory does not consider the more subtle aspects such as readers' interest or preference for a certain style of text; but Son and Metcalfe (2000) found that people allocated time across texts according to their interests. These results are beyond the immediate scope of all other existing studies on the allocation of study time, and suggest that perhaps previous accounts of study time allocation fail to consider important aspects which do affect the learner's decision.

Information foraging theory relates to the experimental tasks in this thesis because the readers choose to forage for the information necessary to complete their post-reading test, and therefore the information sources are already selected for the reader to choose from. The participants do not have prior knowledge about the whereabouts of the information within these texts, and they are not driven by

semantics as they would be if foraging for information as the result of a search engine request on the internet. In the experiments in this thesis, readers forage for the information within a small set of texts, of which they have no prior knowledge. The number of texts available to readers relates to the number of patches available in which they can forage, which informs about movement between patches. The different number of task relevant facts each text provides relates to how relevant/valuable each patch is and informs about within-patch foraging.

1.4.2 Allocation of study time

Since perceived difficulty of texts is likely to affect the way in which people allocate time across texts, it would be beneficial to understand the strategies people use to find the information materials which are most appropriate for them. It has been established (Son & Metcalfe, 2000; Metcalfe, 2002) that when people are under time pressure they are more likely to devote more time to the easier materials (because they do not feel that they have enough time to learn the difficult materials), but it has not been explained, thus far, how people go about selecting these easier items. Taking this into consideration, Experiment 1 explores the possible strategies readers use when faced with multiple texts of varying difficulty from which to learn, and a limited time in which to complete their learning. Following from investigations of study time allocation, it is expected that readers will allocate the most time to the easier texts, or the texts which fall within their region of proximal learning, but what is of interest here is how they go about finding those texts. If readers are to focus their attention on texts from which they will learn most, how do they do this when they seldom have a priori information about the difficulty of the available texts? Experiment 1 aims to address this question.

Research on metacognition in study time allocation is linked with the research on information foraging in the sense that they both aim to understand how people manage to allocate more time to the study materials, or information sources, from which they will most benefit. One important and fundamental way in which the research in this thesis differs from previous research on metacognition in study time allocation is in not allowing the readers time prior to the study phase in which they can make difficulty judgements. As mentioned earlier, previous research has allowed preview time whereas the research in this thesis does not. In this way, one goal of this research is to look at how readers manage and integrate the tasks of evaluating and

studying the texts. In natural learning settings, learners are required to manage the integration of judgements about a) the information gain, and b) whether to continue reading or abandon to search elsewhere, with the reading process. The experiments that I present aim to identify and understand the strategies readers use to allow this integration to occur. These strategies will now be discussed.

1.5 Time allocation strategies

There are a number of strategies which could be used by readers who are required to allocate their limited time across multiple texts. This thesis concentrates on the two main strategies which have previously been identified (Reader & Payne, 2007) as being used by readers foraging for information. However, this is not to say that they are the only strategies which may exist. Other possible strategies will also be discussed and the data will be examined to identify whether any of these alternative strategies are being used by participants. The two strategies that are of primary interest in this thesis are those of sampling and satisficing.

Sampling

The key characteristic of sampling is that reading to evaluate the text is considered a separate process from reading to learn about the information contained in the text. A judgement about the suitability or usefulness of the text is made prior to the learning phase. Sampling is recognised in the study of behavioural ecology as a way in which animals inspect various patches of food before deciding from which one they will feed (Krebs, Kacelnik, & Taylor, 1978). Analogous to this, readers who are allocating limited time among a number of texts may adopt a similar strategy whereby they may rapidly inspect various information sources (texts) before deciding which one will be most beneficial for them to read, depending on their reading goal. Once the texts have been evaluated, one will be chosen to study from. When this chosen text has been read to the reader's satisfaction, then they may go on to sample more texts, or move to the next judged best text. The primary aim of the sampling phase of a sampling strategy is not to learn about the topic, but to learn about the properties of the texts (although it is acknowledged that some learning will inevitably take place during the sampling of the texts). This technique has the advantage that the reader

will find the 'best' text for their learning needs, but the disadvantage of time spent exploring and evaluating the texts that could have been spent learning about the topic. This kind of strategy could be considered as being at the opposite end of the scale to satisficing (discussed next) in the sense that sampling could be viewed as an optimization strategy, whereby the reader is searching for the best or optimal outcome or text source to exploit. This is contrasted to a satisficing strategy, whereby the reader does not search for the 'best' or optimal information source, but rather concentrates on searching for a good enough information source/text to exploit.

Satisficing

Satisficing is a decision-making strategy whereby a person ceases to search for alternatives when they find a resource whose expected utility meets or exceeds a previously determined threshold of what the person deems as satisfactory (Byron, 2004). Satisficing is characterised in the current research by the idea that the evaluation process is integrated into the reading phase so that judgement and learning take place simultaneously. Rather than searching for the 'best'/optimal text the reader will continue to read anything that is deemed to be satisfactory, that is to say, above a pre-set threshold level of information gain. A reader will begin reading a text and continue to read if information gain is high enough, or abandon reading if the text is not seen as useful or informative for the task in hand. The primary aim using this strategy is to learn about the topic and not to pre-judge the texts for usefulness. This strategy has the advantage that all the participant's time is spent learning about the topic, while constantly monitoring that the text is 'good enough' by continually assessing how much information is being learned. However, there is the disadvantage that the reader may spend too much time in lower quality, or non-optimal, texts.

Other strategies

One alternative model of time allocation across multiple texts may be a completely random allocation of time across the information sources available, whereby readers randomly select texts to read from the list and then randomly jump around the patches within a text (the pages) and the patches within a page (the paragraphs or text sections). Environmental issues could also drive a time allocation strategy in the sense that if the texts are listed on a menu bar or if they appear in a

certain sequence, then information foragers may just consult each text in turn according to where it appears on a list. Indeed, people may adopt this kind of strategy in addition to one of satisficing or linear reading (discussed next).

A second alternative model of time allocation across multiple texts may involve linear reading, whereby people read linearly from beginning to end. A variation on this model may even involve a kind of accelerated linear reading strategy whereby readers estimate the amount of time they can afford to spend per page of text, given that they know how many pages/texts are available, and spread their time accordingly. If they underestimated the time it would take to read a page of text then their reading pattern would look like a linear pattern, and if they overestimated the time then perhaps it would appear to be linear reading, but with added revisits to texts in the time they had left over. Similarly, participants may allocate time across the texts by dividing the number of pages available by the time they have to read, and allocate a set amount of time to each page. This may result in the reader having to move to the next page before finishing the current one, or result in the reader having to skim read the pages. At the extreme, under a pro-rata time model, the reader may allocate time by reading the beginning of each paragraph only, so that they can consult all the pages available in the time allowed.

Another possible method for allocation of time across texts which would predict a similar, but not identical, pattern of reading behaviour to the satisficing model is that of a search model. A search model may involve participants reading through the texts in a linear fashion in search of answers to specific questions (the post-reading test). It would predict that readers would quickly scan the texts searching for this target information, and would pause only when this information was encountered. Readers would then read the target information, after which they would begin scanning again. This model assumes two levels of processing; do process A until find a target (i.e., scanning the text searching for target information); then do process B (i.e., stop when identify target and read the text to learn the information); then resume process A (i.e., begin scanning again). The points at which the reader changes processes, from scanning to reading and vice versa, must involve some kind of threshold for the reader to make these decisions. These two thresholds are unspecified and are therefore flexible. It may be that these thresholds are the same, or they may be different. For example, the first may be based on familiarity judgements whereby the reader stops to read when they recognise a target item from the test. The

second threshold may be based on the reader having read all the information associated with the target item/word/phrase, or possibly when the reader has finished reading a certain section of text related to that target item. If the second threshold, when to stop reading and begin scanning again, is specified by the reader having finished reading a certain section of text, then this is not unlike a satisficing strategy which assumes that the text is divided into patches and that readers will jump from patch to patch according to their level of information gain in each patch. Under the search model, however, readers would not jump from one patch to the next, but rather just modify their reading speed from scanning to reading to learn as a transition from one patch to the next. The search model would also predict that participants should spend longer on texts from which they will most benefit (as they would stop scanning and read the text more when texts contain more target information), as does the satisficing model. In this way, the model would appear to elicit adaptive behaviour on the part of the reader in allocating most of their time to the texts from which they will most benefit. This model is discussed in more detail in Chapter 3, Section 3.4.

1.6 Review of Reader and Payne (2007)

Some prior evidence exists which suggests that, under certain experimental conditions, people will adopt a satisficing strategy when allocating time across multiple texts. Reader and Payne (2007) have conducted some research into the effects of time limits and summary outlines on participants' browsing behaviour. Their first study examined a situation in which participants were required to learn from four texts on the topic of the human heart (these were texts used by Wolfe et al., 1998) with the aim of writing a short essay (250 words). Participants were given either 15 minutes or 7 minutes in which to study. Results showed that most participants displayed a satisficing strategy in both time conditions. The simple criterion the researchers used to classify a participant as employing a sampling strategy is that if the person is sampling then they should visit all four texts in the first third of the experiment. They do not necessarily need to leave the fourth text and return to another, as they may decide after looking at the first three that the fourth text is best. In order to identify satisficing, the researchers proposed that the first visit to each text must be the longest visit ever made to that text, because a satisficer should

keep reading a text until it drops below threshold and because they would learn more from a text the first time they visit it.

Reader and Payne (2007) acknowledge that the prevalence of satisficing in their first experiment may be a result of the participants' inability to judge the usefulness of the texts on the basis of a rapid visit to them. Therefore, they conducted a second study using an outline facility which would aid the rapid judgement of the texts. The outlines were constructed by taking the first sentence of every third paragraph in the text to ensure that the outline was representative of the entire text. Their second experiment also examined learning more closely by including a pre-essay writing task and a post-test task. This way they could explore the effects of browsing strategy on learning. Participants were assigned to one of two conditions, with outline or without outline, and were given 15 minutes to read the texts. Results showed that all participants in the outline condition used the outlines at some point during their reading. There were more participants in the outline condition who sampled than in the text only condition. The researchers argue that this shift to sampling when outlines were present could be due to participants not being willing to sample texts by rapidly visiting a small fragment of them (e.g., the first page), because they may not feel that it would be representative of the quality of the entire text. In short, they suggest that the presence of outlines changes the cost-benefit structure of sampling by reducing the likelihood of inaccurate judgements being made. They found no difference in amount of learning according to strategy or condition. Reader and Payne (2007) conclude from their second study that the increase in sampling in the outline condition indicates that sampling is a strategy which is available to most participants, but that it appears to be judged as less useful in adaptively allocating time in text-only situations.

Reader and Payne (2007) conclude that the results from their first study indicating that preferences for texts were influenced by the participants' expertise (i.e., more knowledgeable participants allocated less time to easier texts) supports the idea that study time allocation is strategic and adaptive, in line with previous research in this area (e.g., Metcalfe, 2002). They suggest that these results show that readers are sensitive to text difficulty even under time pressure and when there is no time for prior judgement of the texts.

In short, Reader and Payne (2007) found that, when under time pressure and learning for an essay writing task, the behaviour readers displayed when allocating

time across four texts of varying difficulty was consistent with a satisficing model. However, when they introduced summary outlines for each text, then they found that participants were more likely to sample these outlines before choosing which text to read. Despite these results being interesting, they are not conclusive, and there exist some limitations to the studies. For example, the learning goal in Reader and Payne was to study with a view to writing an essay. This may be considered as being too easy or too vague as a task, since it could be viewed that any text will do if reading to write an essay. Reader and Payne also found that participants favoured the easier text, even though it was taken from a children's encyclopaedia and the population sample consisted of psychology undergraduate students. This preference for the easier text could have been due to the reading goal (to write an essay), or the fact that participants were under time pressure when studying, or that the writing style was more engaging and enjoyable for the reader.

The experiments in this thesis will look at the broader picture and examine factors which may affect browsing behaviour which have not been examined in detail previously. Although it has been seen that satisficing occurs when readers are under time pressure, it may be the case that when readers are not under time pressure they may adopt a different kind of strategy to allocate their time across texts. If a satisficing strategy is a solution to the problem of selecting a satisfactory resource in a limited time, then taking away the time constraint may encourage readers to adopt more of an optimising strategy to find a 'better' resource rather than a satisfactory one.

For the purposes of identifying how readers allocate their time across texts in the experiments in this thesis, the operational definitions that were used in Reader and Payne (2007) for sampling and satisficing will be used here. Sampling is characterised by evidence that the reader has visited all of the texts briefly, and then either remained in the last one visited, or returned to one of the visited texts. In foraging theory, however, sampling may not necessarily need the animal to inspect all sites before foraging, depending on the number of sites available. Thus, to be classified as a sampler in the following experiments, participants need to visit all of the texts available in the first third of the reading time. This rule is based on the fact that if people are sampling before selecting, then this would involve visits to all texts initially within a timescale such that after sampling the person still has a large

proportion of their reading time in which to study the selected text and learn. Although this criterion for a sampling strategy is arbitrary, it is nonetheless useful in helping to classify how readers spend time when reading texts.

The concept of satisficing used in this thesis is also that laid out by Reader and Payne (2007) for use in their studies. The key aspect of the satisficing strategy is that exploration of the texts is integrated into the reading process, rather than a precursor to it (as in sampling). An operational definition for satisficing would be if the first visit to each and every text is the longest visit ever made to that text. This would suggest that the reader is evaluating the text as they read. When they leave the text they do not need to return for long periods of time since they will either have already got the information from it they require, or will have decided that the text did not contain enough of the information they required.

A simple interpretation of satisficing is for readers to treat the whole text as the unit which they evaluate, so when the text falls below their pre-set threshold of information gain, then the reader would leave the current text and switch to the next text. In this way, the unpopular texts would be identified as having had visits to the first page or two of the text only. However, it is possible that the unit evaluated is smaller.

The line of argument in this thesis assumes the satisficing model as one which could conceivably account for the way readers allocate time across multiple texts, and the following experiments therefore aim to provide evidence for or against this model. Browsing behaviour will be observed and analysed to identify whether it can be explained using the satisficing model, or indeed any of the alternative models discussed in this chapter.

This research does not aim to identify *all* possible time allocation strategies, but rather it aims to explain how the majority of people allocate time across multiple written texts, thus focussing on the two main time allocation strategies examined in research conducted in this area (Reader and Payne, 2007) that looked at other factors to those examined in this thesis. However, there are interpretations of the data other than just satisficing or sampling, and this thesis will explore the possibility that in the event that participants' behaviour does not fit into either of these models, then it may fit with one of the alternative strategies discussed above. Indeed, the mere fact that some readers do *not* fall into either of these two categories implies that either there must be other strategies, or a combination of the sampling and satisficing strategies,

that readers use when searching for information from written texts, or that perhaps the categories may have been defined inappropriately. In the event that the readers' behaviour does not fit with either a satisficing or a sampling strategy, for the purposes of this thesis they will be referred to as 'residuals'.

Satisficing is well documented as a decision making strategy (Simon, 1956), but my research focuses on the idea that reading involves decision making, more specifically a process that influences the reader to use a satisficing strategy to find the information that they are seeking. Although prior studies have shown that readers may adopt a satisficing strategy when dealing with multiple texts under time pressure (Reader & Payne, 2007), it has not yet been identified at what level of text readers are making these stay/leave decisions, and thus with what unit of text they are satisficing. Are they satisficing at the level of the entire text, on the basis of one page of text, one paragraph or section of text, or even at a very small unit of text such as a sentence? In order to answer these questions Experiments 4 and 5 will use eye tracking techniques to investigate at what level of text readers are making their judgement decisions. By tracking the movements of readers' eyes it will be possible to see how much of each text is read before making decisions about it. Knowledge about when and where in a text the readers make their judgement decisions is potentially useful for those involved with writing on-line materials, since it gives an idea of where salient information would be best placed.

One of the issues in this thesis concerns the question of whether or not readers are adaptive when allocating time across multiple texts. It is therefore appropriate to define what is meant by 'adaptive' in the current context. From an ecological perspective, adaptive behaviour is when properties in the environment are altered in order for the organism (in this case, the reader) to meet its goal. The field of foraging theory considers optimal behaviours, and looks at observed behaviours compared to ideal solutions. The approach taken here (in line with the approach taken in the studies by Reader and Payne, 2007) considers behaviour to be adaptive if it is preferable to simpler behaviour that is not sensitive to the properties of the environment. It is not known what is optimal in terms of allocating time across texts, but it is assumed that adaptive allocation of time across texts is likely to be sensitive to the differences in usefulness of the available texts, such that readers will be able to distinguish between useful and not so useful texts and allocate more time to the useful texts.

1.7 Aims and objectives of the thesis

The research in this thesis provides an account of the strategies readers use when allocating time across written texts. As well as placing the existing research into how time is allocated across texts under time pressure (Reader & Payne, 2007) in a wider empirical context, the research presented in this thesis examines the problem of how readers select appropriate, or relevant, texts without a priori information about these texts. The eye tracking experiments in Chapter 4 also provide a detailed account of where the reader looks for the information within a page of text. One principal way in which this research differs from previous research by Pirolli and Card (1999) is that it considers those aspects of time allocation across texts that are more focussed, because the number of texts available and the time participants have to study are both less than were studied by Pirolli and Card. This means that participants' preference to allocate more time to one text over another may be associated as being part of the reading process itself, rather than an occurrence that precedes it. Another difference is the fact that the decision to read a certain text must rely on factors other than its relevance, and in many real world situations (e.g., browsing search engine results) most of the texts are likely to be potentially relevant. Thus, this research investigates how time is allocated when all texts are potentially relevant to the task of learning for a specific, pre-seen test.

The main questions addressed in the thesis are a) whether readers are able to adaptively allocate their time across multiple texts in limited time (that is, whether they are able to allocate more time to the texts from which they will most benefit), b) what strategies they use to achieve this, and c) whether the strategies employed depend on the layout of text in the information interface. Information seekers may prefer designs of information sources which improve the returns on their information foraging (considering limited time, and goals of seeking information), which will in turn have an effect on the information sources they select. If the strategies that people use to forage for information are better understood, then it will inform us about the kinds of design from which information seekers are likely to gain most information, and therefore inform us about which sources are likely to be selected. Information sources can then be designed to be selected with this in mind.

1.8 Summary of Experiments

Experiment 1 will assess how readers allocate their time across multiple texts on the same topic but of varying difficulty. Since the amount people are likely to learn from a text depends on the relation between their background knowledge and the level of difficulty of that text (zone of learnability hypothesis, Wolfe, 1998), then it seems important that people are able to select the text most appropriate to their level in order to maximise learning. The time participants have to study (either under time pressure or not under time pressure), and the number of texts available for reading (either four texts or eight texts) will be manipulated. The rationale for this experiment is to investigate whether the patterns found in prior research (Reader & Payne, 2007) are robust under variations in these parameters (study time and number of texts). Since these are factors which inevitably vary depending on the situation, it follows that their effect on time allocation strategies could potentially be important. The results of this experiment will also provide a rich analysis of time allocation across texts, and by increasing the number of texts available, observing participants' behaviour when they are *not* under time pressure, and giving participants a specific learning goal, it is possible to see whether participants' behaviour is consistent with the satisficing model. If so, then these findings will serve to elaborate on existing preliminary findings regarding the nature of satisficing as a strategy to allocate time across multiple texts (cf. Reader & Payne, 2007).

Experiments 2 and 3 will investigate how readers allocate their time across four texts which vary in relevance to the test for which participants are studying. All texts in these experiments are the same level of difficulty but vary in the number of required facts they provide the reader. If the strategies used to allocate time across sources are affected by the apparent relevance of the text, it would be beneficial to observe the behaviour of participants when the relevance of the text is varied. Experiments 2 and 3 will assess whether or not participants are able to identify the texts most useful to them in terms of the study goal, and if they do then what strategies they use to achieve this.

Experiments 4 and 5 will use eye tracking to provide more detailed information about the level of text at which readers are making their judgements about moving from one page/text to the next. The reason for carrying out these

experiments is that they will help to characterise how a text is accepted or rejected on the basis of the information it is providing the reader. These experiments will help to determine what readers are considering as a patch, since they will measure the degree to which paragraphs, sections of text, and pages are visited and subsequently left.

Finally, Experiment 6 will investigate whether specifically advising readers of the existence of one 'best' text for their reading goal will change their time allocation strategy in order to find this text. This experiment will help to identify whether the time allocation strategies employed by readers depend on the experimental conditions or the situation in which the reader finds themselves with regard prior knowledge of the likely usefulness of the texts available to them. These findings will highlight whether satisficing, or the behaviour which manifests the existence of this strategy, is a product of the information environment (that is, the quality or relative usefulness of the available texts), and if so whether being aware of this environment will have any effect on people's allocation of time across texts.

CHAPTER 2

Experiment 1: Time allocation across multiple texts

2.1 Introduction

When faced with multiple texts on a single topic, how does the reader decide how to allocate their time? The possibilities include satisficing and sampling. These are adaptive strategies for allocating time across multiple sources of information, since both allow the reader to allocate more of their time to 'better' texts. Prior research by Reader and Payne (2007) found a stronger tendency for the satisficing strategy rather than a sampling strategy among undergraduates who were reading in order to write a summary essay. Those participants were presented with four texts and given time limits of either seven or fifteen minutes. They were required to complete a pre-test and their learning goal was to write an essay after the reading phase. However, as discussed in Chapter 1, there are other possible interpretations of the data which do not relate to satisficing and sampling. It is possible that some of these alternative models also account for readers' behaviour when allocating time across multiple texts.

Satisficing may have been due to any or all of the three characteristics of the study by Reader and Payne (2007). Readers only had four texts in which to 'forage' or allocate their time across, they were under time pressure, and their reading goal of learning to write a summary essay could be seen as being too vague for easily identifying target material. Experiment 1 addressed each of these issues to explore whether these findings could be generalised to a wider empirical context. The experiment examines whether the readers' behaviour can be explained by a satisficing strategy when there are more than four texts to read, and it also investigates readers' preferred strategy when they *do* have enough time to study the available texts (when they are not under time pressure). There is less need to allocate time wisely when there is no time pressure, and so participants may be encouraged to sample the texts and take their time to choose the best text from which to learn. The reading goal is also changed, and participants in Experiment 1 are asked to read the texts with a view to answering more questions like the ones already answered (in the pre-test). This reading goal is different to that used in Reader and Payne (2007), where readers were told to read in order to write a summary essay. Experiment 1 therefore enriches

research in this area by exploring new factors which could potentially influence how readers allocate their time across multiple texts. These manipulations will identify whether a strategy such as satisficing is a direct reaction to the existence of limited time or limited resources.

The use of pre- and post-tests (as used in Reader & Payne, 2007) enabled investigation into the adaptivity of readers, that is, whether more knowledgeable readers would spend more time on more difficult texts and less knowledgeable readers spend more time on the easier texts. Based on the work of Wolfe et al (1998), it can be assumed that selectivity in browsing is related to expertise in that the more expert the reader, the more time they will allocate to the more difficult texts. If readers *do* spend more time on texts from which they will learn more, then this would suggest that browsing is adaptive.

The time-pattern of browsing and whether participants displayed a satisficing or a sampling strategy when studying the texts, or indeed any other kind of strategy, was of primary concern in this study. All of the texts in Experiment 1 are about the same topic, with varying levels of difficulty, and readers do not have information prior to the reading phase as to the difficulty of the texts. Therefore, participants are presented with the problem of how they will choose the text which will be most conducive to their learning. This choice may be made by either of the browsing strategies described earlier, sampling or satisficing, or indeed any of the alternative strategies discussed in Chapter 1.

If it is assumed that people read selectively, then this must affect their comprehension of the reading materials. The problem remains as to how comprehension can be successfully measured. There are many methods of testing comprehension (e.g., knowledge tests, free recall, writing tasks) and they all have limitations associated. It was therefore thought best to use converging methods, and the two chosen were a written short answer test (pre- and post-reading scores were compared), and the marking of short essays using Latent Semantic Analysis (LSA) because it offered an objective measure of free recall. See Appendix A for information about LSA.

One method for calculating the quality of an essay is to use the Intelligent Essay Assessor, which uses LSA theory to compute the semantic similarity between the text and the essay as a grade which is assigned to the essay to indicate quality. This is a web-based service provided by Pearson Knowledge Technologies which

supplies instant feedback on the content and quality of a piece of writing (*Intelligent Essay Assessor*, 1998-2005). The automaticity and speed of LSA to quickly return measurements of semantic similarity makes it a useful and productive application for research in text comprehension.

2.2 Method

Participants

Sixty-four participants took part, fifteen male and forty-nine female (mean age 20.16, S.D. 1.16, range 19 to 24). Most participants were psychology undergraduates who received course credit for their participation. The others were undergraduates from other university departments, who were paid for their participation. Participants were assigned to experimental conditions according to their order of arrival at the laboratory.

Design and Materials

Experiment 1 was a between-participants design, and analysed the effects of two independent variables. The first of these was the number of texts which were available to read: participants were provided with either four or eight texts to study. The second independent variable was the effect of time constraints on the participants' reading strategies and learning outcomes. Participants were given either 15 minutes (timed) or between 30 and 45 minutes (un-timed) to study the texts. The un-timed condition was set up such that participants were required to read for the first 30 minutes, and when they reached this time they could either proceed to the test or continue reading for up to a further 15 minutes, after which time they were forced to continue to the test. The un-timed condition was a duration which was felt to be long enough for the reader to feel relatively unconstrained by the time. The dependent variables were the browsing behaviour of the readers, identified by analysing the amount of time spent in each text, and the participants' scores on essays and tests before and after the reading phase.

Eight texts on the topic of the human heart and circulatory system were used in this study. The texts used in the four text condition were those used by Wolfe et al. (1998) and since Experiment 1 focussed on altering the factors in Reader and Payne (2007), who also used these texts, they were used here to ensure comparability. The length of the texts ranged from 1,533 to 1,672 words and they varied in level of difficulty. Text A was taken from a children's book (Silverstein & Silverstein, 1983),

Text B was an introductory text for adult readers (Davis & Park 1981), Text C was written for undergraduate anatomy students (Basmajian, 1982), and Text D was taken from a medical school book (Gould, 1953). The additional four texts used in the eight text condition were taken from a variety of sources and the difficulty level was matched as closely as possible to the four original texts: Text E – (Burnie, 1995); Text F – (*Texas Heart Institute*); Text G – (Mohrman, 2003); Text H – (*Collier's Encyclopedia no. 11*, 1992). The difficulty level was matched using the Intelligent Essay Assessor found on the LSA website. It compared each text with the existing four texts on the same topic that are stored in its database and produces cosine values for the similarity between the text input and the four original stored texts. Therefore, the four additional texts were designed such that the similarity cosines between them and the four original texts were as high as possible. Text E and text A had a similarity cosine value of .89, text F and text B had a cosine value of .86, text G and text C had a cosine value of .79, and text H and text D had a cosine value of .82. All the texts were four pages long and ranged from 1275 to 1672 words in length.

The essays were scored using Intelligent Essay Assessor (IEA). When the four texts were analysed it emerged that they differed in the number of text questions they answered. The two intermediate texts (B and C) provided more answers than the easiest text (A) or the most difficult text (D). In terms of questions answered, out of a maximum of 40 points, text A scored 25, text B scored 33, text C scored 33, and text D scored 20. This discrepancy in the texts' ability to supply the reader with the answers to the test questions was not mentioned by either Wolfe (1998) who originally used these texts, or by Reader and Payne (2007). These texts were used in Experiment 1 despite this discrepancy in the quality in order to be able to compare the effects of changing the number of texts available and the time pressure, with the findings from these previous studies.

The texts were accessed via a menu system at the left of the screen which was visible throughout the reading phase, and allowed participants to access any page of any text at any time. This ensured that readers could choose freely which parts of the texts they studied. The menu consisted of four buttons for each text, which were labelled with the text author (for example, 'Silverstein' or 'Gould') and the page number one to four. There was no other information given about the texts. Therefore, readers were unable to use the menu to discriminate between texts according to their difficulty, relevance, or length. Figure 2.1 shows the reading screen which was

presented to participants in the eight text condition. The screen for the four text condition was similar except there were only four texts to choose from rather than eight. The text on each page was presented as a block of text in single line spacing, with no paragraphs. The font used was MS Sans Serif, regular style, Western script, size 10. The average line length was 100 characters. There was an indication at the bottom of the screen detailing how many minutes they had been reading so that participants could keep track of how long they had been reading and how much time remained.

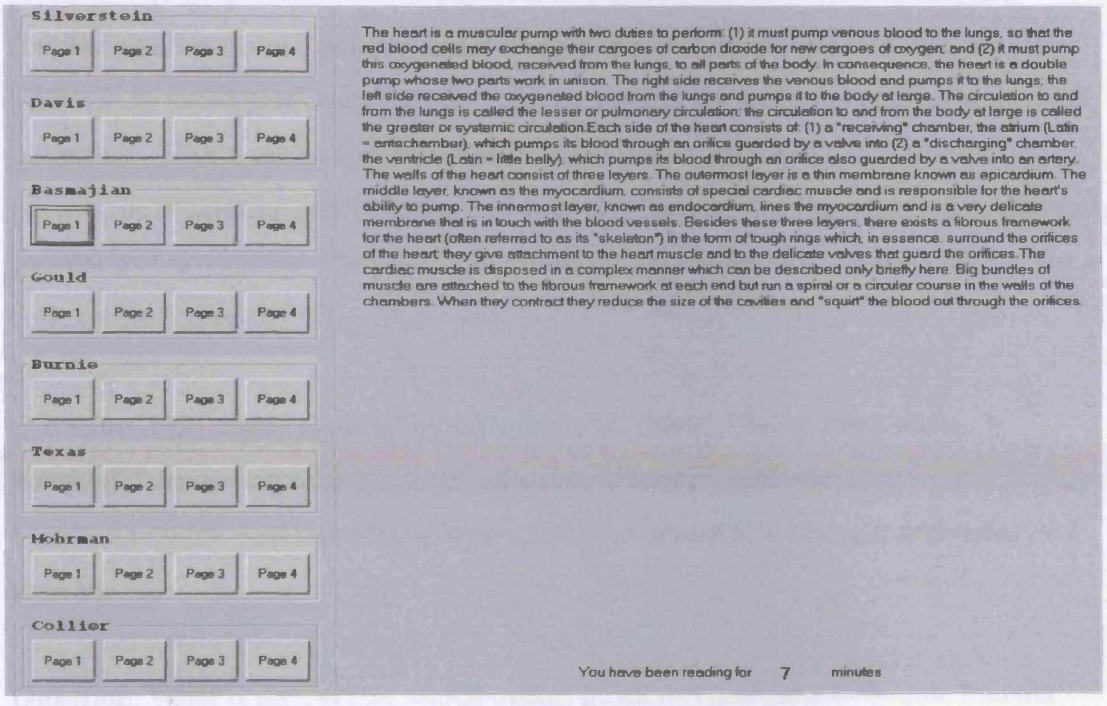


Figure 2.1: The reading interface for the eight text condition (page 1 of the text by Basmajian, 1982, is selected here).

The order in which the texts were presented in the menu (i.e., from top to bottom) was varied for each participant. Twenty-four different orders were used for the four text condition, and thirty-two for the eight text condition (this meant that each participant in the eight text condition was presented with a different text order). The orders for the four text condition were derived using a Latin square design (24 combinations in total) and so all of these combinations were used and eight were used again. The eight combinations which were used twice were selected from across the twenty-four already generated. Two orders were randomly selected from each set of

combinations which began with text number 1, 2, 3, and 4 in the first position. This was so that each text appeared in the first position four times each (twice in each condition timed and un-timed). As participants clicked on the text pages the software recorded the time. There was no facility on the screen to inform participants which pages had been viewed, only which page was currently being viewed.

An on-screen short answer test about the heart was given to the participants before and after the reading phase. Questions appeared on the screen and participants were required to type the answers below each question (see Appendix B for test layout). The 17 questions were split over two pages, and once participants had finished the first screen page and moved onto the next, they were unable to return. This was to avoid participants finding the answers to the earlier questions in the later questions. Within each page participants were free to answer the questions in any order. The maximum score for the test was 40. The test was the one used in the study by Wolfe at al. (1998). Two example questions appear below (See Appendix C for a copy of the full test):

Question. How many types of blood vessels are there? Please name them.
Answer. *5 types - capillaries, arteries, veins, arterioles and veinules.*
Mark. *6 points - 1 for number of types, 3 for the names (6 if also get arterioles and venules)*

Question. What is the protein which makes quick oxygen/carbon-dioxide transfer possible? How many molecules of oxygen can each such protein carry?
Answer. *hemoglobin. each protein can carry 4 mollecules of oxygen.*
Mark. *2 points, 1 for hemoglobin, 1 for 4 mollecules.*

Procedure

Participants were first asked to write a 250 word essay on the anatomy, function and purpose of the human heart. There was no time limit for this, but participants were required to write within 10 words of the word limit before they were able to proceed to the next stage of the experiment. This restriction was imposed to enable easier analysis of the essays after the experiment, using the Intelligent Essay Assessor. To help them to keep track of how many words had been written, they were

provided with a word count facility on the right of the screen which had to display a number between 241 and 259 in order for the proceed button to be enabled. There was no such restriction imposed in the Wolfe et al (1998) study, and therefore although the instructions indicated that participants should write 250 words, there was much variability in essay word count. Studies show that when essay length is less constrained, essay word count is strongly related to knowledge (Rehder, 1998). The question of how long an essay needs to be in order to provide an accurate estimation of the participants' knowledge was investigated by Rehder et al (1998), who examined essays from participants in the Wolfe et al study. They looked at how effective LSA was as a measure in predicting the relation between the pre-questionnaire score with the number of words in the participant's essay. It was discovered that the first 60 words of an essay are non predictive, and between 70 and 200 words was increasingly predictive, but with decreasing marginal returns. Based on this analysis, a word count of 250 words was chosen for our study.

If participants' knowledge of the heart was very limited and they were unable to write 250 words, the set up of the experiment made it possible for them to write anything just to make up the number of words. However, all essays were read through by the experimenter after the experiment to ensure that this had not occurred. Once completed, the test was presented and participants were required to type the answers in the spaces provided. Once the test had been completed the participants were presented with the menu for the text reading phase. They were given either 15 minutes or between 30 and 45 minutes to study either 4 or 8 texts, depending on the condition. These will subsequently be referred to as the timed condition (15 minutes) and the un-timed condition (30-45 minutes). Participants were told to read with a view to answering more questions *like* the ones they had already answered in the pre-text. They were also told how much material there was from which they could study (number of texts, number of pages, and number of words). They were not advised that they would be required to write a second essay. In the timed condition, once the 15 minutes had passed the screen went blank and participants were asked to proceed to the next stage. In the un-timed condition, the 'proceed to next stage' button was enabled after 30 minutes had passed, but participants could still access the texts until 45 minutes had passed, when the screen went blank. Participants were then asked to write another essay of 250 words. The same rules applied as in the pre-reading phase essay. Once they had completed the second essay they were presented with the

original test and asked to complete it. Again, the same rules applied as for the pre-reading phase test.

2.3 Results and discussion

The results will be split into the following sections; time allocation strategy (including data on time spent on texts and pages, and number of visits to pages and texts), systematicity, effects of text difficulty, adaptivity and effects of prior knowledge, and learning scores. Data from three participants had to be excluded from analysis as they failed to follow the instructions properly: one participant in the four text timed condition (4T), one in the four text un-timed condition (4UT), and one in the eight text un-timed condition (8UT). One participant was randomly chosen and eliminated from the eight text timed condition (8T) so that there were fifteen participants in each experimental group.

Time Allocation Strategy

One of the main interests in this study was the time allocation strategy of the participants, that is, whether they used a sampling strategy or a satisficing strategy. The figures in Table 2.1 result from classifying browsing behaviour using the following operational definitions (see Chapter 1 for more details about these definitions):

Satisficing: the first visit to each and every text must be the longest visit ever made to that text.

Sampling: participants make short visits to all of the texts before making a choice of which text to read. This is characterised by all of the texts being visited in the first third of the experiment time, so for those in the timed condition this would be 5 minutes and for those in the un-timed condition this would be 10-15 minutes, depending on how long they spent reading.

Residual: this label was assigned to behaviour which did not fall under either of the other two categories, satisficing or sampling. An example of a residual may be if a participant did not quite visit all of the texts in the first third of the time (so not

classed under sampling), or that not all of their first visits to each text were the longest visits made to that text (so not classed under satisficing).

Analysis of the participants' browsing behaviour shows that there were 76.7% satisficers, 6.7% samplers, and 16.7% residuals across the four experimental conditions. This implies that the browsing patterns of over three-quarters of the participants can be explained by the satisficing model. Table 2.1 shows the breakdown of reading strategy for each condition:

Table 2.1: Number of participants displaying each defined reading strategy in each condition

Condition	Using original criteria for sampling			Binomial probability of satisficers versus samplers (original criteria):	Using 4-text-sampling criteria		
	Satisficer	Sampler	Residual		Satisficer	Sampler	Residual
4T	10	2	3	0.02			
4UT	12	1	2	0.01			
8T	12	1	2	0.01	12	1	2
8UT	12	0	3	0.001	9	5	1

Table 2.1 shows a significantly higher number of participants adopted a satisficing strategy than a sampling strategy across conditions (all $p = 0.02$). Neither increasing the number of texts nor removing the time limit changed the dominant browsing strategies.

A more detailed analysis was carried out on the ten participants who were classed as residuals. There were five participants in the four text condition who were classed as residuals, and three of these satisficed with three texts, and two satisficed with two texts. That is, the first visit to three/two texts was the longest visit ever made to those texts but for the other texts the re-visits were longer than the initial first visits. There were five participants in the eight text condition classed as residuals, and three of these satisficed with six texts, one satisficed with seven texts, and one who did not satisfice at all and whose behaviour appeared relatively random. Overall, although these participants were classed as residuals, they did seem to be displaying a satisficing type of behaviour, since their pattern of browsing indicated that they satisficed with some or most of the texts they viewed.

If for the eight text conditions the criterion for sampling is relaxed in terms of the number of texts which need to be visited in the first third of the time to qualify as

a sampler, then more participants in the un-timed condition could be counted as samplers. This analysis is interesting because perhaps expecting participants to sample with eight texts is too much, and if they were sampling but with a smaller number of texts then this would be missed under the original criteria for sampling. Thus, working on the criterion that participants need to visit four out of the eight texts in the first third of the reading time in order to be classified as a sampler, then there were five '4-text-samplers' in the un-timed condition (the term '4-text-sampler' being used to refer to participants visiting four texts in the first third of the experimental time). Comparing the number of samplers in the eight un-timed condition according to the original versus the new 4-text-sampling criterion, there were significantly more 'samplers' under the new criterion than the original, as would be expected (Fisher-Yates Exact test, $p < .05$). There was no change in the number of samplers in the eight timed condition under the 4-text-sampling criterion. A Fisher-Yates Exact test also revealed that there was no significant difference between the two eight text conditions (timed versus un-timed) with participants classified as per the new '4-text-samplers' criterion ($p = 0.16$).

As well as trying to identify the strategy being used it is also interesting to characterise the participants' behaviour in a more theory neutral way, to check for differences between conditions in terms of behavioural outcomes, whatever strategies might be used. Data regarding number of texts and pages visited is summarised in Table 2.2 where it can be seen that, as expected, more pages were visited when there were 8 texts rather than 4 texts (because there were more texts to visit), and the time limit (15 or 30/45 minutes) had most effect on mean time per text visit (see Figure 2.2). This data on simple time per page/text and number of pages/texts visited during reading can indicate whether readers were exhaustive in their reading, and also whether the existence of time pressure or additional texts caused them to alter how and when they judged the value of a text (i.e., did they read one page and move onto another text, or consult all pages of a text before leaving?).

Table 2.2: Number of pages and texts visited, and time (sec) allocated to texts across conditions

	Timed condition		Un-timed condition	
	4 texts	8 texts	4 texts	8 texts
	mean (*S.D.*)	mean (*S.D.*)	mean (*S.D.*)	mean (*S.D.*)
number of unique pages	12.40 (*3.52*)	21.67 (*7.59*)	15.60 (*1.55*)	23.0 (*7.19*)
mean number of page visits	16.33 (*6.79*)	25.87 (*9.13*)	27.20 (*10.32*)	31.80 (*8.07*)
mean time per page visit	64.57 (*29.60*)	39.25 (*16.36*)	79.09 (*32.31*)	64.65 (*22.44*)
number of unique texts	3.73 (*0.59*)	6.46 (*1.55*)	4.0 (*0*)	7.20 (*1.42*)
mean number of text visits	5.67 (*2.35*)	9.20 (*4.21*)	8.47 (*3.44*)	11.80 (*4.84*)
Percentage of time in individuals' favourite text.	45.82 (*16.25*)	36.13 (*9.80*)	37.76 (*8.94*)	37.99 (*19.03*)
z score	1.12 (*0.27*)	1.66 (*0.29*)	1.13 (*0.24*)	1.76 (*0.4*)

Analysing the number of pages and texts visited helps to determine the browsing strategies used by readers. Table 2.2 suggests that, as would be expected, participants made more text visits in the eight text conditions and in the un-timed conditions and participants made more page visits in the eight text conditions and in the un-timed conditions. Also, as expected participants in the eight text conditions visited more *unique* pages and participants in the eight text conditions made more *unique* text visits than those in the four text conditions. Analysis of variance (ANOVA) with factors of time (timed, un-timed) and texts (four, eight) showed a main effect of text, $F(1,56) = 12.07$, $p < .01$, a main effect of time, $F(1,56) = 7.47$, $p < .01$, and no interaction, $F < 1$. A parallel ANOVA of the number of page visits, with factors of time and text showed a main effect of text, $F(1,56) = 10.08$, $p < .01$, a main effect of time, $F(1,56) = 14.25$, $p < .01$, and no interaction, $F(1,56) = 1.2$, $p = .27$. An ANOVA of the number of unique pages visited, with factors of time and text showed a main effect of text, $F(1,56) = 33.55$, $p < .01$, no effect of time, $F(1,56) = 2.48$, $p = .12$, and no interaction, $F < 1$, and an ANOVA of unique texts visited with factors of time and text showed a main effect of text, $F(1,56) = 110.23$, $p < .01$, no main effect of time, $F(1,56) = 3.13$, $p = .08$, and no interaction, $F < 1$.

Analysing the time participants spent on text visits and page visits shows, as expected, that participants read texts for longer when there were fewer and read texts for longer when they had more time (see Figure 2.2). An ANOVA with factors of

time and text supported this by showing a main effect of text, $F(1,56) = 1003.98$, $p < .01$ and a main effect of time, $F(1,56) = 1088.1$, $p < .01$. There was also an interaction between time and text, $F(1,56) = 101.27$, $p < .01$. Simple effects showed that all differences between conditions were significant (all $p < .01$) except between the 4 timed and the 8 un-timed conditions.

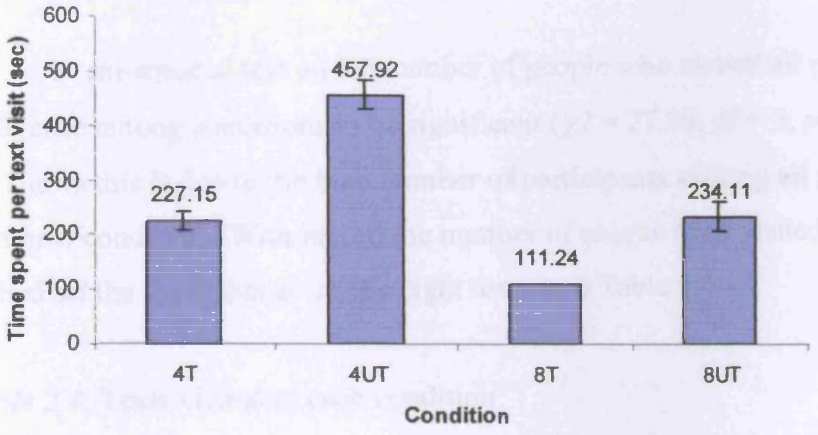


Figure 2.2 The mean time (with +/- S.D.) participants spent per text visit for each condition in Experiment 1.

Participants in the four text conditions and the un-timed conditions spent longer on average on a page visit (a main effect of text number, $F(1,56) = 8.81$, $p < .01$, a main effect of time, $F(1,56) = 8.88$, $p < .01$, and no interaction, $F < 1$). Looking at the amount of time participants spent in any single text also gives information about the browsing strategy used. Participants in the eight text conditions spent more time studying their favourite text (relative to the number of texts available to them, measured as a z score in Table 2.2) than those in the four text conditions (a main effect of text, $F(1,56) = 56.6$, $p < .01$, no effect of time, $F < 1$, and no interaction, $F < 1$). This suggests that the more texts that are available to the reader, the longer they will spend in their favourite text (single text in which they spend most time).

If people are judging small units of text then they would visit more pages than if they were judging large units of text. Therefore, looking at the number of unique pages visited by participants gives an initial indication of the manner in which they make the stay/leave decision. Table 2.3 shows the number of participants who visited all available pages and the mean number of pages visited in each condition.

Table 2.3: Pages visited in each condition

Condition	People (out of 15) visiting ALL available pages	Number of pages available	Mean pages visited (*S.D.*)
4T	5	16	12.4 (*3.5*)
8T	2	32	20.7 (*8.3*)
4UT	14	16	15.6 (*1.6*)
8UT	2	32	23 (*7.2*)

A chi-squared test on the number of people who visited all pages revealed the difference among conditions to be significant ($\chi 2 = 27.29$, df = 3, $p < .001$). Table 2.3 shows this is due to the high number of participants visiting all pages in the four un-timed condition. With regard the number of unique texts visited, many people visited *all* the four texts or *all* the eight texts (see Table 2.4).

Table 2.4: Texts visited in each condition

Condition	People (out of 15) visiting all available texts	Mean texts visited (*S.D.*)	*p* value for visit all versus not visit all
4T	12	3.7 (*0.6*)	< .05
8T	6	6.5 (*1.6*)	= .15
4UT	15	4 (*0*)	< .01
8UT	10	7.2 (*1.4*)	= .09

Again, a chi-squared test revealed the difference between the number of texts visited between conditions to be significant ($\chi 2 = 14.04$, df = 3, $p < .01$), but now this seems due to the low number of participants who visited all eight texts when reading time was limited to 15 minutes. The binomial test shows that significantly more participants in the four text conditions visited all the available texts than failed to visit all available texts, but the difference in the eight text conditions was not significant.

If participants are satisficing, then in order to determine at what level of the text they are doing so (i.e., are they satisficing with small units of texts such as sentences or paragraphs, or with larger units of text such as pages or even the entire text), data on the number of pages visited in a first visit to a text will help. If participants are satisficing at the level of the text, then first visits to unpopular texts would be characterised by single page visits. The data in Table 2.5 suggests few participants make single page first visits to texts.

Table 2.5: Number of pages visited in the first visit to a text

Condition	Strategy	Number of pages read at first visits to text			
		1 page	2 pages	3 pages	all 4 pages
4 timed	satisficers (10)	1	0	2	7
	samplers (2)	1	1	0	0
	residuals (3)	1	0	1	1
4 un-timed	satisficers (12)	0	0	0	12
	samplers (2)	0	0	0	2
	residual (1)	0	0	0	1

The data in Table 2.5 also shows that there is an effect of time on the number of pages visited during reading a text. A Fisher-Yates Exact test revealed that there is a significant difference between the number of people who did not read all of the pages of a text (people who read less than four pages) in the four timed and the four un-timed conditions ($p < .01$). Table 2.5 shows that more participants in the timed condition failed to visit all the pages of a text than did participants in the un-timed condition, where they all visited all four pages.

Systematicity

In order to examine how systematic participants were when presented with multiple texts to read, data was collected on the order in which participants visited the texts available to them. Table 2.6 shows the order in which texts were vertically listed in the menu, with text number 1 being the text at the top of the screen, and the order of texts visited by each participant in the four text conditions.

Table 2.6: Text orders presented and visited (first repeat visited in bold)

4 timed		4 un-timed	
order texts presented	**texts visited**	**order texts presented**	**texts visited**
1234	1234	1243	124312
2134	2134341	2143	2413
3124	2314	3142	2413
4123	234132342	4132	2132
1324	1423	1342	12341323
2314	31241	2341	4123
3214	31	3241	4213132
4213	324143	4231	4132434
1423	21342132	1432	1432
2413	213414141	3421	43214243
3412	3412	4321	432132343
4312	3421	4123	2341
1243	12342	3214	3214214
2341	412	2413	31423413
3421	421	1432	1432

The table shows that all participants in these conditions were systematic in the sense that they did not revisit any text until all texts had been visited at least once, with one exception in the 4 un-timed condition where one participant visited three texts and then returned to the first text visited without visiting the fourth text. The data also showed that browsing was exhaustive at the level of texts. Only three out of fifteen participants in the four timed condition failed to visit all available texts at least once, and in the four un-timed condition only one participant failed to make at least one visit to each of the available texts. In the timed and un-timed conditions, seven and eight participants respectively visited the text in the top position on the menu bar first. Presentation order of the texts is another tool that readers may use to decide how they will allocate their time across the available texts. Surprisingly, only half of the participants in Experiment 1 started reading the text at the top of the list. Reasons as to why this may be are discussed later (see Section 3.2.3).

In the eight text condition, participants were also systematic. Only 3 out of 15 participants in the eight timed condition (all satisficers) and 4 out of 15 participants in the eight un-timed condition (two samplers and two satisficers) ever revisited texts without having visited all other texts first. Table 2.7 shows the order of texts presented and the order participants visited the texts. For simplicity, only the first ten text visits are shown, but the data still clearly shows the tendency for participants to visit all of the texts before revisiting any.

Table 2.7: Order of the first 10 text visits

8 timed		8 un-timed	
order texts presented	**texts visited (first 10)**	**order texts presented**	**texts visited (first 10)**
12358674	1238467576	14675382	1862
25634718	7145	23516784	4128356478
42586371	8261358	36417582	4863285786
51624837	2475271	58346712	78341562
65187423	37862154	61483527	2753
74256318	7381542	73246815	7251632478
85234176	6871345	84561372	58623471
15862347	156213	17264358	13657428
21748356	21647	27814536	4175682317
34861572	583762	38421576	5413687251
47153286	3651482725	48621573	5481637274
52431786	5243186	56183247	36218436
64857213	7682	63815274	462851734
73524861	84253	71825436	2424765812
86573241	86573241	87253614	7358462135

In the eight timed condition only four out of the fifteen participants visited all the eight texts at least once during their reading time. In the eight un-timed condition ten out of fifteen participants visited all the eight texts at least once during their reading time. Only three participants in the un-timed and six in the timed condition began reading the text in the top position of the menu bar.

Effects of text difficulty

Four texts condition

An analysis was carried out to examine the proportion of time spent in the easiest and the hardest texts (texts A and D respectively) to ascertain whether available time had any effect on the amount of study time allocated to these texts. The hypothesis was that the proportion of time spent on text A relative to text D should be higher in the timed condition, because time pressure leads one to favour easier texts. On this basis, it is predicted that an interaction effect will occur.

Table 2.8: Proportion of total time spent on easiest text (A) and hardest text (D) (4 text condition)

Text	Condition		
	4 Timed	4 Un-timed	Overall mean (*sd*)
A	39.79 (*18.18*)	26.49 (*13.86*)	32.94 (*17.35*)
D	11.33 (*9.34*)	17.11 (*8.75*)	14.22 (*9.37*)

Table 2.8 shows the time spent in A and D only (the two extremes) but for the sake of completeness, Appendix D shows the means for all texts. More time was spent on the easy text in the timed condition than in the un-timed condition, but there were no differences in the proportion of time spent on the difficult text across the timed versus un-timed conditions. More time was spent overall on the easy text than the difficult text. An ANOVA with factors of text (A, D) and time (timed, un-timed) confirmed this description of the data by revealing that there was a main effect of text, $F(1,28) = 20.73$, $p < .001$, an interaction between text and time, $F(1,28) = 5.62$, $p < .05$, but no main effect of condition, $F(1,28) = 2.62$, $p > .05$. This suggests that people overall prefer text A, but this preference is greater in the timed condition (as supported by the simple main effect of A v D for timed, $F(1,28) = 23.97$, $p < .001$, but not un-timed, $F(1,28) = 2.38$, $p > .05$).

Eight text condition

The time participants spent in the four original texts A, B, C, and D in the eight text condition was analysed to investigate whether they spent more time on texts within their zone of learnability. The mean time spent on text A averaged across timed and un-timed conditions was 182.58, (*S.D. 174.2)*, for text B the mean was 154.7, (*S.D. 151.1)*, for text C the mean was 207.3, (*S.D. 273.0)*, and for text D the mean was 95.9, (*S.D. 128.8)*. These figures suggest that there was no significant difference in the time participants spent on these four texts. A 2 by 4 ANOVA with factors of time and text confirmed this description by revealing that there was no main effect of text, $F(3,84) = 1.66$, no interaction between text and time, $F < 1$, but, as expected, a main effect of time, $F(1,28) = 18.07$, $p < .001$. Clearly, participants would spend more time on each of the texts in the un-timed condition because there was more time available to spend on them.

As with the four text condition, an analysis was carried out to examine the proportion of time participants spent on the easy texts and the hard texts. Since there were two easy texts and two hard texts in the eight text condition, the analysis was based on the mean of the two easy and the two hard texts for each condition, timed and un-timed. Again, the prediction was that participants would spend proportionally more time on the easy texts than the hard texts in the timed condition, based on the idea that time pressure results in readers spending more time on the easier texts. Table 2.9 shows the mean proportion of time spent in the easy and the hard texts for the eight text condition.

Table 2.9 Proportion of time spent on the easy texts, A and E, and the hard texts, D and H for timed and un-timed (eight text conditions)

Text Category	Text	Condition		
		8 Timed	8 Un-timed	Overall mean (*sd*)
Easy	A	16.77 *(16.93)*	11.68 *(10.85)*	81.32 *(127.62)*
	E	15.15 *(12.50)*	6.38 *(8.59)*	71.38 *(102.54)*
	Overall	**15.96** *(14.65)*	**9.03** *(9.98)*	**12.49** *(12.91)*
Hard	D	8.41 *(11.13)*	15.80 *(13.37)*	12.10 *(12.66)*
	H	11.85 *(13.54)*	16.24 *(21.74)*	14.04 *(17.94)*
	Overall	**10.13** *(12.31)*	**16.02** *(17.73)*	**13.07** *(15.42)*

Inspection of the data in Table 2.9 suggests that more time was spent on the easy text when participants were under time pressure than when they were not. An ANOVA with factors of text (easy, hard) and condition (timed, un-timed) revealed that there was no main effect of text difficulty ($F < 1$), no main effect of time pressure ($F < 1$), but a significant interaction between text difficulty and time pressure ($F(1,58) = 6.07, p < .05$). Analyses of simple main effects revealed that within the timed condition there was no difference between the time participants spent on easy or difficult texts ($F(1,58) = 2.5$), and within the un-timed condition there was also no significant difference between the time spent on the easy and difficult texts ($F(1,58) = 3.61, p < .063$). For the easy text, more time was spent reading it in the timed than the un-timed condition ($F(1,58) = 4.59, p < .05$) and for the difficult texts, no differences were found between the timed and un-timed conditions ($F(1,58) = 2.24$). The finding

that more time was spent on the easy texts in the timed condition is consistent with the prediction that time pressure leads to more time spent reading easier items than when the reader is not under time pressure (see also Son and Metcalfe, 2000).

Adaptivity and effects of prior knowledge

It was expected that people with higher pre-test scores would spend least time in the easiest text (A). However, there was no significant correlation between the pre-test score and the time spent in text A in the four timed condition, and a significant negative correlation (r = -.666, N = 15, $p < 0.01$,) existed in the four un-timed condition: readers evidently showed more adaptivity when they had more time available, spending less time in the easiest text the more knowledgeable they were about the topic.

For the eight text condition the mean time each participant spent on the two easy texts was correlated with the pre-test score and revealed that there was a significant negative correlation for both the timed (r = -.565, N = 15, $p < .05$) and the un-timed conditions (r = -.523, N = 15, $p < .05$). When more texts were available, participants were adaptive both when they were under time pressure and when they were not.

In order to achieve a more powerful statistical test of the extent to which participants were adaptive, a median split was done on the pre-test scores for all participants. The median split was 12.

Condition: 4 texts.

The time participants spent reading texts A and D was analysed according to whether they were in the timed or un-timed condition, and whether they were in the high or low pre-test score group. Table 2.10 shows the mean time participants in each group spent in text A and text D. For the sake of completeness, Appendix E shows the time spent in each of the four texts, A to D, according to the high/low pre- test score group divide.

Table 2.10: Time spent in texts A and D in both four text conditions by people with high/low pre-test scores

Text	Time	Pre-test	Mean	*S.D.*	N
A (easy)	Timed	Low	370.13	*120.85*	10
		High	319.35	*235.69*	5
	Un-timed	Low	665.93	*283.05*	6
		High	351.39	*133.84*	9
D (hard)	Timed	Low	91.01	*86.25*	10
		High	119.31	*80.99*	5
	Un-timed	Low	276.37	*178.18*	6
		High	342.23	*160.61*	9

A three way ANOVA with factors of text (easy v difficult), time, and pre-test score (high v low) showed that participants spent more time on the easy text A than they did on the hard text D ($F(1,26) = 16.79, p < .001$) and a significant interaction between text and pre-test knowledge, ($F(1,26) = 4.6, p < .05$). In terms of between-participants effects, there was a significant main effect of pre-test score group, $F(1,26) = 5.13, p < .05$ (those in the low score group spent more time on the easy text A than the difficult text D), and, as would be expected, participants spent longer reading the texts in the un-timed condition ($F(1,26) = 37.8, p < .001$).

Comparing pre-test knowledge (high v low) at each level of time (T v UT) for each text (A v D), participants in the low group spent more time reading text A than did participants in the high group ($F(1,26) = 10.32, p < .01$) in the un-timed condition. No other differences were significant ($F < 1$). Those with low pre-test scores also spent more time reading text A in both the timed and un-timed condition than they did on text D (timed, $F(1,26) = 9.8, p < .01$; un-timed, $F(1,26) = 11.46, p < .01$). There were no significant differences for time spent on text A and D in the high group (largest $F(1,26) = 2.52$).

As expected, participants spent more time reading the texts in the un-timed condition than in the timed condition. Simple effects revealed these differences to be significant for all except for text A at the high group level (text A, low group, $F(1,26) = 9.5, p < .01$; text D, low group, $F(1,26) = 7.31, p < .05$; text D, high group, $F(1,26) = 9.06, p < .01$).

Condition: 8 texts

Again, the time participants spent reading texts A and D in the eight text condition was analysed according to whether they were in the timed or un-timed condition, and whether they were in the high or low score group. Table 2.11 shows the mean time participants in each group spent in text A and text D. Appendix F shows the time spent on all four original texts, A, B, C, and D.

Table 2.11: Time spent in texts A and D in eight text condition

text	Time	Pre test	Mean	*S.D.*	N
A8 (easy)	Timed	Low	244.96	*173.17*	7
		High	68.71	*64.64*	8
	Un-timed	Low	205.36	*212.17*	9
		High	227.48	*180.6*	6
D8 (hard)	Timed	Low	41.23	*42.88*	7
		High	105.78	*127.6*	8
	Un-timed	Low	158.5	*180.39*	9
		High	52.34	*74.42*	6

An ANOVA with factors of text (easy, A, v difficult, D), time, and pre-test score (high v low) showed that participants spent more time reading text A than they did reading text D ($F(1,26) = 6.38$, $p < .05$) and a significant interaction between text, pre-test knowledge and time ($F(1,26) = 5.75$, $p < .05$). There were no between participants effects of group ($F(1,26) = 1.57$) or time ($F(1,26) = 1.37$).

Comparing time spent on texts A and D at each level of time for each pre-test knowledge group revealed that participants in the low group in the timed condition spent more time reading text A than text D ($F(1,26) = 6.69$, $p < .05$), and that participants in the high group in the un-timed condition also spent more time reading text A than text D ($F(1,26) = 4.24$, $p = .05$). All other differences (i.e., between texts in the un-timed condition for low group and in the timed condition for high group) were not significant ($F < 1$).

Learning Scores

Essay Measure

The essay score analyses are based on the data of fifty-two participants. The Intelligent Essay Assessor (IEA) was unable to score eight of the essays (either pre- or

post-reading). This affected three in the four timed condition, one in the four un-timed, one in the eight timed, and three in the eight un-timed condition. This problem could be for a number of reasons. Before scoring the essay, the IEA checks to see if it would be confident in its assessment of the essay and if it is not then no score is returned for the essay. The software gives the following reasons as possibilities for why it cannot assess an essay: the essay may not be written in complete sentences, the essay may be off topic, or the essay may be highly original or creative or use language that the system has not been taught. The maximum score the IEA can give is five.

Pre- and post-scores

As Figure 2.3 shows, the post-essay scores were higher than the pre-essay scores, as would be expected if the participants were learning. A 2(time) by 2(text) by 2(pre/post score) repeated measures ANOVA revealed a main effect of pre/post score, $F(1,40) = 26.02$, $p < .001$. There were no interactions. There were no between participant effects of time ($F < 1$) or text ($F(1,40) = 2.02$, and no interaction ($F < 1$).

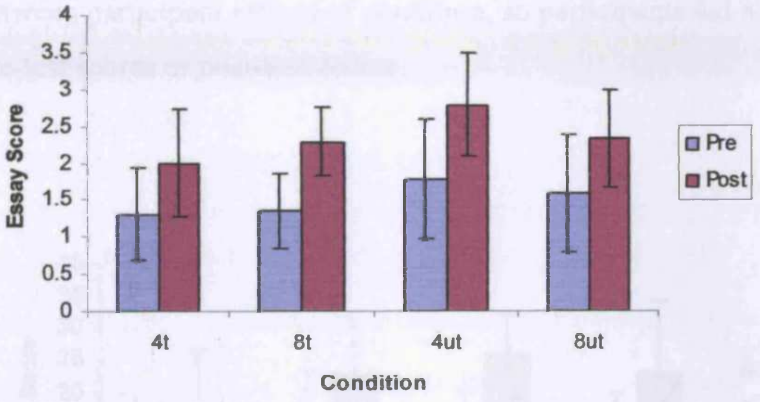


Figure 2.3 Mean pre- and post-essay scores (with +/- S.D.), according to IEA software, for participants in each condition, 4t, 8t, 4ut, and 8ut, for Experiment 1.

Learning score

Following Wolfe et al. (1998), learning scores were computed to take into account the amount participants actually learned from their reading, rather than relying on measures of pre- and post- essay score (see Table 2.12). The learning score used is the proportion improvement between the pre-score and the post-score, calculated by dividing the difference in the pre- and post-scores by the difference

between the maximum score (40 in this case) and the pre-score. An ANOVA revealed that there was no significant difference in the amount participants learned (the learn score) across conditions ($F < 1$).

Table 2.12: Learning scores

Condition		Essay learn score		
Time	Texts	Mean	*S.D*	N
Timed	4	0.19	*0.16*	12
	8	0.26	*0.08*	14
Un-timed	4	0.26	*0.32*	14
	8	0.19	*0.19*	12

Test Measure

Pre- and post-scores

The post-test scores were higher than the pre-test scores, as would be expected if the participants were learning (see Figure 2.4): $F(1,56) = 155.53$, $p < .001$. A 2(time) by 2(text) by 2(pre/post score) repeated measures ANOVA also showed an interaction between pre/post score and time, $F(1,56) = 7.99$, $p < .05$. There were no between participant effects of condition, so participants did not differ across either pre-test scores or post-test scores.

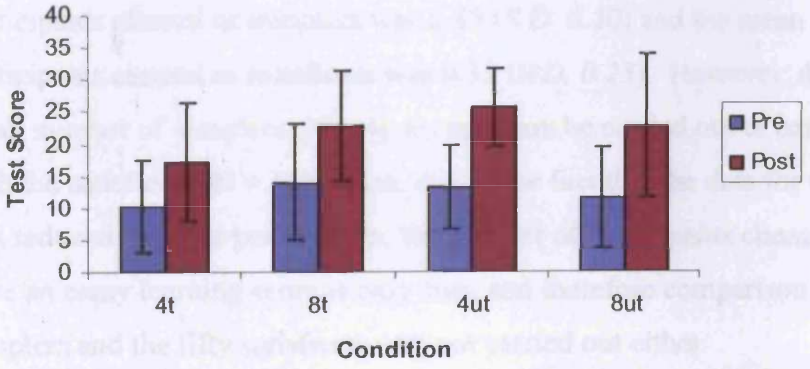


Figure 2.4 Mean pre- and post-test scores (with +/- S.D.) for participants in each condition, 4t, 8t, 4ut, and 8ut, for Experiment 1.

Learning score

Again, a learning score was computed to minimize any effect of variation in pre-test scores. Figure 2.5 shows the mean learning scores for each of the four conditions. An ANOVA revealed that there was no significant difference in the amount participants learned (the learn score) depending on time or number of texts ($F < 1$).

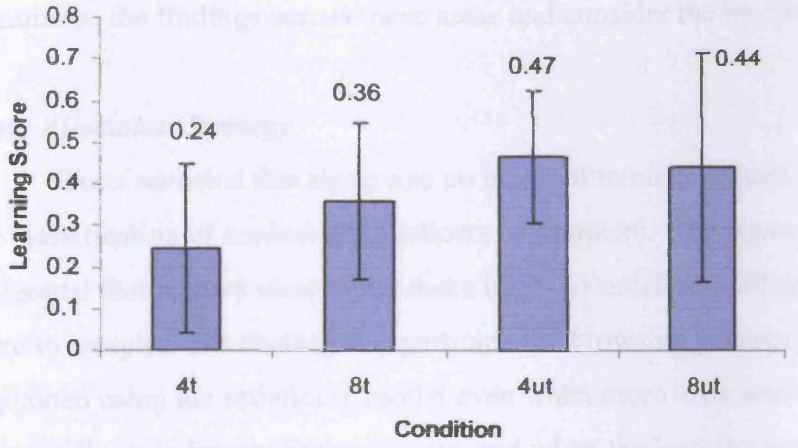


Figure 2.5 Mean learning score (with +/- S.D.) according to the test measure for each condition, 4t, 8t, 4ut, 8ut, for Experiment 1.

Learning score and browsing strategy

According to the short answer test measure, the mean learning score for participants classed as samplers was 0.45 (*S.D. 0.20*) and the mean learning score for participants classed as satisficers was 0.37 (*S.D. 0.23*). However, due to the very small number of samplers (N = 4), no tests can be carried out to compare this data with the satisficers (N = 56). Also, due to the fact that the data for the essay measure was reduced by eight participants, the number of participants classed as samplers who have an essay learning score is only two, and therefore comparison between these two samplers and the fifty satisficers was not carried out either.

2.4 General Discussion

This chapter has investigated the effects of time constraints and text number on participants' browsing strategy and learning outcomes. The main focus of the experiment was to investigate the effects of time and resource factors on browsing strategies which had not previously been considered. This discussion is separated into the following sections: time allocation strategy, selectivity/adaptivity, and learning score. After discussion of these issues, there will be a general conclusions section to summarise the findings across these areas and consider the broader picture.

Time Allocation Strategy

Tests revealed that there was no effect of number of texts or time available on the classification of readers as satisficers or samplers. The browsing patterns suggested that readers were much more likely to satisfice in all conditions than they were to sample. The finding that participants' browsing patterns could still be explained using the satisficing model even when more texts were introduced, when they were not under any time pressure, and when the learning goal was different, builds on the preliminary findings of Reader and Payne (2007), showing that this strategy for allocating time across texts is not an artefact of number of texts available or time pressure.

The pattern of behaviour of the participants did not fit with most of the alternative strategies laid out in Chapter 1 either. It is clear from the data that participants were not simply reading linearly, since the time participants spent across easy and difficult texts differed significantly. If they had been reading linearly, or even following a model of accelerated linear reading, then the pattern of data would reflect a much more even distribution of time across texts (see Tables 2.3 and 2.4). It is also clear from the systematic manner in which participants visited texts and pages that their time allocation strategy was not random. It is problematic trying to apply the search model to data from Experiment 1, because participants were not aware of the specific information they needed to learn for the test. It could be that new information is seen as target information for readers in Experiment 1, as they were asked to learn for an unseen test, but this suggestion is speculative, and therefore fitting the data to the search model seems problematic.

Based on the results of Reader and Payne (2007), it is perhaps unsurprising that satisficing was the more frequently used strategy under the conditions of four texts and limited time, despite the fact that it is, in a sense, counter intuitive. One might assume that people briefly assess what is available to them before making a decision about which text to settle on, rather than simply choosing a source and reading it as long as it meets their level of satisfaction. It may be that sampling is not considered a desirable strategy as people are not prepared to judge the value of a text on the basis of the first page or paragraph. It may also be the case that the participants thought that all of the texts could be read in the 15 minutes reading time (although they were advised about the number of texts and pages available).

Another factor potentially contributing to the preponderance of satisficing is the way in which the texts were presented to the participants. They were only able to view one text at any one time, so the cost involved in switching texts versus remaining in the current text in the hope that it will get better as they read on may be influencing participants in their time allocation strategy. However, the cost involved in switching texts in these experiments seems low compared to what it could be in other situations.

Assuming a simple interpretation of satisficing (discussed in Chapter 1) where readers treat the whole text as the unit which they evaluate, the unpopular texts would be identified as having had visits to the first page or two of the text only. This is not supported by the data in Experiment 1. In the four timed condition, for participants classified as satisficers, only 10% of first text visits were to a single page whereas 70% of first text visits were to all four pages. Similarly, in the four un-timed condition for participants classified as satisficers, 100% of first visits were to all four pages (see Table 2.7). There was therefore a tendency for participants to consult all or most of the pages of a given text, with the less favoured texts being scanned, rather than completely abandoned after visiting a single page.

There was also found to be an effect of time on the number of pages of a text visited, as more people read less than all four pages in the timed condition than in the un-timed condition. This could reflect how the participants may be feeling under pressure to hurry through the texts, since in the un-timed condition all four pages were always consulted. The skimming behaviour evident in Experiment 1 can be explained by the satisficing model when it is assumed that the unit of text which is being evaluated is smaller than the entire text, for example a page or section of a page (e.g.,

a paragraph, or since in Experiment 1 a page was the same as a paragraph, then a quarter or third section of the text). It is possible that readers use several small sections of text as a basis for satisficing because they do not feel that they can judge the quality or usefulness of the entire text merely on the basis of one small section. If this is the case, then it would explain the infrequency of sampling, because if readers cannot quickly judge the value of a text on the basis of the first page or section, then sampling is not useful.

For most participants in the four text conditions, browsing was exhaustive at the level of texts. In the timed condition, three out of fifteen participants failed to visit all available texts at least once, and in the un-timed condition all participants made at least one visit to all available texts. In the eight text conditions, more participants failed to make visits to all available texts; timed condition nine out of fifteen failed to visit all texts and in the un-timed condition five out of fifteen failed to visit all eight texts at least once during the reading phase. This suggests that participants were more likely to visit all available texts when there were fewer texts on offer, and that having more time did not result in participants in the eight text condition necessarily visiting all texts. Looking more closely at the participants in the four timed condition who failed to visit all the available texts, the text visits they had made would not have resulted in them acquiring all the answers to the test, so having obtained the information they required would not have been a reason for them not consulting all available texts.

Browsing behaviour was also systematic, in the sense that participants did not revisit texts until all texts had been visited once. All participants in the four timed and four un-timed conditions were systematic in this respect. Only three out of fifteen participants in the eight timed condition (all satisficers) and four out of fifteen in the eight un-timed condition (two samplers and two satisficers) ever revisited texts without having visited all other texts once first. This systematic behaviour is especially compelling in the eight text condition since it shows that most participants will go through each of the eight texts before deciding to revisit any of them. However, this does not imply that participants were using a sampling strategy, because participants were not making rapid visits to all available texts earlier in the experimental session. Rather, they were using a satisficing strategy by skimming over and rejecting unpopular texts and spending more time on the popular texts.

It is possible that the order in which they were searching the texts was a result of environmental factors, such as the menu order of the texts. Their time allocation strategy may be driven by such factors but the time they spent in the different texts and pages is more likely to be based on the strategy they were using.

Selectivity/Adaptivity

The extent to which participants were selective was variable. One index of selectivity, or adaptivity, is the amount of time participants spent reading their favourite text (that is, the one text in which they spent the greatest proportion of their time). Participants in the eight text condition spent significantly more time in their favourite text, relative to the number of texts available to them than did participants in the four text condition. Thus, despite having double the number of texts available to them, readers with eight texts still spent on average, across the timed and un-timed conditions, 38.5% of their time studying one particular text. Participants in the four text condition spent on average 42% of their time reading one particular text. With more available texts it was predicted that less time proportionally would be spent on the favourite text. However, this result was contrary to what was predicted. Although this will not be examined in this thesis, further investigation is necessary to ascertain why participants with more texts available still spend a similar proportion of their time reading one particular text.

However, participants did not devote a very large proportion of their time to any single text (four timed, 45.8%; eight timed, 39%; and four and eight un-timed both 38%). Furthermore, in the four text condition, most participants spent some time in all four texts (the mean number of texts visited in the four timed condition was 3.73 and in the four un-timed condition was 4). This tendency to consult most or all of the available texts suggests that participants were not willing to forgo information which may be contained in other texts, because in general it is probable that each text contains *some* unique information, albeit very little in some texts.

In terms of how adaptive readers were at allocating more time to easier texts when under time pressure, Table 2.3 showed that participants who were under time pressure spent more time reading the easy text than the difficult text in the four text condition. This analysis also showed that more time was spent reading the easy text for participants who were under time pressure than for participants who were not under time pressure. Reader and Payne (2007) also found that participants who were

under time pressure and reading with the goal of writing a summary essay favoured the easiest text. The results of both Reader and Payne and Experiment 1 are consistent with the idea that when people are under time pressure, they devote more time to the easier sources (Son & Metcalfe, 2000). However, these results were not fully replicated in the data for the eight text condition. In the eight text condition, there was no significant difference between the time participants spent on average on the two easy versus the two difficult texts when they were under time pressure. It would have been expected that more time would be spent on the easy texts. However, the data did show the tendency for participants under time pressure to spend more time reading the easy text than participants who were not under time pressure.

However, Son and Metcalfe (2000) also found that when studying for a test people spent more time on the more difficult items, an effect which was not replicated in Experiment 1. The high proportion of people in Experiment 1 who spent most of their time reading the easiest text also contradicts what Dunlosky and Hertzog (1998) would have predicted according to their discrepancy reduction model. This model states that people will spend more time reading materials with the largest discrepancy between the level of their existing knowledge and the level of the text, thus resulting in people spending more time on the more difficult texts. This does not happen in Experiment 1.

The reading goal in this experiment was very specific and it is possible that this had an effect on the strategy employed by the readers. Participants were instructed to 'read with a view to answering questions', and this may have influenced the way in which they searched for the information required. Learning for a test may engender a different reading strategy than learning in order to write an essay, as each of these goals requires a different type of knowledge. However, Reader and Payne (2007) gave instructions to participants to study with a view to writing a summary essay, and they also found an overwhelming majority used a satisficing strategy.

The findings from Experiment 1 also indicate that the length of time participants spent reading their 'favourite' text does not fit with what might be expected. The pattern of data regarding time spent on texts suggests that participants in the four and eight text conditions spent a similar proportion of their study time reading their 'favourite' text, despite the fact that participants in the eight text condition had double the number of texts to browse through if they desired. With more texts available, it was predicted that less time proportionally would be spent on

the favourite text. More studies (perhaps involving asking participants after the experiment which was their favourite text and why they chose to spend however long they spent reading it) would be necessary to determine why participants with more texts available would still spend the same proportion of their time reading one particular text.

Learning Score

As reported in Section 2.3, there was no significant effect of time available or number of texts on the learning score under the essay measure. This may be due to the 250 word limit set on the essay, or the essay assessment method used. In terms of the test measure, analyses revealed that participants learned more in the un-timed condition, so time available did affect their learning score. This, however, is what would be expected, as the participant had more time to digest the information in the texts available. The data regarding the number of points of improvement from pre- to post-test also support this conclusion.

There appeared to be no effect of reading strategy on the learning score, although there were not enough samplers to draw this conclusion with any confidence. Although it seems that the behaviour evident when readers were allocating time across multiple texts fits with one of satisficing, it did not necessarily result in more effective learning, according to the learning measures employed in the analysis of data from this experiment to compare learning scores of satisficers versus samplers.

Using Latent Semantic Analysis to grade the essays in Experiment 1 did not add much to the results, and effects of learning were found only using the test measure. Despite its useful and productive nature, LSA does carry with it a few limitations which may call into question its reliability as an accurate measure of learning. The representations induced by LSA are done through analysis of text alone. Thus, no knowledge is derived from perceptual information about the world, or from instinct, or from emotions and intentions, making its representation of reality "sterile and bloodless" (Landauer et al., 1998, p 261). There are several facts about the way in which LSA operates which could be deemed as limitations in its effectiveness. LSA is sensitive only to word usage, and therefore cannot detect logical errors and misconceptions (Wolfe et al., 1998). It also makes no use of word order, and therefore syntactic relations or morphology. Instead, it ignores how word

order produces meaning and captures only how the differences in word choice and in passage meaning are related to one another (Landauer et al., 1998). A simpler way of viewing this is that LSA represents a word meaning as an average of the meaning of all the passages in which that word appears, and a passage meaning as an average of all the words which are contained in that passage. It was therefore decided that in subsequent experiments the test measure alone would suffice to provide the information necessary on learning scores, and so the essay measure will not be used again.

General conclusions

Experiment 1 was inspired by the idea that readers are generally faced with more than one source of information when learning for a test or essay writing task, and therefore investigated how readers decided to share their time across sources of information (texts) when they were faced with too many texts to read in the time they had available. Experiment 1 focussed on two research questions, a) whether readers can adaptively allocate time when studying texts of varying difficulty under time pressure or not, and b) what type of strategies readers adopt in order to read under such conditions. Two strategies were outlined, satisficing, which involves setting a threshold of satisfaction and continuing to read the chosen text provided satisfaction does not fall below this threshold, and sampling, which involves evaluating the available texts prior to the study phase. Both strategies are adaptive, but results from this study indicated that satisficing was more common than sampling. That is, the behaviour of the participants was such that it was more frequently consistent with a satisficing model than a sampling model. However, in a sense, a reader must decide which strategy to use without knowing which would be best in the current situation (although this is not to suggest that readers consciously choose between the strategies). Employing a sampling strategy is likely to be better in the event that more than one of the texts meets the readers' satisficing threshold but one of these texts is better than the others. Although the texts used in Experiment 1 were set up in this way (on the basis of research by Wolfe et al. 1998), participants were not aware of such facts about the texts prior to the study phase. Therefore, participants had to 'make the decision' of how to allocate their time across these available texts

(behaviour which fits with sampling or satisficing) based on prior expectations and the extent to which the texts available supported a sampling style strategy.

Perhaps one explanation of why satisficing is more popular than sampling is that since sampling separates evaluation of texts and learning, it is only an effective strategy if readers are able to judge the value of the text quickly and accurately. If they feel that the first page is not representative of the entire text, then sampling seems less likely as their time would be better spent learning than evaluating texts. Similarly, participants are not likely to sample if they think that the information they are required to learn is scattered across all the texts.

Experiment 1 thus indicates that when faced with multiple texts to read and a time limit in which to do so, readers have a tendency to adopt a satisficing strategy. Furthermore, participants still consulted most pages of a text before rejecting it to move onto the next text. Although adaptivity has been considered in this experiment, it has been measured by comparing participants' scores on the pre-test and the amount of time they spent in what the LSA website classed as the easiest text or the hardest text of the four. One important issue which follows from these results is whether readers have the ability to be adaptive when reading, that is, are they able to allocate more time to the texts from which they will gain the most valuable information for their learning task? In order to investigate this, manipulation of the texts is necessary to ensure that some contain relevant information to the task and some do not.

In Experiment 1 the texts differed in terms of difficulty (and to some extent frequency of information targets). However if texts differed in terms of direct relevance to the task, then this would enable a more accurate measure of the level of adaptivity readers display under these conditions. Manipulating the texts, to make them differ in terms of relevance to the experimental task, also eliminates any problems associated with the texts originally used by Wolfe et al. (1998) and subsequently used in Experiment 1. Since the four texts (A, B, C, and D) used did differ in the number of answers to the test they provided, it is not completely certain that preference for one text over another in Experiment 1 is not attributable to this aspect of the texts rather than their difficulty level. Although, considering that the instructions were not to search for specific facts, it is argued that this issue with the texts is unlikely to have had any influence. Also, these original materials were used in Experiment 1 so that direct comparisons could be made regarding specific factors between this and previous research (Reader & Payne, 2007). In the next experiments,

Experiments 2 and 3, the texts are designed such that the difficulty level is kept constant, while their direct relevance to the test for which participants are studying is varied. Since people's browsing strategies are likely to be influenced by the direct relevance of the available texts, it is important to understand the strategies people use in order to identify the most relevant texts, so that this selectivity based on relevance can be maximised.

CHAPTER 3

3.1 Introduction

The following pair of studies was designed for two main reasons. The main reason for these studies was a more direct test of whether readers have the ability to be adaptive when reading, that is, whether they are able to allocate more study time to the texts from which they will gain the most valuable information. By manipulating the reading material it was possible to determine which texts contained the most valuable information for the task in hand. If people are required to learn for a specific task, then it is reasonable to assume that their choice of information source (text) will be influenced by the direct relevance of that source/text to the task in hand. If relevance is important, as it inevitably is in real world situations (people usually search for a specific reason, although they may at times leisurely browse for nothing in particular), then it seems necessary that information seekers are able to develop strategies to enable them to identify relevant sources to exploit. It follows then, that those strategies most often used will be those which identify target sources most efficiently and easily, with the least effort possible.

If it can be identified which strategies enable people to allocate attention to the desired information most quickly and easily, then it is plausible to suggest that these strategies could be taught to people as an effective way to read, or allocate attention, selectively. Experiments 2 and 3 offer a more direct test of adaptivity which was not possible in Experiment 1, which tested adaptivity by comparing the time participants spent on texts of varying difficulty with their knowledge about the topic, as measured by a pre-test. This method of testing adaptivity is, however, very indirect. Experiments 2 and 3 will test adaptivity in a more direct manner by altering the relevance of the texts to the task in hand and monitor the amount of time participants spend reading the relevant (termed 'good' in the following experiments) and non relevant (termed 'bad') texts. If they are displaying adaptive behaviour, then they will spend more time in the relevant (good) texts.

Another reason for the following experiments was to see whether or not the satisficing model applied with a more defined, specific reading goal. Experiment 1 found that readers' behaviour was consistent with a satisficing strategy when they were asked to learn for a test from either four or eight texts, and when either under

time pressure or not restricted by time. This strategy was more frequent than a sampling strategy, even when the number of texts available to the reader increased, the time available for studying was un-limited, and also the learning goal was more specific than had been used previously. Experiments 2 and 3 will examine whether the findings from Experiment 1 extend to a situation when the sole aim of reading is to study for a test to which readers have already been exposed, either by studying the questions or attempting to answer them. In Experiment 1, participants were studying 'with a view to answering more questions *like* the ones already answered', so they were not aware that the post reading questions would be exactly those encountered before reading. In Experiments 2 and 3, however, participants are explicitly told that they are learning for that specific test. This tests whether satisficing is still prevalent when the task is to learn *specific* facts, rather than more general learning (Experiment 1).

However, changing the task to one where the participants are aware of the information which they are required to learn, rather than merely reading to learn for an unknown test or essay, may have an effect on the way participants allocate their time across the texts. Due to the task becoming more like a search task, this may discourage a sampling strategy because if participants are searching for specific facts then they are more likely to search all the texts for these facts. The possibility that people are using a search model (as discussed in Chapter 1) will be addressed, and attempts will be made to fit the data to this model.

3.2 Experiment 2 – Adaptivity 1

3.2.1 Introduction

Experiment 2 was designed to investigate the extent to which readers are adaptive in their allocation of time across multiple texts, and to further explore the use of satisficing and sampling when allocating time across multiple texts under time pressure. Participants were presented with a set of questions on the topic of the human heart and told that they would be required to answer these questions after they had had a chance to study texts on this topic for twelve minutes. Four texts on the topic of the human heart were used, which were constructed from a variety of sources and adapted to ensure that the difficulty and readability of all texts were as equal as

possible. They were then manipulated so that two of the texts, classified as 'bad', provided one or two required facts per page, and the other two, classified as 'good', provided four required facts per page.

The learning goal in this study was more specific than in Experiment 1. Rather than being asked to learn for an unknown test, the present study required participants to learn specific facts (the answers to the questions in the pre-test). Previously, it was not possible to determine with great accuracy the adaptivity of readers, due to the fact that it was not possible to classify the texts as being an adaptive choice for the individual reader, or not, because the texts differed in difficulty level rather than in how useful they were for the test. That is, two of the four original texts in Experiment 1 did in fact answer more questions in the test than the other two, but participants were not told to study for a specific test so it was argued that this feature present in the materials had no bearing on the results of Experiment 1. However, in Experiment 2 an adaptive choice of text for all participants would be a good text.

The prediction was that participants would continue to read the 'good' texts and abandon the 'bad' texts as they realised that they were not providing them with enough of the facts they were required to learn. This is what would be predicted according to a satisficing model, whereby the participant constantly monitors the level of information gain from the text and compares it to their desired level of information gain. Texts are abandoned when the participant feels that they are no longer 'good enough'. In Experiment 1 the aim was to provide texts which all contained the same information but with differing difficulty levels. In the current study the texts were manipulated so that they did not contain the same information, thus enabling a more accurate method of studying the theory of satisficing as an adaptive method of selecting texts, and a more accurate measure of the readers' ability to be adaptive (in selecting appropriate texts). Thus, Experiment 2 looked at the browsing behaviour of readers (the dependent variable) when learning for a pre-seen test by measuring the time spent in each of the available texts.

3.2.2 Method

Participants

Twenty-four participants took part in the study, twenty-three of whom were Cardiff University students who were paid for their participation. The other

participant was an employee of Cardiff University, who also got paid for their participation. There were fifteen females and nine males. The mean age was 21.6 years, with a S.D. of 5.4 years and a range of 18 to 44 years. All the participants had a GCSE in Biology (or Double Science), and four of them also had an A level in Biology. Eleven participants were studying sciences, nine studying humanities, two studying arts degrees, and one studying social sciences.

Design and Materials

Experiment 2 was a within-participants design. The dependent variable was the browsing behaviour of the participants and the independent variable was the text quality (good or bad).

Four texts on the topic of the human heart were used in this experiment. These texts were constructed from the four original texts on the human heart used in Experiment 1, which were taken from the study conducted by Wolfe et al. (1998). Each text was restructured for Experiment 2 such that two of them could be classed as relatively bad texts and the other two classed as relatively good texts. The classification of the texts as 'good' and 'bad' does not relate in any way to the semantics or the structure of the text, and these labels are used solely to indicate whether the content was useful, or not, for the goal test. When text quality is discussed throughout this thesis it is referring to these 'good' and 'bad' texts in terms of how useful/relevant the content was for the participants with regard the end test. To be classed as a 'bad' text, the text was re-written so that only one or two of the required facts were available per page of text, as opposed to the 'good' texts where four required facts were provided per page of text. Manipulations were made so that both the bad texts provided 1 required fact per 305 words, and the good texts provided 1 fact per 88 and 79 words. All texts were four pages long and between 1258 and 1413 words.

These texts were then tested with the LSA (Latent Semantic Analysis) Intelligent Essay Assessor (http://lsa.colorado.edu) to ensure that the difficulty level was constant across texts. All of the four texts were similar in difficulty to either text B or text C of Experiment 1 (see Table 3.1). This was to avoid participants favouring one text over another due to differences in difficulty, and allowed the study to focus on differences in relevance to the reading goal. Readability assessments were also carried out on the four texts to ensure this was kept constant (see Table 3.1).

Readability measures are mainly based on the surface characteristics of the text, such as the number of words in a sentence and the number of letters or syllables in a word (for a reflection of word frequency). Two commonly used measures are the Flesch Reading Ease and the Flesch-Kincaid Grade Level. Both of these were used on the texts. The former produces a number between 0 and 100, with higher numbers indicating easier reading. The latter converts the Reading Ease Score to a U.S. grade-school level.

Table 3.1: Difficulty and readability measures of texts

TEXT	A	B	C	D
Most similar to text (from Exps 1&2):	B	B	B	C
Cosine:	0.99	0.91	0.89	0.82
Flesch Reading Ease	54.8	55.9	52.1	32.9
Flesch-Kincaid Grade level	10.5	9.2	10.4	12

The texts were accessed via a menu system at the left of the screen which was visible throughout the reading phase, and allowed participants to access any page of any text at any given time. This ensured that readers could choose freely which parts of the texts they wished to study. The menu consisted of buttons which were labelled with a fictitious author, and a title of 'The Heart', and page numbers 1 to 4. There was no other information given about the texts. Therefore, readers were unable to discriminate among texts according to their relevance.

The order in which the texts were accessed (from top to bottom on the menu bar) was varied. Eight different text orders were used, with a combination of blocked (good texts and bad texts presented next to each other) and mixed (good texts and bad texts alternating). The orders used were: ADBC, BCAD, DACB, CBDA, ABDC, BACD, DCAB, CDBA. The eight chosen combinations were arrived at following a simple rule: As texts A and D are 'bad' and texts B and C are 'good', then if text A precedes text D, then text B would precede text C, and if text D precedes text A then text C would precede text B. This rule ensures that there is no order effect of the presentation of the texts, and since the differences between the two good texts themselves and between the two bad texts themselves are minimal, then this arrangement was sufficient to avoid such effects.

The test used in Experiment 2 was the one used in the Wolfe et al. (1998) study and consisted of 17 short answer questions on the topic of the human heart (see Section 2.2 for sample questions and Appendix C for the full test). Points were assigned to each question depending on the information required for the answer. There was a maximum of 40 points available.

Procedure

Participants were given five minutes to study 17 questions on the human heart. These were presented to them over two pages on a computer screen and a digital ascending timer was visible for them to keep track of their five minutes study time. Participants were asked not to answer the questions yet, but were advised that they would be given the identical questions after the reading stage of the experiment.

Once the five minutes had passed, the screen went blank and participants were instructed to proceed to the next stage of the experiment, where they were given twelve minutes to study for the test. Twelve minutes was used as it was pro rata the same as Experiment 1.

Participants had access to four texts on the human heart and were advised that these texts had been selected from various internet sites. Therefore, they were told that it was possible that all of the texts were useful for the current task, but that it was equally plausible that some of the texts would be more useful and relevant than others. They were advised to bear this in mind and allocate their time across the texts in whichever way they thought suitable for the current task. They were advised that they were not *required* to read all of the texts, that they may read one, some, or all of them, provided they were using their time productively given the task requirements.

A time stamp was taken when participants moved pages so that it would be possible to identify patterns in reader behaviour which would inform about reading strategies and the allocation of time when reading multiple texts. Again, there was a digital ascending timer visible on the screen to inform participants about how long they had been studying. Once the twelve minutes had passed, participants were asked to proceed to the next stage of the experiment, which was the test. There was no time limit for completion of the test, which was also used in Experiment 1 (see Section 2.2), and participants typed the answers to the questions in the boxes provided. Once they had completed these questions, participants progressed to a screen where they were asked to answer three questions about their educational background (i.e.,

whether they had GCSE and A level biology, and what degree course they were following).

3.2.3 Results

Adaptivity

This section will analyse the data to examine whether participants were being adaptive in their reading, that is, did they manage to allocate more of their study time to the texts from which they would most benefit, given the task goal of reading for a test. Examining the time participants spent on the good and bad texts will give an indication of whether they were adaptively allocating their time across the texts. Participants spent more time reading the good texts (mean = 510.7 second, *S.D. 175.79*) than they did the bad texts (mean = 203.5 seconds, *S.D. 174.25*); $t(23) = 4.31$, $p < .001$.

It should be pointed out here that there are different patterns to look for as an index of adaptivity. One is the amount of time spent in texts overall, where more time in the good texts will indicate that the participant is behaving adaptively. Another factor which is not independent of the first, but can be measured separately, is the favourite text of the participant (the single text in which most time is spent). If this is a good text then this would indicate adaptive behaviour. These factors are partially dependent on one another but can be tested separately.

Most of the participants were adaptive according to both the overall time in texts and the favourite text measures, but two participants were only adaptive according to one of these measures (the overall time in texts measure). Since they are both measures of adaptivity, just displaying different patterns, then it is argued that these two participants would be considered adaptive. However, there were four participants classed as satisficers who were not adaptive according to either of the discussed measures. Such maladaptive behaviour could be a result of memory problems encountered during the experiment. The post-reading test relied on participants remembering the questions for which they were learning the answers. There is plenty of room for recognition judgements to be different, and searching for answers to a test is not like searching trees for fruit, as the potential to misrecognise items is more prevalent. Participants may come across a required fact but not

remember that it is a target item, or they may even misrecognise a fact as being one which is required when it is in fact not.

For seventeen participants (70.8%) their favourite text (single text they spent most time reading) was a good text. The finding that for eight participants the favourite text was good text B and for nine participants it was good text C argues against any particular advantage either text may have had over the other. For seven participants their favourite text was a bad text, and overall, participants spent around 60% of their time in their favourite text. Nineteen participants (79.1%) spent more time in the two good texts, and five participants spent more time in the two bad texts (binomial test $p < .01$).

The satisficing model makes predictions about the length of first visits to texts, in that the first visit to each text will be the longest visit made to that text. Therefore, looking at the length of first visits to texts for participants depending on whether they read a good or a bad text first will help to determine both their browsing behaviour and the degree to which they were reading selectively. Since more predictions can be made about the satisficing strategy than any other described in this thesis, then the following analyses will be carried out first looking at the satisficers only, and then, for completeness, looking at all participants together.

An analysis was carried out to investigate the length of the very first text visit for readers who viewed a good text first compared to participants who viewed a bad text first. Table 3.2 shows that the mean time spent on the first visit to the first text was higher when the first text presented in the menu bar was a good text than when it was a bad text. An independent t test confirmed that the difference was significant, $t(16) = 3.23, p < .005$. Similarly, a within-participant comparison of the duration of all first visits (34 separate time values in total) revealed that the duration of first visits to good texts was significantly higher than the duration of first visits to bad texts (see Table 3.3): $t(33) = 3.37, p < .005$.

Table 3.2: Mean time spent on first text visits for good versus bad texts presented first on the menu bar (satisficers)

	Text	
	good	**bad**
Mean	424.97	214.00
S.D.	*129.43*	*143.52*

Table 3.3: Mean time spent on *all* first visits to good and bad texts (satisficers)

	Text	
	good	**bad**
Mean	237.71	110.46
S.D.	*158.28*	*110.88*

People who displayed satisficing behaviour were being adaptive not only in terms of the first text they visited, but throughout their reading, and spent more time in a good text the first time they visited it than in a bad text the first time they visited it. Participants were continuing to read, as opposed to abandoning the text in favour of reading elsewhere.

There is evidence for adaptivity whatever strategy was used, as all three participants displaying sampling behaviour were adaptive (according to both tests for adaptivity), and two out of three of the residual participants were also adaptive. The pattern of significance for time spent on first visits to texts depending on whether a good or a bad text was read first is maintained for all participants, even though the predictions were for satisficing behaviour. The analysis comparing the length of very first visits to first texts depending on whether a good text (mean 362.46, *SD* 178.75) or a bad text (mean 189.44, *SD* 147.52) was viewed first was also carried out including all the participants (thus, those who were classed as samplers and residuals as well as those classed as satisficers) and revealed a significant result, $t(22) = 2.6, p < .05$. Similarly, the analysis looking at the length of all first visits to good texts (mean 207.59, *SD* 170.28) versus bad texts (mean 92.06, *SD* 108.17) was carried out for all participants and this also revealed a significant result, $t(46) = 3.61, p < .01$. In six out of the seven participants who were not adaptive, their favourite text (a bad text) was the text presented to them in the top position on the menu bar, and for the other one it was presented in second position.

Time Allocation Strategy

All browsing behaviour in this thesis will be determined using the definitions laid out in Section 2.3. Out of 24 participants, 18 exhibited satisficing behaviour (75%). That is, for 18 participants, the first visit to each and every text was the

longest visit they made to that text. Three participants exhibited sampling behaviour, by visiting all the texts briefly at the beginning of the reading phase before settling on a preferred text (this may be the fourth text they visit and never leave). Finally, three participants did not fit into either of these categories, and so were classed as residuals (12.5%).

The three participants in the residual category displayed behaviour which was very similar to the satisficing model, and in some cases *almost* fitted with the categorisation for the satisficing strategy. For one of the residuals, all of their first visits to texts were the longest visits they made to those texts with the exception of text C whereby the first visit was 62.1 seconds and a later visit was 75.6 seconds. For another residual participant, 99% of their time was spent on the first text they visited, but they briefly visited (for 4.4 seconds) another text in the middle of their reading (after 156.3 seconds) of this first text. This also appears to be a kind of satisficing strategy, but with a slight deviation to check out another text briefly. The third residual also had the first visits as longest visits except for one text which was visited for 1.8 seconds initially and then revisited later for 187 seconds. It could be argued, however, that brief visits of a few seconds, as shown by two of these residual participants, may not be enough to make a reliable judgement of the quality or usefulness of the text, and that during these few seconds the participants may not even have been paying attention to the text. Therefore, if we ignore a few seconds here and there (assuming perhaps that they could be clicking errors), it could be argued that these residuals are displaying a satisficing strategy. In any case, the satisficing criteria is characterised by an arbitrary set of rules, and although this serves to classify most participants, it is plausible to assume that some will behave slightly out of the range of the specified rule.

As shown in Experiment 1, it was found that browsing behaviour was exhaustive, that is, readers generally consulted all available texts and did not tend to revisit texts already viewed. Eighteen out of twenty-four participants visited all of the texts available, but only eight out of twenty-four participants (33.3%) visited all the available pages. It is suggested that not revisiting texts until all of the available texts had been visited at least once first would indicate that participants were systematic in their search for an appropriate text. Nineteen out of twenty-four participants (79.2%) did *not* revisit texts before having visited all the available texts at least once first. Most participants consulted the texts in the order they appeared on the menu list and

worked their way down the list. This is what would be expected from participants. However, these results differ from Experiment 1, where far fewer participants used the menu bar as a guide to the order in which they would visit the texts. It is not known why participants in Experiment 1 did not visit the texts in the order they appeared on the menu bar, but it s possible that it is related to the learning goal of the experiment, as participants in Experiment 1 were learning for an un-known test as opposed to participants in Experiment 2 who were learning for a pre-seen test. This is just speculative and a full explanation for why there is a difference across the experiments is not known.

Examining the visits to texts and pages made by the participants will inform us about the type of strategy they are using, as a sampling strategy would be indicated by initial brief visits to all texts, and a satisficing strategy would show longer visits to texts and not necessarily all available texts being consulted (if a preferred text is opened first the participant may spend most of their time in that and only visit one other text). Table 3.4 shows how participants allocated their time amongst the texts available to them. It includes the means for duration of each text and page visit (in seconds), the number of unique pages and texts visited by the participant, the percentage of time they spent in their 'favourite' text, and the total time spent in the two good texts and the two bad texts (in seconds).

Table 3.4: Mean number of texts/pages visited and mean time per text/page

		Number of text visits	Number of page visits	Unique pages visited	Unique texts visited	% time in favourite text	mean time per text	mean time per page
GOOD	Mean	**3.4**	**11.5**	**6.5**	**1.8**	**65.9**	**289.2**	**76.1**
	SD	*1.4*	*4.3*	*1.6*	*0.4*	*18.0*	*117.9*	*137.2*
BAD	Mean	**2.9**	**7.5**	**5.2**	**1.9**	**45.6**	**103.9**	**33.3**
	SD	*1.7*	*6.8*	*2.8*	*0.3*	*14.7*	*86.3*	*29.0*

As would be expected following the finding that participants spent more time overall in the good texts, Table 3.4 shows that more time is spent per text visit in the good texts than in the bad texts, $t(23) = 4.81, p < .001$. Also, participants made more visits to pages of good texts than they did to pages of bad texts, $t(23) = 2.85, p < .01$.

In terms of how many pages per text were visited for each text type, good and bad, on average across participants, there were 2.80 pages visited out of 4 in the bad

texts (*S.D. 1.36*) and 3.48 pages visited out of 4 in the good texts (*S.D. 0.98*). In order to identify when participants were making their stay/leave decisions with regards the quality of the texts, an analysis was carried out examining the time participants spent over the four pages of each text type, good and bad. Since participants did not always visit *all* the pages of a text, the time values for the pages they did not visit were substituted with the mean of all of the other times spent in that page (1, 2, 3, or 4) for that text type during the experiment. This was to ensure that the overall mean for each page represented actual time that the participants spent reading that page, rather than putting zero time which would lower the mean, suggesting that the time spent in that page was less. The data also excludes pages of texts which participants were reading when their time expired, as this would not be a true reflection on how they would have spent their time had the time not run out. This will be applied in all subsequent experiments for missing values of page time data.

The data showed that overall, participants spent more time in each of the pages of the good texts than the bad texts, and there were few differences between the times spent in pages within a text type, good or bad (see Figure 3.1). An ANOVA with factors of page (page 1, 2, 3, and 4) and text type (good and bad) confirmed these findings by revealing that there was no main effect of page ($F(3,234) < 1$), a main effect of text type ($F(1,78) = 68.10, p < .001$), and an interaction between page number and text type ($F(3,234) = 6.50, p < .001$). Simple effects revealed that at each level of page, readers spent longer reading the good texts than the bad texts (smallest $F(1,78) = 24.5$, all $p < .001$). However, there was an exception between the good and bad texts for page 1, where there was no significant difference in time spent ($F(1,78) = 2.6$). With regard between-page differences within a text type, for the bad texts and the good texts, the time was significantly different between pages 1 and 3 only (bad texts, $F(3,76) = 5.12, p < .01$; good texts, $F(3,76) = 3.14, p < .05$).

Due to the fixed study time in these experiments this analysis may violate the independence assumption, and therefore an extra analysis was carried out on the time on pages 1 to 4 of *one* good and *one* bad text visited (the first good/bad text visited). The pattern of results was the same (no main effect of page ($F < 1$), a main effect of text type ($F(1,46) = 30.03, p < .001$) and an interaction between page and text type ($F(1,138) = 3.32, p < .05$).

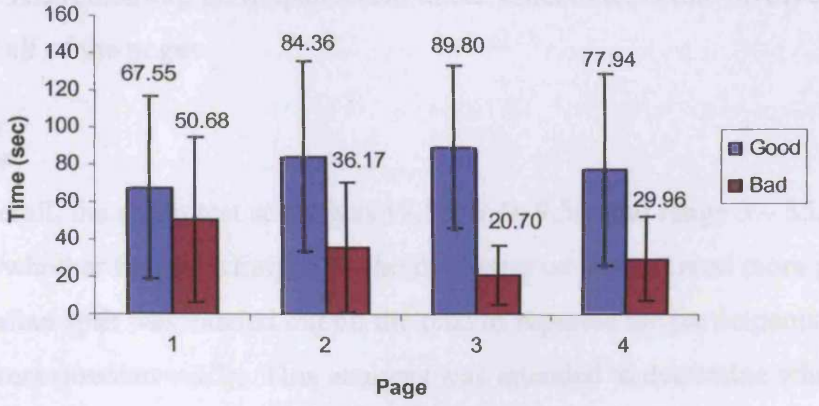


Figure 3.1 The mean time in seconds (with +/- S.D.) participants spent reading pages 1 to 4 in both the good and the bad texts in Experiment 2 (for pages that were visited).

If participants enter a page of the text, then overall they spend similar amounts of time reading that page, except in the case of page 3 where it is significantly higher than page 1 for the good texts and significantly lower than page 1 for the bad texts. For pages 1 to 3 for the good texts, the pattern of data indicates that there may exist some kind of dynamic criterion. If readers are getting the information they require from the good texts, then they may adjust the criterion of information gain as they progress through the pages, which would result in them spending more time on later pages of good texts. Similarly, with the bad texts, if participants are not getting the information they require, they may dynamically adjust the criteria resulting in them spending less time on later pages. This would help to argue for a satisficing strategy with a dynamic criterion, and suggests models which predict the same pattern of behaviour as satisficing but without a criterion level would not fit this data.

However, sometimes people leave documents early and do not visit the later pages of a text. In terms of how many pages participants visited, for the good texts, 4 participants made a 1-page visit, 3 participants made a 2-page visit, 5 participants made a 3-page visit, and 21 participants made a 4-page visit. For the bad texts, 7 participants made a 1-page visit, 4 participants made a 2-page visit, 4 participants made a 3-page visit, and 13 participants made a 4-page visit. Eleven participants made only 4-page visits to good texts and ten participants made only 4-page visits to bad texts. 87.5% of participants visit all four pages of at least one of the good texts they visit, and 54.2% of participants visit all four pages of at least one of the bad texts

they visit. The remaining participants also make visits to texts that involve viewing fewer than all of the pages.

Test scores

Overall, the mean test score was 18.83, *S.D.* 9.56, and range 3 – 35. In order to identify whether those participants who did better on the test read more good or bad texts, a median split was carried out on the data to separate the participants into high or low scorers (median = 17). This analysis was intended to determine whether adaptive reading led to better test scores. It was revealed that in the low score group for five participants out of thirteen (38.5%) the favourite text was a bad text and four participants spent more time in bad texts (30.8%). In the high score group, for two out of eleven (18%) the favourite text was a bad text, and one of these spent more time in the bad texts (9%). A chi square test showed that these differences were not significant (frequency of favourite text: $\chi 2 = 1.19$, *n.s.*, and frequency of more time in good/bad texts: $\chi 2 = 1.70$, *n.s.*). A correlation was carried out to determine whether the post- test score was in any way related to the time participants spent in the good texts. The results showed that time in good texts did not predict a better score in the test ($r = .142$, N = 24, $p = .507$, 2 tailed).

3.2.4 Discussion

The behaviour of three quarters of the participants who took part in this study was consistent with a satisficing strategy. The fact that nearly 80% of the participants spent most of their reading time in the 'good' texts as opposed to the 'bad' texts indicates that readers were being adaptive in the way in which they allocate their time amongst the texts. Further evidence that is consistent with the view that it was the classification of the texts as 'good' which caused the participants to be spending more time in them (and not any other factor e.g., the style of the text) is the fact that there was no difference in the number of participants who preferred the first good text (eight readers) and the second good text (nine readers). Twelve of the seventeen participants who preferred a good text preferred the first good text they encountered and the other five participants preferred the second good text they encountered (the term 'preferred' being used to mean their favourite text – single text in which they spent most of their time).

The search model could also account for the participants spending more time on the good texts than the bad texts, as they would pause to learn more often where there were more target items (i.e., in the good texts). However, there are aspects of the data which cannot be fully explained by the search model, but can be explained by the satisficing model, and these will be discussed in Section 3.4.

The time participants spent across the four pages of a text within the good or the bad text categories were not significantly different except for page 3, which was significantly higher in the good texts than page 1 and significantly lower in the bad texts than page 1. As expected, participants spent more time overall on each of the pages of a good text than a bad text apart from on page 1 where there was no significant difference in time spent reading it for the good versus the bad texts. One reason why the time on page 1 of texts is not different across the good and bad texts may be due to the fact that all participants set out with the same expectations for each of the texts, since although they were advised that the texts may differ in quality, they have no a priori information about which texts are better. The finding that participants visited each page within a text type for approximately the same amount of time also suggests that they were not prepared to judge the value of later pages on the basis of the earlier pages. However, this is only the case if the pages are actually visited, since some participants do not visit the later pages of a text, especially the bad texts.

A comparison of the length of time participants spent in the text viewed first (i.e., in the top position on the menu bar) revealed that if this text was a good text they would spend significantly more time reading it than if it were a bad text. So, the reader began to read the text in the top position and continued to read this text if it was providing them with a satisfactory rate of return, or abandoned reading it in favour of reading elsewhere if the information gain fell below their level of satisfaction. This is further evidence of the adaptivity, or selectivity, of readers in allocating their time only to texts that are deemed 'good enough'.

To focus on the participants classified as satisficers, an analysis was carried out on the length of time spent on the first visit to each text for the good and bad texts. Tests revealed that there was a significant difference in the amount of time readers spent on first visits to the good versus the bad texts. This suggests that they were being adaptive and staying in the text if the rate of return was good, or abandoning it in favour of reading elsewhere if the rate of return was not satisfactory. This

strengthens the adaptivity theory as it shows that satisficers were being adaptive throughout their reading and spending more time in the good texts the first time they visit them than they do in the bad texts the first time they visit them. Since it was found that all of the three samplers, and two out of three of the participants classed as residuals, were also exhibiting adaptive behaviour, it is consistent with the idea that adaptivity (as measured by more time spent in the good texts or a good text as the favourite text) was evident whatever strategy was used.

However, it was observed that a third of the satisficers were not being adaptive in their time allocation due to the fact that either their favourite text was a bad one, or they spent more time overall in the bad texts than they did in the good texts. One reason for this maladaptive behaviour, as mentioned earlier, may be due to the participants not remembering the test questions properly and perhaps misrecognising facts in the text as facts that they needed to learn for the test. As discussed, there is plenty of room for recognition judgements to be different, and this seemingly maladaptive behaviour may be a result of memory problems for the test questions. Another possible reason for this is the positioning of the readers' favourite text on the menu bar. Almost all of the satisficers who were not showing adaptive behaviour chose to spend most of their time in the first text viewed, which happened to be a bad text. Since most participants visited the text in top position first, this may account for some of them spending more time in a 'bad' text. However, under the satisficing model, readers should abandon reading this 'bad' text despite its spatial location on the menu bar. The fact that they do not abandon it and continue reading suggests that their threshold level with regard rate of return of relevant information may be much lower than it was assumed to be here (in order to make predictions, the satisficing model assumes a moderate threshold level). Another possible explanation for why participants may spend more time in bad texts is that they may be expecting a 'required fact' to appear if they keep reading. In this way, rather than moving to the next available option in the hope of finding facts there, they may be working on the premise that if they keep reading then some useful information will emerge soon. This is purely speculation.

If participants were following a patch-leaving rule that stated they would remain in the text until a certain level of information gain was achieved, then this could account for participants staying longer in bad texts. Similarly, if the threshold of information gain is relatively low, then participants will stay longer in the bad texts

because they would be working to the rule that they should leave the text only when information intake has dropped below this already low threshold level. If they are working alongside a patch-leaving rule centred round time, then they may stay in a text for *t* seconds before leaving, regardless of the information yielding properties of the patch. The fact that most participants managed to spend more time in the good as opposed to the bad texts suggests that most participants were not following a patch-leaving rule based on time or pure gain, but rather working with a rule stating that they should leave if gain is less than a certain threshold after a certain amount of time. The few participants who do not spend more time in the good texts may well be following an alternative assessment rule.

In terms of the test scores, there was no significant correlation between the time that participants spent on the good texts and their post-test scores. Although readers were adaptive, it appears that this had no bearing on how well they performed in the test. This is contrary to what would be expected, since if the participants read the good texts surely they would perform better on the test than if they had read the bad texts? However, the level of knowledge of the participants before the reading phase was not measured, so it is impossible to determine how much participants learned during reading, and therefore whether the participants who spent more time in the good texts actually did better in the test relative to their previous knowledge level. In order to ascertain whether adaptivity is at all related to learning it is necessary to incorporate a pre-test measure within the design of the study. This is addressed in Experiment 3.

Overall, the data from Experiment 2 shows that most people were satisficing. The pattern of reading durations is consistent with the possibility they were remaining in a text as long as they were being rewarded in terms of encounters with the information they needed for the test. It has also shown that people were being selective in the way in which they were allocating their time to texts.

Two readability measures and an LSA analysis were carried out in order to ensure that all of the texts were as equal as possible in terms of their difficulty level, overall length, number of words per page, and style. Despite these precautions, it is difficult to claim with absolute certainty that the 'good' texts were favoured because of their content and not for any other reason, such as them being easier to read, or more interesting to the reader. This is addressed in Experiment 3.

3.3 Experiment 3 – Adaptivity 2

3.3.1 Introduction

Experiment 3 was designed to complement Experiment 2, and deal with the uncertainty surrounding the issue of the classification of the texts as 'good' and 'bad'. As discussed previously, it cannot be claimed with any certainty that the preference for the 'good' texts evident in Experiment 2 is solely on account of their content and not a result of these texts being easier to read or more interesting to the reader. Despite conducting readability and LSA measures to eliminate this doubt, I decided that some refinements were necessary in order to strengthen this claim. The most reliable way to ensure that it is the content of the texts which attracts the reader and nothing else would be to change the short answer test so that the information the readers are seeking is more available in the texts previously classed as 'bad'. Therefore, the previously 'bad' texts (in Experiment 2) would now be the 'good' texts (in Experiment 3) and vice versa. This would address the issue of whether texts were favoured for any reason other than for their content.

In addition to changing the short answer test, Experiment 3 was designed to address the issue of learning, which was not possible in Experiment 2. Adaptivity, as measured by the overall time spent in texts and the quality of the favourite text, was evident in Experiment 2 but this did not seem to have any effect on the participants' test scores. A participant who spent more time in the good texts did not necessarily have better post-reading test scores than those who spent most of their time in the bad texts. However, the post-test score alone is not an indication of how much each participant actually learned during reading; a participant may have a lower final test score than another but still have learned more during reading, depending on their previous level of knowledge on the topic in question. No measure of prior knowledge was taken, so computation of learning scores was not possible. Experiment 3, however, required the participants to take the test both before and after reading, thus enabling a learning score to be calculated and used to assess whether reading adaptively enhanced learning, or not. However, changing one of the experimental tasks may have some other effect on the behavioural outcome of the study. If participants are required to complete the test, rather than simply study the questions,

then this changes the dynamics of the experimental goal and could perhaps have an alternative effect on reading strategy or adaptivity.

3.3.2 Method

Participants

Thirty-two participants took part in this experiment, 11 of whom were male and 21 of whom were female. The mean age was 21.6 years old, with a standard deviation of 1.52, and a range of 18-25. All participants were students at Cardiff University, and all participants were paid for their participation. All participants had a GCSE in biology (either in double science, single science or straight biology) and only 4 had an A level in biology. Sixteen participants were psychology students and sixteen were from various departments across the university.

Design and Materials

The design of Experiment 3 was similar to that of Experiment 2, with an added dependent variable of learning score, which was measured by comparing the pre- and post-test scores and calculated as it was in Experiment 1 (see Section 2.3 for an explanation of how the learning score was calculated).

The texts used for Experiment 3 were exactly those used in Experiment 2. Again, they were accessed by a menu system at the side of the screen, and the same eight variations of text order were used (see Section 3.2.2 for details).

Participants were given a test to complete both before and after the reading phase. The new test comprised 17 questions (the same number as in the original test); see Appendix G for the list of test questions. In order to make the previously good texts bad and the bad texts good, the questions were based on the information contained in the two previously 'bad' texts, thus making them 'good' for the purposes of the new test. A 'good' text in this experiment provided 8 or 9 answers out of 17 in total, and the 'bad' text provided only 1 or 2 of these answers. In terms of how many answers occurred per number of words in each document, the good texts provided 1 fact per 182.9 and 174.37 words, and the bad texts only provided 1 fact per 839.3 and 838.6 words. There are differences in the good text/bad text distinctions that were present in Experiment 2 and those that are present here. Firstly, the two good texts in Experiment 2 answered more or less the same questions. The good texts in

Experiment 3 answer different questions, making the reading of the second good text beneficial for a different reason than in Experiment 2. It would be advantageous now due to the reader finding more required information in the second good text, whereas it was advantageous before because reading the second good text was reinforcing required information (rather than providing new information). Secondly, the difference between the good and bad texts is more pronounced in Experiment 3 than it was in Experiment 2. The good texts were previously only 3.2 times better in terms of the information they provided relevant to the task at hand than the bad texts, whereas now the good texts are 6 times better than the bad texts (i.e., answer 6 times more questions). The set-up of the texts during the experiment and the text properties are exactly the same as Experiment 2 (see Section 3.2.2).

Procedure

Participants were asked to complete the test on the human heart and told that they would be required to answer the same questions later in the experiment, after the reading phase. There was no time limit for the completion of the test. Once the test had been completed, participants moved onto the reading phase. The procedures for the reading phase and the post-reading test were exactly the same as in Experiment 2 (see Section 3.2.2 for details). Once again, after the final test, participants moved onto a screen where they were asked to answer three questions about their educational background (whether they had GCSE and A level biology, and what degree course they were following), which ended the experiment.

3.3.3 Results

This section will outline the important findings from this experiment in the same order as in Experiment 2: the browsing behaviour of the participants and whether their browsing was systematic, the time participants spent on the good and bad texts and the pages of these texts, and finally the learning scores for readers who spent more time in the good versus the bad texts.

Adaptivity

Comparing the time participants spent in the good and the bad texts will provide an indication of how adaptive they were in their allocation of time across

texts. Participants spent more time reading the good texts (mean = 426.6, *S.D. 112.6*) than they did the bad texts (mean = 285.0, *S.D. 111.3*); $t(31) = 3.58$, $p < .01$.

Only six participants out of thirty-two failed to show adaptive behaviour according to the two measures used here (i.e., more time in the good texts, and a good text as the favourite text). Four of these participants were classed as satisficers, the other two as residuals. Three additional participants were adaptive according to one of these measures. All other participants (71.9%) were adaptive according to both overall time in texts and favourite text measures. As discussed in Section 3.2.3, lack of adaptivity in participants could be due to errors in recognition judgements.

For twenty-five participants out of thirty-two (78%) their favourite text was a good text, and the fact that good text A was favourite for fifteen participants and good text B was favourite for eleven participants argues against any particular advantage either text may have had over the other. Seven participants had a bad text as their favourite text. Overall, participants spent about 44% of their time in their favourite text. Twenty-four participants (75%) spent more of their time reading the 'good' texts than the 'bad', and the remainder (25%) spent more time reading the 'bad' texts (this difference is significant according to binomial distribution, $p < .01$).

As discussed in Section 3.2.3, the satisficing model makes predictions about the length of first visits to texts in that the first visit to each text will be the longest visit made to that text. For this reason, the following analysis looks at the lengths of first visits to texts when participants visit a good and a bad text first. This will help to determine whether participants are adaptive and also give an indication as to the extent of their selectivity. Since the predictions made here specifically concern a satisficing strategy, the following analyses will first be conducted on the satisficers alone, and then for completeness look at all the participants in the experiment.

An analysis was carried out to investigate the length of the very first text visit for readers who viewed a good text first compared to readers who viewed a bad text first. The data from one of the participants who was classed as a satisficer had to be excluded due to the fact that they did not read the text in the top position first (they read the last text first). This participant had been presented with a bad text first, but the time they spent on this particular text was not included in our analysis here, because it was not the text which they visited first. Therefore, the analysis was done on 23 satisficers – 11 who were presented with a good text in top position, and 12 who were presented with a bad text in the top position. Table 3.5 shows that the mean

time spent on the first visit to the first text was higher when the first text presented was a good text than when it was a bad text ($t(21) = 2.53$, $p < .05$).

Table 3.5: Mean time spent on first text visits for good and bad texts presented first on the menu bar (satisficers)

	Text	
	good	**bad**
Mean	287.31	174.63
S.D.	*110.65*	*103.05*

A comparison of the duration of *all* the first visits to each of the four texts was made to find out whether the first visits to the good texts were longer than those to the bad texts, as would be expected if the participants were being adaptive in their choice of texts. Table 3.6 shows that the mean time spent on first visits to good texts was higher than the time spent on first visits to bad texts throughout the reading time ($t(63) = 3.09$, $p < .01$). This analysis, together with an analysis comparing the first visits to texts 2, 3, and 4, so, excluding the first text (good: mean 167.31, *SD* 104.39; bad: mean 111.57, *SD* 78.57, $t(51) = 2.91$, $p < .01$) shows that participants were not only being adaptive in terms of their visit to the first presented text, but throughout their reading, as they spent longer on the first visits to the good texts than on first visits to the bad texts.

Table 3.6: Mean time spent on all first visits to good and bad texts (satisficers)

	Text	
	good	**bad**
Mean	185.32	124.25
S.D.	*115.73*	*84.36*

As explained earlier (Section 3.2.3), there are different patterns which indicate that a participant has been adaptive in their reading. These are the overall time spent in the good texts and the quality of the favourite text (single text in which most time is spent). In this experiment there were six participants who did not spend more time in the good texts overall *and* did not have a good text as their favourite text, and therefore seemed to show maladaptive behaviour. Two of these participants were under the residual category and four were under the satisficing category. Again, this maladaptive behaviour could be a result of participants not recognising target facts in

the texts as being required for the test, or not recognising target facts when they read them due to problems remembering the test questions. Adaptivity was evident for all strategies used.

Time Allocation Strategy

Time allocation strategy was measured by examining the time participants spent in the various texts and the manner in which they visited these texts and pages. Participants were classified according to the definitions used previously (see Section 2.3). Out of 32 participants, there were 24 (75%) who showed behaviour which was consistent with the satisficing model, so the first visit to each text was the longest visit made to that text for 75% of participants. 1 participant (3%) showed evidence of a sampling strategy, so having visited all texts briefly before settling on a preferred text. Finally, 7 participants (22%) were classified as residuals because they did not fit into either of the other two categories.

There were a relatively high number of residuals in this experiment (22%) and so the behaviour of these participants was explored in more depth. On closer analysis of the data, it transpired that one of the residuals displayed sampling behaviour, and visited all texts briefly within the first 283 seconds of the experiment before choosing and settling on one. This participant was not grouped under the samplers because in order to qualify as a sampler the participant must visit all texts in the first third of the experimental time (240 seconds). This rule for sampling is arbitrary and this participant is still arguably exhibiting sampling style behaviour. The other six residuals appear to be behaving in a satisficing manner, with the exception of later visits being slightly longer at times than earlier visits, thus resulting in them not being classified as satisficers under the satisficing criteria. Nevertheless, all six of these participants satisficed with all but one text, and even then the later visits to these texts were as little as 21 seconds longer than the earlier first visits (the maximum length difference between first visits and second visits was 69 seconds).

As in Experiment 2, the extent to which browsing behaviour was exhaustive was analysed. Three participants (9%) did *not* visit all the available texts, and twelve participants (37.5%) did visit all the available pages (i.e., all 16 pages). Twenty-six participants (81.3%) did *not* revisit texts before having visited all the available texts at least once first. In terms of the order of texts visited, six participants out of thirty-two

did *not* consult the texts in the order in which they were presented on the menu bar, but the remaining participants did follow the menu bar order when visiting texts.

The time participants spent in the individual texts and pages of the texts gives an indication of the strategy they used to allocate their time across the texts. Table 3.7 shows how participants spent their reading time with regard to the number of texts or pages visited, and how long they spent reading the good and bad texts (in seconds).

Table 3.7: Mean number of texts/pages visited and mean time per text/page

		Number of text visits	Number of page visits	Unique pages visited	Unique texts visited	% time in fav text	mean time per text	mean time per page
GOOD	Mean	**3.5**	**10.2**	**6.9**	**2.0**	**43.6**	**218.5**	**52.2**
	SD	*1.8*	*6.2*	*1.6*	*0.2*	*7.2*	*59.2*	*22.8*
BAD	Mean	**3.0**	**8.8**	**6.6**	**1.9**	**43.2**	**147.1**	**35.4**
	SD	*1.5*	*4.3*	*1.6*	*0.2*	*6.3*	*58.1*	*17.9*

Again, following the finding that participants spent more time in the good texts, Table 3.7 shows that more time is spent per text visit in the good texts than the bad texts ($t(31) = 3.62, p < .01$), and also more time is spent per page visit for the good texts ($t(31) = 4.08, p < .01$). As in Experiment 2, participants also made more visits to pages of the good texts than they did to the pages of the bad texts ($t(31) = 2.06, p < .05$).

In terms of how many pages per text were visited for each text type, good and bad, on average across participants, there were 3.31 pages visited out of 4 in the bad texts (*S.D. 1.10*) and 3.47 pages visited out of 4 in the good texts (*S.D. 1.04*). As in Experiment 2, in order to identify when participants were making their stay/leave decisions with regards the usefulness of the texts, an analysis was carried out addressing the time participants spent over the four pages of each text type, good and bad (see Figure 3.2). The data showed that overall, participants spent more time in each of the pages of the good texts than the bad texts, and there were few differences between the times spent in pages within a text type (good or bad), although a difference does appear between pages 1 and 3 for the good texts.

An ANOVA with factors of page (page 1, 2, 3, and 4) and text type (good and bad) confirmed this description of Figure 3.2 by revealing that there was a main effect of page ($F(3,357) = 4.75, p < .01$), a main effect of text type ($F(1,119) = 30.96, p$

<.001) and an interaction between page number and text type ($F(3,357) = 2.74$, $p <$ .05). Simple effects revealed that at each level of page number there were significant differences in the time participants spent reading these pages between the good and the bad texts (smallest $F(1,119) = 5.27$, $p < .05$) except for page 3 where the time spent in good and bad texts was not significantly different ($F(1,119) = 1.19$). For the good texts there was a significant difference between time spent on pages 1 and 3 only ($F(3,117) = 6.36$, $p < .001$), and there were no differences in time spent between the four pages in the bad texts ($F(3,117) = 1.15$).

As in Experiment 2, because of the independence assumption, an additional analysis was carried out on the data of the first good and bad texts visited only. The pattern of results was the same (a main effect of page ($F(3,138) = 8.59$, $p < .01$), a main effect of text type ($F(1,46) = 24.13$, $p < .01$) and an interaction between page and text type ($F(3,138) = 5.65$, $p < .01$).

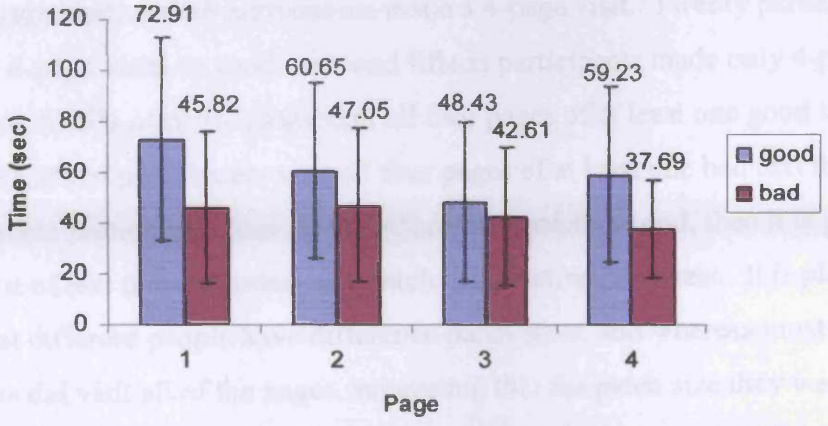


Figure 3.2 The average time in seconds (with +/- S.D.) participants spent reading pages 1 to 4 in both the good and the bad texts in Experiment 3.

Figure 3.2 shows that if participants enter a page of the text, then overall they spend similar amounts of time reading that page wherever it is in the document (i.e., whether it is the first page or the last page), except for the good texts where there is significantly less time spent on page 3 than on page 1 overall. The pattern of data in this experiment is slightly different from that found in Experiment 2 in that previously time tended to increase across pages (although the only significant difference was between page 1 and 3) for the good texts whereas in this experiment the time tends to decrease (again, only significantly different between page 1 and 3). This may be due to the subtle difference in the experimental set-up. Experiment 2 required participants

to simply study the questions and then read, whereas Experiment 3 required participants to attempt the questions, then study them, then read. The act of attempting to answer the questions may have caused the participants not to study them quite so well (they may have been eager to get on with the reading, or they may have felt as though they did not need to study them because they had already read through them), and therefore leading them not to remember the questions as well when reading.

However, the data showing the time participants spent on pages also indicates that sometimes people leave documents before they have visited all of the pages of the text. In terms of how many pages participants visited, for the good texts, 5 participants made a 1-page visit, 5 participants made a 2-page visit, 5 participants made a 3-page visit, and 28 participants made a 4-page visit. For the bad texts, 4 participants made a 1-page visit, 7 participants made a 2-page visit, 10 participants made a 3-page visit, and 26 participants made a 4-page visit. Twenty participants made only 4-page visits to good texts and fifteen participants made only 4-page visits to bad texts. 87.5% of participants visit all four pages of at least one good text they visit, and 81.3% of participants visit all four pages of at least one bad text they visit.

If some participants leave texts before they reach the end, then it is possible that the unit of text they are using as a patch is the entire document. It is plausible to assume that different people have difference patch sizes, and whereas most participants did visit all of the pages, suggesting that the patch size they were using was less than or equal to a page, for the participants who did *not* visit all of the pages, their patch size could have been the document. That way, these participants would leave the document when information gain fell below their threshold, and rather than jumping to the next page (which would happen if the page was the patch) they would jump to the next document.

Test Scores

The pre- and post-test scores for participants in Experiment 3 are shown in Table 3.8. There is no difference in the pre-test scores for participants who spent more time reading the good or the bad texts ($t(30) = .06$, *n.s.*), but participants who spent more time in the good texts had significantly higher post-test scores than participants who spent more time reading the bad texts ($t(30) = 3.00$ $p < .01$). This

indicates that participants reading the good texts learned more than participants reading the bad texts.

Table 3.8: Pre- and post-test scores for all participants, participants who spent more time in the good texts, and participants who spent more time in the bad texts

	Pre	Post
Overall	3.22 (*1.70*)	12.83 (*5.89*)
Participants who spent more time in good texts	3.23 (*1.72*)	14.4 (*5.88*)
Participants who spent more time in bad texts	3.19 (*1.77*)	8.0 (*2.14*)

Table 3.9: Learning scores for participants who spent more time in good or bad texts

	Text	
	good	**bad**
Mean	0.47	0.2
S.D.	*0.23*	*0.1*

A correlation was carried out between the time participants spent on the good texts and their learning score for the test (shown in Table 3.9). This showed a significant relationship between time spent on good texts and learning score ($r = .681$, N = 32, $p < .001$) in as much as the more time participants spent reading the good texts the higher their learning score was. This shows that those participants who were adaptive in spending more time in the good texts did in fact learn more from their reading than did those participants who spent more time reading the bad texts.

Overall, the mean learning score was 0.41, *S.D. 0.24*, with a range of 0.04 to 0.87. A median split of the learning scores was carried out to divide participants into low and high learner groups (median = 0.33). Data shows that in the low score group eight participants out of sixteen (50%) spent more time in the bad texts and seven out of sixteen (44%) had a bad text as their favourite text. However, in the high score group none of the participants spent more time in the bad texts and none had a bad text as their favourite text. This reinforces the claim that being adaptive in terms of text choice had a beneficial effect on learning.

3.3.4 Discussion

Experiment 3 was designed to provide a more robust test of adaptivity by showing that those texts classed as 'good', and which were more popular amongst readers, were good due to their content only (answering the questions contained in the test for which participants were studying), and not for any other reason, such as the texts being more readable, or more interesting.

As would be predicted on the basis of Experiments 1 and 2, the browsing patterns of 75% of the participants in Experiment 3 can be explained by the satisficing model. The fact that only one participant out of thirty-two used a sampling strategy strengthens the claim that sampling is not a popular strategy for allocating time across texts, at least under these conditions. However, since the task in this experiment, as in Experiment 2, could be viewed as more of a search task (because participants were aware of the facts for which they needed for the test), then it would be argued that this would not encourage a sampling strategy anyway. It is evident from these results, and from those obtained in Experiment 2, that the behaviour evident when readers are allocating time across texts in order to learn for a test fits well with the idea that they are using a satisficing strategy. Again, it is conceivable that the search model can account for the data regarding more time being spent on the good texts than the bad texts, but as will be discussed in Section 3.4, the model still fails to account for some aspects of the data.

With regards the amount of time participants spent in the good and bad texts, it is promising that three quarters of participants spent more time in the now 'good' texts, and only a quarter in the now 'bad' texts. This deals with the uncertainty surrounding the desirability of the good/bad texts as readers were adaptive in choosing to spend more time in the texts which contained the answers to their test questions. This establishes that the previously 'good' texts were only good due to their factual content, and not due to any other factor, since this experiment showed less preference for these texts. Again, the fact that there was no difference in the number of participants who favoured (i.e., spent most time reading) one good text over another further supports the case that it was the classification of 'good' texts having this effect, rather than a preference for any individual text.

The time participants spent over the four pages of a text within each text category appeared to be similar (see also Experiment 2). For the bad texts there were

no differences in time spent on pages, suggesting that each page was treated as a separate unit and judged on its own merits and not as a function of the page which preceded it. For the good texts, there were no differences between pages except for one, where significantly more time was spent on average reading page 1 of good texts than page 3. There is no specific explanation for why this may have occurred, but it is possible that page 3 of the good texts was not as close as the other pages to the remembered topic of the test questions, and this may have caused participants to spend less time reading it. Between the good texts and the bad texts, more time was spent reading each of the pages 1, 2, and 4 of the good texts than the bad texts, but no difference between pages 3 for each text type. Experiment 2 found that there was no difference in time spent on page 1 for good versus bad texts, but there was a difference between pages 3.

In terms of how long participants spent on the very first visit to their first text, this again depended on whether a good or bad text was in top position on the menu bar. Although participants could begin by reading any of the four texts, or indeed, any of the pages of any of the four texts, all but one participant began reading the first page of the text at the top of the menu bar. This can be explained by the fact that starting to read at the top, on the first page, is the natural way for native English speakers to begin reading. Those participants who were presented with a good text in the top position spent significantly longer reading that text than did those who were presented with a bad text in the top position. Readers spent longer reading the first text they visited if this text was a good text. If the text was meeting their level of satisfaction in terms of facts returned, then the reader continued reading, whereas if they found that the rate of return was not sufficient, then they abandoned reading and chose another text. This indicates quite clearly that these participants were being adaptive in their reading, and they were doing this by means of a satisficing strategy. Furthermore, the time spent on pages and texts suggests that the pages were being skimmed, in line with the 'satisficing then skimming' strategy being proposed here. This is because the time participants spent on pages of the bad texts were too short for all the words on the page to be read by the participant. The pattern of results is consistent with the idea that participants were monitoring how much useful information they were learning as they were reading (integrating judgement and learning) and if this fell below their threshold level then they abandoned reading the text and moved onto the next.

An analysis which strengthens the adaptivity hypothesis further still is a comparison of the length of all the first visits to texts throughout the reading phase of the experiment. A significant difference was found between the mean length of a first visit for the good texts and the bad texts, with participants spending longer on their first visits to good texts. Also, it has been shown by way of a correlational analysis that the more time participants spent in the good texts, the better their learning scores were. Therefore, being selective, or adaptive, in choice of text did have a bearing on the amount learned.

There are a few plausible reasons for why six of the participants were not adaptive in their allocation of time across texts. Of these six participants, four were presented with a bad text in the first position and only two with a good text presented in first position. It is possible that, considering most of the participants began by visiting the text in top position, that they spent more time in bad texts because they started off in a bad text. There are several possible explanations of why participants may display maladaptive behaviour, and these were discussed in Section 3.2.4.

Another analysis which followed this experiment was investigating the degree of systematicity in participants' browsing behaviour. Very few participants (9%) failed to visit all four available texts. Eighty-one percent of participants did not revisit any text before having visited all the available texts. This result shows that people are systematic in their reading behaviour. Choice of which texts participants read is not just random, but part of a systematic approach which appears to take into account the number of texts available to them and the time in which they have to study these texts, as well as the goal of their reading.

One element which was excluded from Experiment 2 was the requirement that participants complete the test before they began the reading phase. The inclusion of this pre-test in the present study allowed comparison of how much participants learned in relation to the time they spent on various texts. Tests revealed that those participants who spent more time in the good texts learned more than those who spent more time on the bad texts. This is the result that would be expected, since the answers to the test questions were mainly present in the 'good' texts, if participants spent more time reading the bad texts then it follows that they would not learn nearly as much as if they spent more time in the good texts.

3.4 General Discussion

Experiments 2 and 3 discussed in this chapter, along with Experiment 1 in Chapter 2 which focused on the allocation of time across multiple texts of varying difficulty but similar content, all point towards the fact that the satisficing model can explain the way in which people allocate time across texts under these conditions. That is, the data in these experiments show that the behaviour exhibited by readers allocating time across multiple texts fits with what a satisficing strategy would predict. Participants have been seen to use a satisficing strategy throughout their reading, and furthermore, have been seen to exhibit adaptive behaviour. Participants seem able to be adaptive in their reading, spending more time on resources which are useful and relevant to the experimental task, and less time on the resources which are not as profitable in terms of their information gain. Indeed, the findings from these experiments fit with what would be predicted according to Green's model of patch leaving decisions. According to Green's law (1984) more time would be spent in the texts where required information is being found than in the texts where information is not being found. Here, there is no pre-set threshold as such, but rather people monitor how much they are learning from the text and each set of required facts they come across increases the time they will end up spending in that text.

One model which may be argued could account for some of the data in these experiments is that of the search model, as mentioned in Chapter 1. This model assumes that participants search the texts for target items, that is, the answers to the test questions for which they are studying, by reading the texts linearly and stopping to learn information when target information is found and continuing to read when target information is not encountered. To show that participants were not reading through the texts and pausing to gather information, the time spent per page and the number of words on a page were analysed. The fact that each page of text contained between 315 and 353 words per page, and considering that the mean time spent reading a page for the bad texts was 46, 47, 42, and 38 seconds for pages 1, 2, 3, and 4 respectively, indicates that participants were not reading through the texts, because at a normal adult reading speed of 250 words per minute (Rayner & Pollatsek, 1989) they clearly did not spend enough time per page to read it.

Perhaps, then, this search model could suggest that participants are scanning the texts quickly, and then pause for some unspecified amount of time when target

information is found. Two threshold levels are needed for this model; one which tells the reader when to stop scanning and learn (i.e., when they make the judgement that they have found some target information), and one to tell the reader when to stop reading to learn and when to begin scanning again (i.e., when the reader has read all the information concerned with the target identified and needs to resume scanning the texts for further targets). The former could be based on familiarity judgements, and so when the reader recognises a word from the test, then they may stop to read about it. The second threshold of when to begin scanning again after reading may be when the reader has read the information associated with the word they recognised, or could be when they have finished reading a certain section of text (in which case it is not unlike a satisficing strategy which works on the premise that text is divided into patches). The one difference with this model and the satisficing model is that the search model does not involve readers jumping from one patch of text to another. The search model, like the satisficing model, is one which is flexible according to the definitions which identify when one process (e.g., scanning) stops and another (e.g., reading to learn) begins.

This search model could conceivably account for the data in Experiments 2 and 3 which show that more time is spent on the good texts than the bad texts; following the search model, participants would stop scanning and pause to learn more in texts which contain more target information, the good texts, than in texts which contain less target information, the bad texts. However, it does not account for the fact that some participants leave a text before they reach the end. In Experiment 2 just over half of the participants made a visit to a bad text which involved looking at all four pages of the text, and in Experiment 3 81% of participants visited all of a bad text. In both experiments nearly 88% of participants made a visit to a good text which involved looking at all four pages. The search model predicts two more or less uniform reading speeds, one when they are scanning and one when they are learning, but it would predict also that the participant would consult all sections of the text. If the model does not acknowledge that the text is divided into patches, then it would predict that all sections of the text are scanned.

The search model does not account for the data in Experiment 1 either, since participants were not searching for target information, because they were not aware of the test questions for which they were studying. However, the satisficing model does account for Experiment 1 data. The search model thus appears more complicated than

the satisficing model, as it requires two threshold levels or two instances when processes are changed, at which points some kind of judgement must be being made. The search model is also less plausible in the sense that it requires readers to scan the text quickly and still be able to make judgements between target and non target information in a text where the target information is very similar to non target information (as the topic of the test and the topic of the texts are both the human heart). It predicts therefore, that people are able to make semantic judgements about individual words without reading the text properly and that meaningful units of text can be identified by scanning quickly through the text. It has been suggested that a skimmer cannot fixate words sufficiently in order to extract a semantic interpretation from those words (McConkie & Rayner, 1975, as cited in Masson, 1982). Masson (1982) also notes that when readers are skimming a text for specific information, it is difficult to accurately discriminate between relevant and irrelevant information when deciding which sentences should be read and which should be skimmed over. Indeed, in the materials in Experiments 2 and 3, the bad texts and the good texts were both likely to contain target words from the test, just the good texts answering the questions and the bad texts not answering the questions. How do people discriminate between these without reading the text? This must be defined within the model for it to be applied to data.

Therefore, although it appears that the search model may be able to account for some of the data in Experiments 2 and 3, it seems relatively implausible as a model and it does not account for the pattern of data in Experiment 1. The satisficing model, however, accounts for the data from Experiments 1, 2, and 3. Nevertheless, more precise evidence may be needed to fully separate the two models. The studies in the next chapter will attempt to furnish this evidence.

It could be argued that the interface used for these experiments was equally supportive of a text sampling strategy and a satisficing strategy, as the buttons to access the texts are arranged in such a way that the cost of switching texts is exactly the same as the cost of switching pages within a text. One view may be that this interface design encourages a reader to sample the pages of another document, since there is no cost in doing this. This view provides reinforcement for the theory that readers favour a satisficing strategy, as there was no preference for sampling despite the interface layout.

The results from these experiments suggest people prefer satisficing to sampling, with only 7% of participants across both experiments displaying sampling behaviour. Perhaps the reason why sampling is not popular in these experiments can be found in the distribution of information within the texts. If the distribution of facts required for the test is not even, then sampling a patch of the document will not give the opportunity for a reliable judgement of the whole document. The bad texts still deal with the same topic, so how does a reader know that they are not about to come across what they are looking for any minute? A piece of required information may be 'just around the corner'. Also, a reader does not know that the next text they view will not be even worse than the current one.

Sampling may be pertinent to the allocation of time across multiple texts because one's information needs are likely to change after reading any relevant text. Sampling is only possible to the extent that text quality for the task at hand can be judged reliably and quickly relative to the time available to read the text. Sampling is likely to be better than satisficing if there is more than one text that is 'good enough' and yet significant and perceptible differences still exist among these texts. Unless both of these conditions are met, then satisficing may be the better strategy. People are perhaps reluctant to make a judgement about a whole text merely on the basis of the first paragraph or page, and this may be why sampling is not a popular strategy. However, it appears that sampling is the preferred strategy when outlines are used (Reader & Payne, 2007). This may be due to the fact that participants find it easier to make judgements about the text based on the outline, and therefore are able to choose which text best suits their needs in little time, thus not infringing on their learning time.

Whilst the discovery that there is a clear and strong preference on the part of readers to use a satisficing strategy when allocating time across texts is an important addition to the field of information foraging, it is not known exactly *how* readers are satisficing. The experiments presented in Chapters 2 and 3 provide evidence that implies that people do appear to satisfice when allocating time across texts, but what needs to be investigated next is the way in which people implement this strategy. The satisficing model assumes that the text is divided into patches, and in order to refine the model it is necessary to determine what constitutes a patch. To do this we need to determine the level of text at which people are satisficing; are they making the decision to stay in a text or leave it on the basis of the first few sentences, the first

paragraph, the first page, or the first paragraph of each page? Are they skim-reading a paragraph and then making their decision, or are they skimming through the whole text? One method which enables the study of such behaviour is through the use of eye tracking equipment. Once the eye movements of readers when they are using this satisficing behaviour to allocate attention are tracked, then the satisficing theory can be developed, leading to a deeper understanding of the way in which people use this time allocation strategy which appears so popular and widespread. Eye tracking will also allow further consideration of the search model as a competing explanation of satisficing browsing patterns.

CHAPTER 4

4.1 Introduction

One method for investigating what readers are doing when allocating time across written texts is by using eye-tracking. Some research using eye tracking has concentrated on how users browse websites. Schroeder (1998) claims that eye tracking is useful because it can tell us whether the user is reading or scanning the information, and how long a user has been looking at a particular section of the website. A recent study by Russell (2005) focused on first time usage of websites, where eye tracking data was analysed to provide information about how users view a web page the first time they visit it, and what aspects of that page attracted their visual attention. Eye tracking has also been used in research into visual search of menus, (e.g., Byrne, Anderson, Douglass, and Matessa, 1999), where it was found that the first eye fixation relates to the initial menu item, search is generally top to bottom and rarely random, and some items are 'skipped' during the top to bottom search. Byrne, John, Wehrle, and Crow (1999) also conducted a study to look at the way users spend their time when browsing on the web, and they concluded that websites should be designed for readability *and* scannability.

Although a lot of the research carried out using eye tracking as a tool has focussed on website usage and internet browsing, the benefits of eye tracking data and the information this can provide is still useful for the current research. Goldberg et al. (2002) suggest that eye-tracking can provide us with an insight into the way in which users make decisions when browsing by observing the patterns of fixations and saccades which produce scan paths that illustrate the distribution of visual attention on a screen. It is this aspect of the eye-tracking data that is of interest in these experiments, in order to help in understanding how people allocate their time across information sources in the form of written texts.

This chapter presents two experiments that collected data from an eye tracking system in order to investigate the way in which readers appear to use a satisficing strategy as a means of adaptively allocating their time to reading the texts or documents which are most beneficial to them, given their specific task requirements. Experiments reported in Chapters 2 and 3 have indicated that when readers are faced with too many texts to read in the time available, that they tend to allocate more time

to those texts from which they will most benefit, although they do not always manage to do this (in the case of Experiment 1 these would be the texts within their zone of learnability, and in Experiments 2 and 3, these texts would be those which answered the test questions for which they were studying). In addition to showing that readers were adaptive, the data in Experiments 1, 2, and 3 suggest that these readers managed to allocate their time adaptively through using the satisficing strategy. That is, the data has been interpreted using the satisficing theory, which fits with the behaviour participants displayed. By employing this satisficing strategy, readers would have been able to adaptively allocate more time to texts which were most beneficial to them in terms of their task goals. The data in these experiments fits with the assumption that the reader began reading a patch of text, and if their reading of this patch was fulfilling their information needs then they continued, but if the patch of text was not providing the information necessary for them to complete the post reading test, then they left that patch and began reading another. The browsing patterns evident in Experiments 1, 2, and 3 can be explained by this satisficing model and suggest that participants are using a satisficing strategy to allocate their time across the texts rather than a sampling strategy, or an alternative strategy such as linear reading.

Another reason for conducting eye tracking studies is to help clarify whether the search model is a plausible competing account of the behaviour evident when people are allocating time across written texts. The search model predicts that people scan through texts and pause to learn when target information is found. This model assumes that people are able to make judgements about information value by scanning each word or phrase of a text and stopping when a target is identified (a target being a word or phrase they have seen in the test questions). By looking at readers' eye movements it will be possible to determine what they are doing when they are reading within a page of text. The design of the following experiments favour the search model, as participants were asked to learn for a specific test and were shown the test questions before the reading phase. This set up was used to try to prompt participants to use a search style model.

If readers <u>are</u> satisficing, we do not know at what level of the text the readers are satisficing (i.e., what is the patch size). It is necessary to determine what unit of text is being treated as a patch (is it small, such as a sentence or a paragraph, or larger such as a page or even the entire document?). Although it has been seen from

previous experiments that participants consulted most of the text before leaving it to read elsewhere, it is not clear what unit of text was being judged. Masson (1982) suggested that readers skimming a story were satisficing at the level of the paragraph in so much as they were reading the first sentence of each paragraph in order to get a gist of the text. However, he found this out from descriptions of what the participants thought they were doing when skimming.

In order to investigate *how* readers are satisficing, the following experiments make use of eye-tracking equipment to reveal the movements made by the readers' eyes whilst allocating their time across the texts. The aim here was to look into the descriptive nature of skimming that arises from satisficing. This could potentially inform about how skimming a page and rejecting a page/text is played out.

4.2 Experiment 4: Eye tracking with blocked text

4.2.1 Introduction

The first of the eye-tracking experiments focused on the eye movements of the readers when they were faced with the task of allocating time across four texts which were presented as one large block of text per page, as in previous experiments.

4.2.2 Method

Participants

Sixteen participants were paid for their participation. Nine females and seven males took part, and the mean age was 24.06 years, SD 2.29, range 21 – 29.

Design and Materials

The experiment employed a within-participants design. The dependent variables in this study were the browsing behaviour of the participants as measured by the time spent in each text and the number of visits made to each text, and also the time spent in each quarter of text for each text type, good and bad. The independent variable in this experiment was the text quality (good or bad).

The four texts on the human heart were used in this experiment were those used in Experiments 2 and 3 (see Section 3.2.2 for details). The set-up and layout of the texts on screen was also the same as used in Experiments 2 and 3. Figure 4.1

shows the interface layout. The order in which the texts were displayed is the same as in Experiments 2 and 3 (see Section 3.2.2 for details).

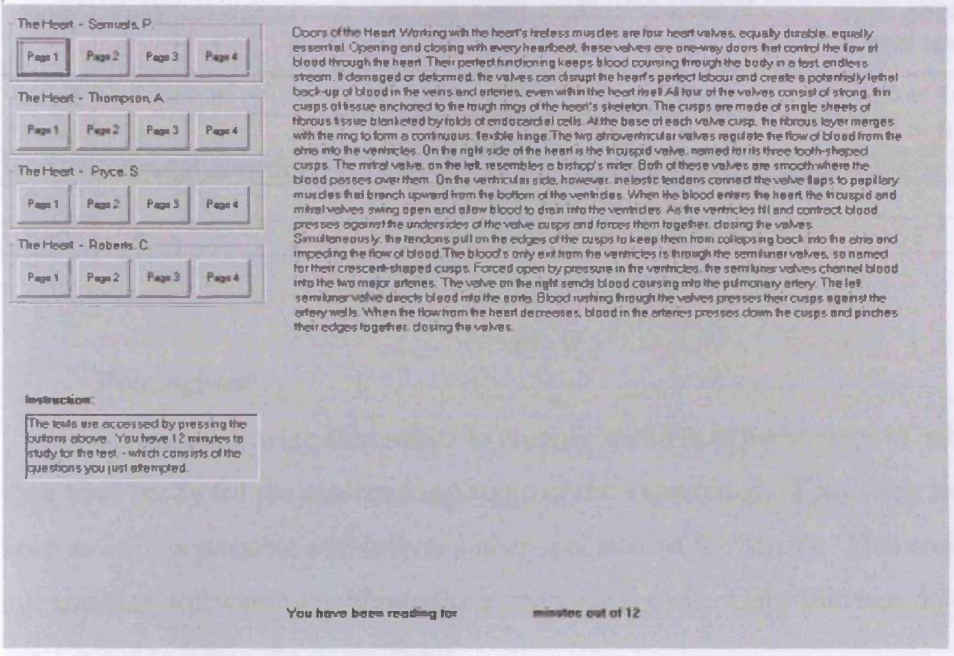


Figure 4.1 The reading interface for Experiment 4.

A Tobii 1750 eye tracking system was used to record the eye movements of the readers, and ClearView 2.1.5 analysis software package was used to analyse the data. One of the benefits of the Tobii 1750 eye tracking hardware is the fact that it is unobtrusive and therefore does not place unnatural restrictions on the user. It consists of a high resolution camera and near infra red light emitting diodes and records the position of the users' eyes roughly every 20ms. The Tobii 1750 eye tracker takes measures from various angles, which means that users can move their heads slightly during reading and the data is still collected. The analysis software package allows the experimenter to create Areas of Interest (AOIs) on the screen so analysis can be confined to these areas. The eye fixations can then be analysed in terms of these areas, and can be compared by looking at the order of fixations, the dwell time of fixations, and the total number of fixations in each AOI.

The texts (two good and two bad) and the two tests were those used in Experiments 2 and 3. Table 4.1 shows the number of facts on each page and in each quarter for the two good versus the two bad texts.

Table 4.1: Number of facts per page and per quarter for good and bad texts

Text	Test	Page				Quarter (averaged across pages)				Total facts in each good or bad text
		1	2	3	4	1	2	3	4	
Good	1	8	8	9	7	9	7	10	5	15 or 16
Good	2	5	3	5	4	3	4.5	6	3.5	8 or 9
Bad	1	3	3	2	2	2	6	2	0	5
Bad	2	1	2	1	0	0	1	0.5	2.5	2

Procedure

Participants were first asked to comply with the experimenter in calibrating their eyes ready for the eye tracking stage of the experiment. They were asked to keep as still as possible and follow a blue spot around the screen. This enabled the eye tracking software to calibrate the participants' eyes. Once this was done they were asked to begin the experiment.

As in Experiment 3, participants were given 17 on-screen short answer questions about the human heart. They were told that they would have to answer the same questions after the reading phase, so they knew what information they would be looking for when reading the texts. As before, the questions were laid out over two pages, and once participants had moved onto the second page they were unable to return to the first page. There was no time limit for this stage of the experiment.

The procedure for the reading phase and the post-reading test is the same as used in Experiments 2 and 3, as are the instructions given to participants (see Section 3.2.2). The eye tracking recorder was activated and ran whilst the participants were reading the texts. Once the reading phase was over the eye tracking recorder was deactivated and the participants were asked to proceed to the next phase of the experiment.

4.2.3 Results

This section will deal with data of interest in the following order: it will firstly show the time participants spent in the various texts and pages of the texts, which will inform about how adaptive participants were in their choice of texts. It will then look

at the time allocation strategies used. Then, in order to help identify the level of text being judged, the number of pages and texts visited will be analysed, along with the amount of time spent in each quarter of a page of text, which will reveal how readers behave *within* a page of text. It will then show gaze plots which illustrate the pattern of eye fixations during reading and give a visual representation of readers' behaviour.

Adaptivity

The time readers spent on the various texts and pages of a text is interesting to the extent that if readers spent more time reading the good texts, then this suggests that they were able to identify the more appropriate information sources for their needs. This is a replication of analyses done in Experiments 2 and 3. In terms of the effect the text quality had on the amount of time readers spent reading that text, it is apparent that all of the participants spent more time reading the good texts, and all of them had a good text as their favourite text (single text in which the participant spent most of their time: binomial test significant, $p < .001$). Table 4.2 shows that the times spent in the good and bad texts.

Table 4.2 Time (sec) spent in texts

Time in Good texts		**Time in Bad texts**	
Mean	592.7	**Mean**	116.58
S.D.	*77.4*	***S.D.***	*78.35*
Minimum	455.72	**Minimum**	0
Maximum	711.27	**Maximum**	261.45

Since comparing the total time on good versus bad texts violates the independence assumption (since the total time for reading is fixed), an analysis was conducted on the time taken to switch from the first text visited. The time taken to switch from the first text visited to the next text was compared for those participants who had a good text presented first in the menu and those who had a bad text first. If the time taken to make the first switch is significantly different, it would suggest that the reason for time differences throughout the reading time are attributable to the difference in text quality rather than the way in which time on one text relates to time on another. Table 4.3 shows that the mean time taken to switch away from the first text for readers who were presented with the good text first was significantly higher than for those presented with the bad text first ($t(14) = 16.75, p < 0.001$).

Table 4.3: Time to switch from text 1 to text 2 for good and bad presented first

Text presented first		Time taken to switch to next text (sec)
Good	Mean	426.06
	S.D.	*44.61*
Bad	Mean	83.85
	S.D.	*36.7*

The average time participants spent reading a page of text was also analysed in terms of whether the participant had been presented with a good text or a bad text in the first position on the menu (although participants can choose which text to begin reading, all participants did in fact begin by reading the first text). This was to determine whether readers were influenced by their experiences of previous texts. Table 4.4 shows these times.

Table 4.4: Time spent reading a page of text for good and bad text presented first

		bad text first	good text first
Mean time per page	**good**	80.63	91.46
	bad	28.24	16.64
S.D.	**good**	*10.22*	*12.21*
	bad	*6.23*	*15.22*

The data in Table 4.4 suggests that there is no difference in the time spent reading a page of a good text when a good text was presented first than when a bad text was presented first ($t(14) = 1.92, p = .07$). This suggests that for the good texts, readers were not influenced by the order of the texts in the menu bar, since they tended to spend a similar time per page on the good texts when a good text was presented first and when a bad text was presented first. Similarly for the bad texts, there was no difference in the mean time spent per page of a bad text between the participants who were presented with a good text first or those presented with a bad text first ($t(14) = 1.99, p = .06$.

Time allocation strategy

All of the sixteen participants in this study used a satisficing strategy when allocating time across the four available texts in order to learn for a post-reading test.

All participants began reading the text in the first position, perhaps purely because it was at the top of the list of texts, and they consulted texts systematically. Only one participant revisited any texts, and even this was only done once all texts had been visited once each first.

An analysis was carried out to examine the time participants spent on each of the visited pages of a text, good and bad. Figure 4.2 shows the times spent on pages of the good and bad texts to which readers made visits. The figure suggests that the time spent on each page of a good text was higher than the time spent on each page of a bad text. An ANOVA with factors of page (1 to 4) and text type (good and bad) revealed that the time spent on pages within a text type were not significantly different (no main effect of page, $F < 1$), a main effect of text type, $F(1,40) = 91.23$, $p < .001$), and an interaction between time on page and text type ($F(3,120) = 2.8$, $p < .05$).

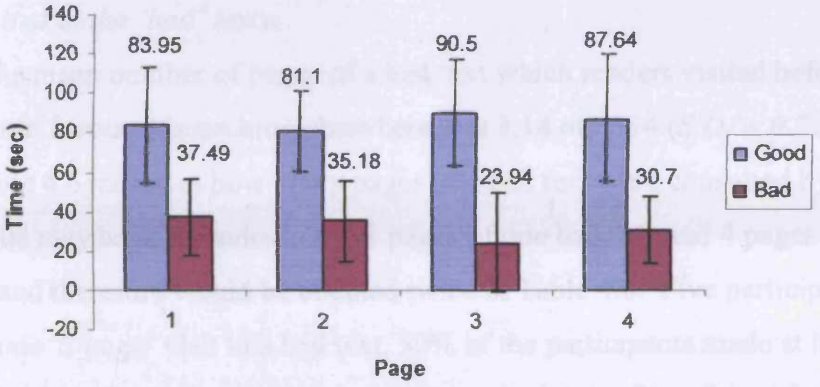


Figure 4.2 The time spent on pages visited for all pages of good and bad texts in Experiment 4.

Pairwise comparisons showed that there were differences for each page between the good and bad texts (smallest $F(1,40) = 35.79$, all $p < .001$) but there were no significant differences between any of the four pages within each text type, good ($F < 1$) and bad (largest $F = 1.34$).

Since the above analysis may violate the independence assumption, due to the fact that study time was fixed, the same analysis was carried out on the time participants spent on pages 1 to 4 of the first bad and the first good text visited. The pattern of results was the same (no main effect of page ($F < 1$), a main effect of text

type ($F(1,27) = 82.16, p < .001$), and an interaction between page and text type ($F(3,81) = 4.88, p < .01$).

The number of pages and texts readers visited was analysed in order to determine whether participants were exhaustive in terms of consulting all of the available texts/pages or whether they only made visits to some. Table 4.5 shows that over half of the participants (9) visited all four texts during their reading time.

Table 4.5: Unique texts visited overall

number of texts visited:	number of participants:
1	0
2	3
3	4
4	9

Pages visited in the 'bad' texts:

The mean number of pages of a bad text which readers visited before deciding to reject it in favour of searching elsewhere was 3.14 out of 4 (*S.D. = 0.79*, range = 2 to 4). Table 4.6 indicates how many pages of a bad text were consulted by readers. Participants may have consulted, say, 3 pages of one bad text and 4 pages of the other bad text, and therefore would be counted twice in Table 4.6. Five participants out of 16 made one '2 page' visit to a bad text, 50% of the participants made at least one '3 page' visit and 50% made one '4 page' visit to a bad text. Out of the 16 participants, 12 made either a 3 or 4 page visit (or both, or together with a 2 page visit), 1 participant made a single '2 page' visit, and 3 did not visit a bad text at all. If participants do visit a second bad text, then the number of pages they visit does not differ significantly from the number of pages they visit on the visit to the first bad text (first visit, mean 3.27, *S.D. 0.90*, second visit, mean 3, *S.D. 0.67*). A Mann-Whitney U test revealed that this difference was not significant, $U = 43, z = 0.81$, N1 = 11, N2 = 10, $p < 1$. Thus, it appears that the bad texts were not treated differently when the reader had already been exposed to a bad text during their reading. Three participants did not visit any bad texts and were therefore not included in the analysis. This behaviour is interesting because it suggests that each text is treated as separate and judged on its own merits rather than on the basis of the previous text.

Pages visited in the 'good' texts:

The mean number of pages of a good text which participants visited was 3.39 out of 4 (*S.D. = 1.08,* range 1 to 4). Table 4.6 shows that all participants made a visit to one of the good texts which involved looking at all four pages. The seven participants who made 1-2-or 3-page visits to a good text did this on the second good text consulted and ran out of time while reading that text.

Table 4.6: Pages of texts consulted by readers

Number of pages consulted	Number of participants who made a visit of this kind	
	Bad texts	Good texts
1	0	3
2	5	1
3	8	3
4	8	16

Table 4.7 shows the order in which texts were presented to the participants, and the number and order of texts the participants consulted during the reading phase. The far right columns in Table 4.7 show how many texts were visited by the participants according to the text sequence type they were given (two good texts followed by two bad texts (2G, 2B), two bad texts followed by two good texts (2B, 2G), alternating good and bad texts starting with a good in the first position (alt GB), or alternating good and bad texts with a bad text in the first position (alt BG). This shows that when a bad text was in the first position *all* the participants visited all texts, whereas when a good text was in first position *none* of the participants visited all of the texts.

Table 4.7: Order of text visits

text order (good or bad)	order texts visited	text order (good or bad)	order texts visited	Mean number of text visits for each text sequence type:	
1423 bbgg	1423	1423 ggbb	14		
2314 ggbb	231	2314 bbgg	2314	2G, 2B	2B, 2G
4132 bbgg	4132	4132 ggbb	41	3, 2, 2, 2	4, 4, 4, 4
3241 ggbb	32	3241 bbgg	3241		
1243 bgbg	124343	1243 gbgb	124		
2134 gbgb	213	2134 bgbg	2134	alt GB	alt BG
4312 bgbg	4312	4312 gbgb	431	3, 3, 3, 3	6, 4, 4, 4
3421 gbgb	342	3421 bgbg	3421		

As can be seen, only one participant revisited any of the texts after having already visited them once. All other participants only ever visited a text once, and always consulted the texts in a systematic order (i.e., corresponding to the vertical order of presentation).

An analysis was carried out to investigate whether participants treated a bad text differently (i.e., spent more or less time reading it) depending on whether it was preceded by a good text or not. Again, this was to determine whether times spent on certain texts were influenced by the quality of preceding texts. So, the time spent on the first bad text encountered (bad text 1) for the order of text alt BG was compared with the time spent on bad text 1 for the order of text alt GB (alt BG - mean, 74.43, *SD 32.62*; and alt GB - mean, 105.69, *SD 35.86*). An independent samples t test revealed that this difference was not significant ($t(6) = 1.29, p = .12$).

Time in each quarter page of text

In order to further refine the satisficing model Experiment 4 used an eye-tracker to identify what readers were doing within a page of text. The time participants spent reading each quarter of a page of text was used to analyse whether participants satisficed at the level of the text, page or sections of page. In order to analyse how readers were skimming and rejecting a text, the text on each page was split into quarters horizontally and the duration of their eye fixations in these quarters was taken as a measure of how long they spent reading them. Figure 4.3 illustrates the time participants spent reading each quarter of a page of text for the good and bad texts. The four data points for each text type were arrived at by averaging across pages, for the two good or the two bad texts, the time participants spent looking at each quarter one to four *if* the quarter was visited (i.e., only times for quarters that were actually visited were included in the calculation). It can be seen that a similar time was spent in all quarters visited for the good texts, but for the bad texts the reading duration lessened as the reader went down the page.

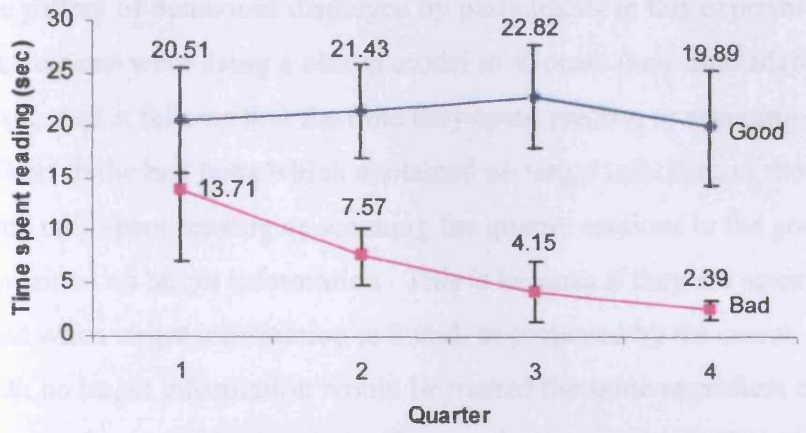


Figure 4.3 The mean time (with +/- S.D.) participants spent reading each quarter page of text for good and bad texts for Experiment 4.

Three participants did not visit any bad texts during their reading (all three of these participants had a menu order with 2 good texts appearing first and second), and therefore because no comparison could be made between the times spent on quarters of good and bad texts, the times for these three participants spent on quarters of the good texts were not included in this analysis. This analysis was thus carried out on the thirteen participants who made visits to both text types.

Figure 4.3 indicates that more time was spent on each quarter in the good texts than in the bad texts, and that there were more significant pairwise comparisons across quarters within the bad texts than within the good texts. An ANOVA with factors of text (good and bad) and quarter page (Q1 to Q4) confirmed this description of the data by revealing that there was a main effect of text [$F(1,12) = 296.28$, $p < .001$], a main effect of quarter [$F(3,36) = 6.65$, $p < .001$] and an interaction between text and quarter [$F(3,36) = 23.78$, $p < .001$]. Simple main effect analyses revealed that for each quarter (1 to 4), there was a significant difference in the time spent reading these quarters between the good and the bad texts (all $ps < 0.001$). Simple main effect analyses within text type (good and bad) at each quarter level revealed that for the good texts there was a significant difference in time spent in Q3 and Q4 only ($F(3,10) = 5.65$, $p < .05$). For the bad texts, there was no significant difference in time spent between Q3 and Q4, nor between Q1 and Q2. All other differences between quarters within the bad texts were significant (all $ps < .001$).

The data on time spent per quarter page of text was examined more closely with the aim of using it to clarify whether or not the search model could account for

the pattern of behaviour displayed by participants in this experiment. If the participants were using a search model to allocate their time adaptively across the four texts, then it follows that the time they spent reading or scanning over quarter sections of text in the bad texts which contained no target information should be similar to the time they spent reading or scanning the quarter sections in the good texts which contained no target information. This is because if they are scanning and pausing to read when target information is found, as proposed by the search model, then quarters with no target information would be treated the same regardless of whether they were in a good text or a bad text. Looking at the time spent reading page quarters (a mixture of quarters 1, 3, and 4) which contained no target information when they were good texts and when they were bad texts (depending on which test the participants were studying for) revealed that significantly more time was spent reading/scanning the quarters when they were in the good texts (mean 18.70, *S.D. 9.*47) than when they were in the bad texts (mean 10.70, *S.D. 9.43*); $t(82) = 3.58, p < .001$. There were more occasions when participants read these quarters when they were good texts (N = 58) than when they were bad texts (N = 26). This result suggests that the search model cannot account for the data in this experiment.

As well as measuring the overall mean time which was spent in each of the quarters for the good and bad texts, an analysis was done to investigate the number of participants who looked at each of the quarters on a page of a bad text. Out of the 65 visits in total to pages of a bad text (this was from 13 participants because 3 participants did not visit any bad texts at all), all of these 65 visits included looking at the first quarter, 63 included looking at the second quarter also, 55 at the third quarter also, and 48 at the fourth quarter. All but one participant visited the quarters in order, and only looked at later quarters if they had already looked at earlier quarters, for example, they would not look at Q4 without looking at Q3, or Q3 without Q2, et cetera. One participant looked at Q1, Q2, and then Q4 without looking at Q3, but this only happened on one page. Overall, this just implies that participants were reading in a conventional way, so this result is not surprising.

Some of the participants who made 4-quarter visits also may have made 3- or 2-quarter visits in a different text or page, and will be counted once for each category in the table below. Table 4.8 shows that 70% of visits to pages of a bad text involved the reader looking at all four quarters, only 3% included 1-quarter visits, 9% included 2-quarter visits, and 16.9% were 3-quarter visits.

Table 4.8: Visits to page quarters of pages in bad texts

Page Section	Number of visits	Number of participants who made a visit of this kind (out of 13 participants)
1 quarter only (Q1)	2	2
2 quarters (Q1,Q2)	6	5
3 quarters (Q1,Q2,Q3)	11	6
4 quarters (Q1,Q2,Q3,Q4)	46	12

There is no maximum cell value for the number of visits due to the fact that participants could re-visit texts and therefore there could be more than one type of visit (1 quarter, 2 quarter, etc) for each text. For each participant visiting each of the two bad texts, a certain type of visit was identified for each text and averaged out across the texts. So, if a participant only made 4 quarter visits to all texts, then they would be counted once, but if a participant made a 4 quarter visit to one bad text and a 2 quarter visit to the other bad text then they would be counted twice in the 'number of participants who made a visit of this kind' column.

A similar analysis was carried out for the good texts. Out of the 111 visits in total to pages of a good text (all 16 participants included), all included looking at the first quarter, all looking at the second quarter also, 108 looking at the third quarter also, and 103 visits looking at all four quarters. All participants visited the quarters in order. A frequency table is presented below to show the number of visits which were made to text quarters of a good text and how many participants made such visits (i.e., 1, 2, 3, or 4 quarter visits) in their reading.

Table 4.9: Visits to page quarters in pages of good texts

Page Section	Number of visits	Number of participants who made a visit of this kind.
1 quarter only (Q1)	0	0
2 quarters (Q1,Q2)	3	3
3 quarters (Q1,Q2,Q3)	5	5
4 quarters (Q1,Q2,Q3,Q4)	103	16

Again, note that some of the participants who made 4-quarter visits also may have made 3- or 2-quarter visits and will be counted once for each category in the

table above. Although there were some visits to pages of good texts where the reader did not look at all four quarters of the page, all of these cases occurred at the very end of the reading time, and so the reader not looking at some quarters of the text can be explained by them running out of time, rather than actively choosing not to look at them.

Gaze plots

A useful tool that can be used to gain an impression of how much time readers were spending looking at various sections of text or indeed which sections of text the reader was looking at is the gaze plot. A gaze plot allows analysis of the fixations and the scan path for each reader at any point in the recording. Each fixation is represented by a dot, and the larger the dot the longer the fixation. In order to gain a deeper understanding of what an individual reader is doing when asked to allocate their time across texts under time pressure, two exemplary participants were chosen and their behaviour was analysed. Due to the fact that all participants in Experiment 4 displayed behaviour consistent with a satisficing strategy, then the examples can only represent readers who were satisficing. The following gaze plots give an impression of what a participant who appears to be using a satisficing strategy is doing when they are reading/allocating time within a page of both good and bad texts.

Figures 4.4 and 4.5 show the scan path of participant 1 when reading a page of a bad text and a page of a good text. The texts were presented to participant 1 in the menu bar in the order of bad, bad, good, good. Participant 1 spent most of their time in the good texts and their favourite text was a good text. They visited the texts in the order they appeared on the menu bar and visited three pages of the first bad text, two pages of the second bad text, and then all four pages for each of the two good texts.

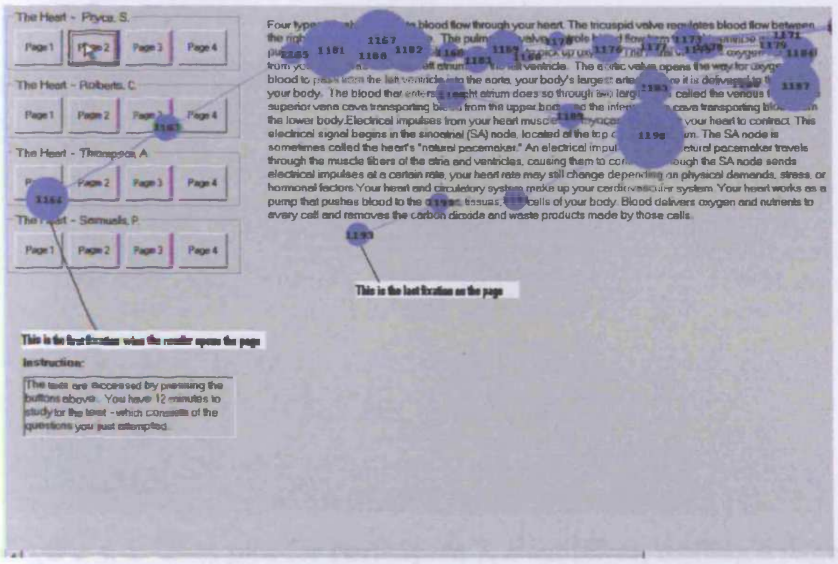


Figure 4.4 Gaze plot for a bad text detailing eye fixations for a representative participant reading page one of the second text in the menu bar (Experiment 4).

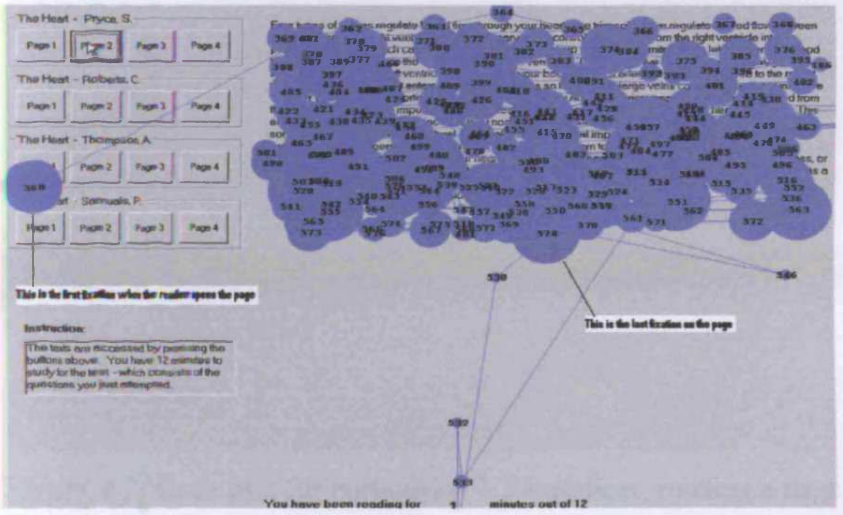


Figure 4.5 Gaze plot for a good text detailing eye fixations for a representative participant reading page one of the third text on the menu bar (Experiment 4).

The gaze plots show, by the distribution and density of the fixation points (the blue dots), that the reader is spending more time reading the good text than the bad text. The scan path in Figure 4.4 indicates that the reader is perhaps skimming the text rather than reading it fully.

Figures 4.6 and 4.7 show another participant who was presented with the texts in the menu bar as bad, good, bad, good. They visited all of the texts and visited all the pages of each text except for the text in first position of which they only visited three pages. They spent most of their time in the good texts and their favourite text was a good text.

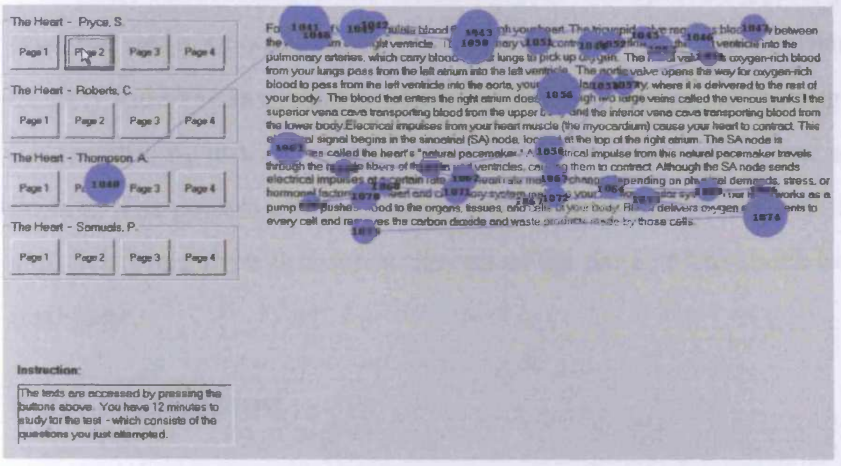


Figure 4.6: Gaze plot for participant 7, a satisficer, reading a page of a bad text (Exp 4).

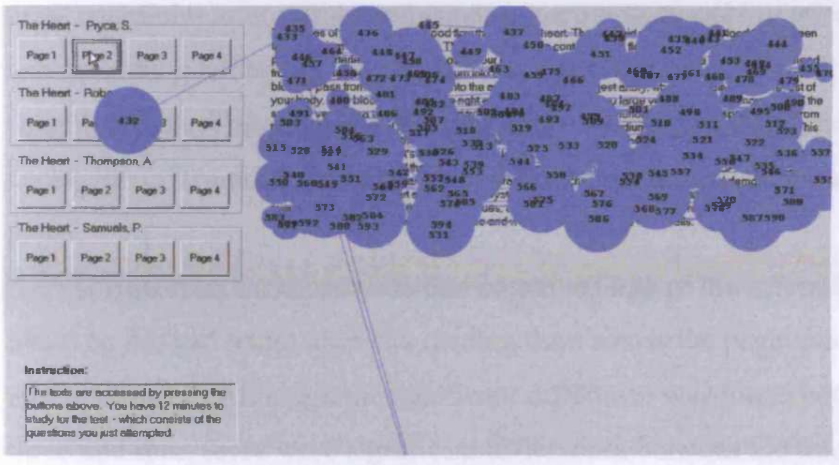


Figure 4.7: Gaze plot for participant 7, a satisficer, reading a page of a good text (Exp 4)

These gaze plots clearly show that a judgement was being made about the value or usefulness of the text near the top of the page, and this judgement influenced whether the participant read the page fully or skimmed through it. Participant 7 did not abandon the page of the bad text after reading the first section of it, which indicates that they were not prepared to judge the entire page of text on the basis of the first few lines. This supports the claim that the unit of text being used by satisficers as a patch is smaller than the page. In fact the pattern of eye fixations in the bad text suggests that the patch they are using for judgement decisions is as small as a section of a page (quarter/paragraph).

However, the pattern shown in these examples of typical participant behaviour does not really fit with the original satisficing model, as participants still appear to quickly view the last quarter of a page before abandoning the page to read elsewhere. The satisficing model would predict that participants either read a patch of text or they do not. The pattern of data evident from Experiment 4 suggests that participants are satisficing and then skimming the rest of the page just to check before moving to the next page.

4.2.4 Discussion

The results from analysing the eye-tracking data, in terms of how long participants spent reading each quarter of a page of text (Figure 4.3), reveal an interesting pattern which informs the theories of adaptivity and satisficing. For those texts containing more of the information participants were required to learn for the test (i.e., the good texts), reading time across the four quarters of the page was roughly equal. This suggests that participants started reading, and as they came across the information they needed to learn, they continued to read through the entire page.

However, for those texts that contained less of the information needed for the test (i.e., the bad texts), then the reading time across the page-quarters decreased monotonically. Although no significant difference was found between page-quarters three and four, there were significant differences between the time participants spent reading all other page-quarters. The pattern of data recorded suggests that readers began by reading the page normally, and then, as they realised they were not getting the information they required, they proceeded to skim the rest of the page before moving to the next page, since most people read at least three pages of the first bad text they encountered. Thus, the assessment rule proposed by Green (1984) would map onto the current data in terms of the participant encountering less information than they require after having looked for it a certain number of times in that patch, and therefore choosing to abandon the page in favour of reading elsewhere.

With regard the number of pages of a bad text visited by readers, on average, participants visited 3.14 out of the 4 pages of a 'bad' text before deciding to reject it in favour of searching elsewhere. This suggests that the unit of text being judged was less than or equal to the *page*, since they were not prepared to reject the entire

document on the basis of one page (the unit of text would be the document/entire text if participants left after visiting the first page only), nor were they prepared to reject the page on the basis of the first paragraph alone (the unit of text would be the paragraph – but in this study there are no paragraphs in the text so it would not be possible for this to occur). Instead, after reading the first section of text, it appears that they skimmed the remaining text before deciding to quit the page. If all or most pages were visited, but the pages were each abandoned quite quickly as they were in the bad texts in this experiment, then this suggests that the unit of text being judged was less than or equal to a page. The patch therefore appears to be a section of text, as indicated by the fact that readers do not leave the page after reading only the first section, but continue to consult all sections of a page before moving to the next page. If the patch-unit was a page, then it would be expected that readers would leave the page before reaching the end, which they did not do in Experiment 4, therefore suggesting that the unit must be smaller than a page (e.g., a section of text). Also, the fact that a visit to a second bad text did not cause the reader to consult fewer of the pages than s/he did on the visit to the first bad text (if a second bad text is viewed in the reading time) suggests that prior experience of a bad text did not affect the readers' strategy, since they skimmed the pages and rejected the text in the same manner. The data in Figure 4.2 showing the time spent on pages 1 to 4 of the good and bad texts presents a similar overall pattern to the data in Experiments 2 and 3.

Also of interest was the order in which participants read the available texts. As is evident from Table 4.7, which details the number of texts each participant visited as well as the order in which they visited these texts, all participants visited the texts in the order in which they were presented on the menu bar from top to bottom. However, what makes this interesting is the fact that less than half of the participants in Experiment 1 behaved in this manner. In Experiments 2 and 3, most participants behaved as did those in Experiment 4 and consulted the texts in the order they were presented on the menu bar, suggesting that it is the data from Experiment 1 that is the exception. It seems, therefore, that when participants were told that some texts may be better than others (as they were in Experiments 2 to 4) they were more likely to consult the texts in the menu order than when they were not advised about the quality of the texts (as in Experiment 1). Furthermore, the majority of participants did not visit any text more than once. One participant visited each text in turn (in the menu order) and then revisited the last two texts, although these visits were relatively short

(29 seconds for the bad text, 54 seconds for the good text) and mainly to consult the pages not viewed in the initial visit to these texts. From this data we can conclude that readers are influenced by the display with regards the order in which they choose to read available texts and that generally, if readers reject a text they do not revisit it. Similarly, the data suggests that if readers find a good text they will read it at that time and not need to return. Such behavioural tendencies point towards a theory of satisficing. Sampling would show evidence of much more text switching than is happening here, at least four initial visits to each text before deciding on the best one to read.

The purpose of the analysis investigating how many visits were made to each quarter of text, and how many participants made such 1, 2, 3, or 4 'quarter consulting visits' was to better understand the skimming behaviour of the majority of the participants. Table 4.9 shows that over 70% of visits to pages of a bad text involved the reader consulting, or looking at, all four quarters of the page (46 out of 65 visits made). Only 3% included 1-quarter visits (where the reader consulted only the first quarter of the page before switching to the next page or text). The table also shows the number of participants who had carried out each of these type visits, indicating that 92% of participants who visited the bad texts made '4 quarter' visits and only 15% (2 participants) made one '1 quarter' visit. For the majority, readers would consult all four quarters of the text on the page before rejecting it in favour of searching elsewhere (either on the next page or the next text). Considering that the time spent on each quarter down the page of a bad text decreases significantly, this further supports the claim that readers are skimming the whole page, reading less and less as they progress down the page, as they are rejecting sections of text in favour of the next, although they reset their satisficing criterion when they begin the next page.

The data from the gaze plots (see Figures 4.4 and 4.5) clearly show that when reading a bad text the eye fixations were shorter and there were fewer towards the bottom of the page than when the same participant was reading a page of a good text. This suggests that readers judged the value of a page of text by reading the first few lines to see if the page was going to contain the information they required. If they realised that it did not contain what they were searching for, then they skimmed quickly through the remaining parts of the page before moving onto the next. An important point to note is that the design of the interface which has been used for previous experiments, and also in this experiment, was such that each page was

presented as a block of text, which could therefore be viewed as one paragraph, as well as being one page. If a block of text on a page is only one paragraph long, then it would be unclear whether readers were satisficing at the level of the paragraph or the page, since in this design a paragraph and a page are the same. Therefore, an interesting question based on the design of the reading interface is what readers would do if the text was presented differently. If each page consisted of more than one paragraph, then would readers skim through each paragraph before rejecting the page, or simply read the first paragraph and use that as an indication of information contained later in the page? It is assumed from the results of this experiment that the unit of text being judged as inadequate when a decision to abandon is made is less than or equal to a page. In order to determine whether this unit is smaller, for example, a paragraph or half page, then it is necessary to look at eye movements when reading text separated into paragraphs. Thus by comparing readers' eye movements when the text is presented in paragraphs we can define more clearly the level at which readers are satisficing.

The results of this study suggest one important design implication: if readers use specific patches of text to make judgements about the value of the text, which in turn determines whether they want to continue reading or not, then it makes sense that text designers apply this principle when designing online texts, such that they place the salient information at positions in the text that readers are most likely to use to judge the usefulness or appropriateness of the information source. This is reflected in the writing style used by journalists, known as the 'pyramid' style (Ricketson, 2004). This will be discussed in more detail in Chapter 6.

4.3 Experiment 5 – Eye tracking with paragraphed text

4.3.1 Introduction

Experiment 4 showed that participants began reading a page normally and would then skim (as shown in the gaze plots, Figures 4.4 and 4.6) the remainder of the page sections if they found that they were not encountering the information necessary for the test. This could indicate that readers are satisficing at the level of the page section, or less, at least under these conditions. However, the fact that participants skimmed the remaining sections on a page before leaving does not quite fit with the satisficing model, and therefore could be interpreted as participants satisficing and skimming to check before leaving. Taking into account the research done by Masson (1982), who found that readers skim the first sentence of each paragraph, and the fact that the pages in Experiment 4 consisted of one long paragraph, Experiment 5 was designed. It was decided that repeating Experiment 4 with a reader interface designed with pages split into three paragraphs of text would help to examine any effect the text layout may have had on the results. Furthermore, as discussed earlier, this manipulation of interface design will also help to clarify exactly at what level of text readers are satisificing. Having large gaps between the paragraphs on a page will increase the accuracy of the eye tracking data.

4.3.2 Method

Participants

Sixteen participants took part in this experiment and were paid for their participation. Ten females and six males took part, and the mean age was 23.38 years, SD 2.09, range 21-27.

Design and Materials

The experimental design was identical to Experiment 4. See Section 4.2.1 for details. The texts used were those used in Experiment 4 (see Section 4.2.2). The only difference with the texts used in this study was the layout of text on the page. Each page was divided into three natural paragraphs and although the length of each paragraph was not identical, effort was made to make them approximately similar,

while still maintaining a natural paragraph structure. The reason why there were three paragraphs in this experiment and not four, made from the four quarters in Experiment 4, was because the text did not fall naturally into more than three paragraphs. Figure 4.8 shows the layout of text on a page.

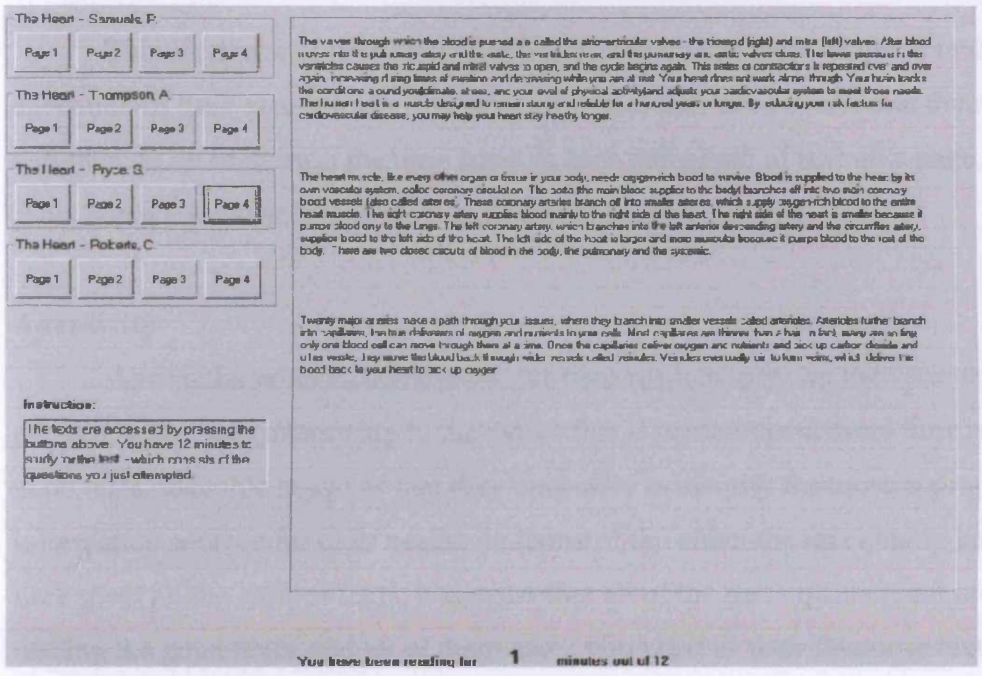


Figure 4.8 The reading interface for Experiment 5.

When the texts were 'good' texts (i.e., texts 2 and 3 for participants studying for test 1, and text 1 and 4 for participants studying for test 2) there were more answers to the corresponding test questions in the first paragraph of these than there were in the first paragraph of the 'bad' texts. Table 4.10 shows the number of facts contained in pages and paragraphs of the two good texts and the two bad texts.

Table 4.10: Number of facts per page and per paragraph for good and bad texts

Text	Test	Page				Paragraph			Total facts in each good or bad text
		1	2	3	4	1	2	3	
Good	1	8	8	9	7	12	10	9	15 or 16
Good	2	5	3	5	4	7	4.5	5.5	8 or 9
Bad	1	3	3	2	2	2	8	0	5
Bad	2	1	2	1	0	0	1	3	2

Procedure

The procedure for this experiment was exactly the same as Experiment 4. See Section 4.2.2 for details.

4.3.3 Results

This section will first discuss whether participants were adaptive in their allocation of time across texts, and what strategies they used to allocate their time. It will then go on to discuss the time spent in each paragraph of text on a page, and show hotspot plots and gaze plots to illustrate reading patterns.

Adaptivity

As mentioned in Experiment 4, the time readers spent on the various texts and pages of a text was interesting to the extent that if readers spent more time reading the good texts, then this suggests that they were able to identify the more appropriate information sources for their needs. In terms of the effect the text quality had on the time spent on the various texts, it appears that all of the participants spent more time reading the good texts, and all of them had a good text as their favourite text (see Table 4.11).

Table 4.11: Time spent in texts

Time in Good texts		Time in Bad texts	
Mean	572.92	**Mean**	137.34
S.D.	*99.04*	***S.D.***	*98.9*
Minimum	411.45	**Minimum**	0
Maximum	712.81	**Maximum**	301.23

As in Experiment 4, an analysis was conducted on the time taken to switch from the first text to the next text, depending on whether the first text was a good or a bad text (see Table 4.12). A two sample t test revealed that people switched faster when the first text was a bad text ($t(14) = 18.57, p < 0.001$).

Table 4.12: Mean time taken to switch away from first text

Text presented first		Time taken to switch to next text
Good	Mean	427.56
	S.D.	*40.83*
Bad	Mean	119.65
	S.D.	*23.1*

The average time participants spent reading a page of text was also analysed in terms of whether the participant had been presented with a good text or a bad text in the first position on the menu (although participants can choose which text to begin reading, all participants did in fact begin by reading the first text). See Table 4.13 for details.

Table 4.13: Time on a page of text with good or bad presented first

		bad text first	good text first
Mean time per page	**good**	93.74	97.68
	bad	30.34	28.68
S.D.	**good**	*8.99*	*6.27*
	bad	*3.64*	*5.05*

Table 4.13 suggests that there is no difference in the mean time spent per page of the good texts according to which texts was presented first in the menu bar ($t(14) = 1.02$, $p = .32$). The table also suggests that there is no difference in time spent per page of a bad text according to which text was presented first in the menu bar ($t(10) = 0.66$, $p = .53$); the degrees of freedom are lower because only half of the participants who were presented with a good text in the first position on the menu bar made visits to bad texts. This pattern of data is consistent with that found in Experiment 4, suggesting that the order of texts presented in the menu bar does not influence the time readers spend per page of each of the text types.

Time Allocation Strategy

The browsing patterns of all of the sixteen participants in this study can be catergorised as satisficing because the first visit to each text was the longer visit ever made to that text during the reading time. All participants began reading the text in

the first position and consulted texts in their vertically displayed order until, as the pattern of data indicates, they found a text which satisfied their information needs.

As in Experiment 4, an analysis was carried out to examine the time participants spent on each of the four pages of a text, good and bad, if visited. Figure 4.9 shows the time participants spent on page visits. The data suggests that more time was spent on each of the pages of a good text than a bad text, but that there are no differences in the time spent on pages within a text type, good or bad. An ANOVA with factors of page (1 to 4) and text type (good and bad) was carried out on data excluding when the participants ran out of time whilst reading a text, and confirmed the description of the data by revealing that there was no main effect of page ($F(3,132) = 1.16$, $p = 0.33$), a main effect of text type (good, bad), ($F(1,44) = 580.00$, $p < .001$), and an interaction between page and text type ($F(3,132) = 2.75$, $p < .05$).

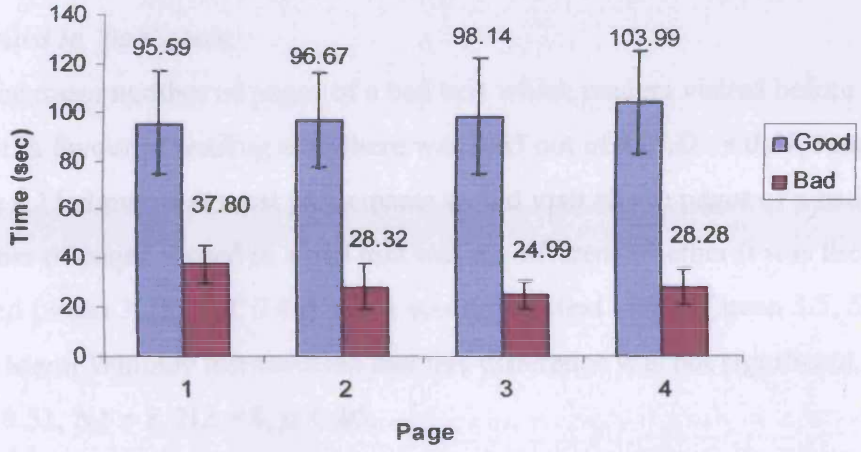


Figure 4.9: The mean time (with +/- S.D.) participants spent on pages visited for the good and bad texts in Experiment 5.

Pairwise comparisons revealed that there were differences for each page between good and bad texts (smallest $F(1,44) = 136.28$, all $p < .001$) but that there were no significant differences between any of the four pages within each text type, good ($F < 1$) and bad ($F = 2.30$). This data on time spent on pages of the good and bad texts is consistent with the pattern of data found in Experiment 4. Also, because this analysis may violate the independence assumption due to the fixed study time, the analysis was repeated with the first good and the first bad text visited only. The pattern of results was the same (no main effect of page ($F < 1$), a main effect of text

type ($F(1,24) = 359.6, p < .001$) and an interaction between page and text type ($F(3,72) = 3.62, p < .05$).

As in Experiment 4, the number of pages and texts readers visited was analysed in order to determine whether participants were exhaustive in terms of consulting all of the available texts/pages or whether they only made visits to some (see Table 4.14). Half of the participants visited all four available texts in their reading time.

Table 4.14: Unique texts visited overall

number of texts visited:	number of participants:
1	0
2	4
3	4
4	8

Pages visited in 'bad' texts:

The mean number of pages of a bad text which readers visited before deciding to reject it in favour of reading elsewhere was 3.65 out of 4 (*S.D. = 0.59*, range = 2 to 4). Table 4.15 shows that most participants would visit all the pages of a bad text. The number of pages visited in a bad text was no different whether it was the first bad text visited (mean 3.75, *S.D. 0.*46) or the second bad text visited (mean 3.5, *S.D. 0.76*). A Mann Whitney test revealed that this difference was not significant, $U = 37.0, z = 0.53$, N1 = 8, N2 = 8, $p < .05$.

Pages visited in 'good' texts:

The mean number of pages of a good text which readers visited was 2.93 (*S.D. = 1.26,* range 1 to 4). All participants visited all four pages of the first good text consulted, and then ran out of time while reading the second good text (which accounts for the lower mean pages read for good texts than for bad texts). All 1-2-or 3-page visits were to the second good text and all participants ran out of time while reading this text.

Table 4.15: Pages of texts consulted by readers

Number of pages consulted	Number of participants who made a visit of this kind	
	Bad texts	Good texts
1	0	6
2	1	6
3	5	2
4	14	16

A comparison was done between the number of pages consulted in visits to a bad text in Experiment 5 and those in Experiment 4. A chi-square test revealed that the frequencies in the two experiments did not differ significantly, $\chi 2 = 4.99$, which is < critical value of 7.81, df = 3, therefore not significant. Table 4.16 shows the order in which texts were presented to the participants and the number and order of texts participants consulted during the reading phase. As for Experiment 4, the far right columns of the table detail the number of texts visited according to the text sequence type given to the participants. Consistent with data from Experiment 4, this shows that participants who had a bad text in the first position visited *all* texts whereas participants who had a good text in the first position did *not* visit all texts. Also evident from the table is the fact that all participants visited a text only once, and always in a systematic manner from the first text in the menu working down.

Table 4.16: Order of texts presented and number of texts visited

text order (good or bad)	order texts visited	text order (good or bad)	order texts visited	Number of pages visited for each text sequence type:	
1423 ggbb	14	1423 bbgg	1423		
2314 bbgg	2314	2314 ggbb	23	2G, 2B	2B, 2G
4132 bbgg	4132	4132 ggbb	41	2, 2, 2, 2	4, 4, 4, 4
3241 ggbb	32	3241 bbgg	3241		
1243 gbgb	124	1243 bgbg	1243		
2134 bgbg	2134	2134 gbgb	213	alt GB	alt BG
4312 bgbg	4312	4312 gbgb	431	3, 3, 3, 3	4, 4, 4, 4
3421 gbgb	342	3421 bgbg	3421		

As in Experiment 4, an analysis was carried out to investigate whether participants treated a bad text differently (i.e., spent more or less time reading it) depending on whether it was preceded by a good text or not. So, the time spent on the

first bad text encountered (bad text 1) for the order of alt BG was compared with the time spent on bad text 1 for the order of alt GB (for alt BG – the mean time spent on first bad text was 113.21, *SD 18.37*; and for alt GB - the mean time spent on first bad text was 108.63, *SD 28.92*). An independent samples t test revealed that this difference was not significant ($t(6) = 0.27, p = .8$).

Time in each paragraph

As in Experiment 4, eye-tracking was used to examine more closely what participants were doing when reading within a page of text. In order to analyse how much time was spent looking at each paragraph on a page of text, the duration of readers' eye fixations in these paragraphs were analysed. Due to the fact that each paragraph was slightly different in length than the next and the area set by the eye tracking software to count eye fixations was static, extra calculations were necessary to accurately compare time spent *viewing* each paragraph. For each paragraph on the page there was a measure of overall time spent looking at it. This figure was divided by the number of lines in that particular paragraph to determine how long the participant was reading the actual text. This gave a figure for the time spent reading each line of text in the paragraph, but it must be remembered that participants may only have read the first few lines (as is evident from other recordings such as gaze replay) and not necessarily all the lines in the paragraph. Figure 4.10 shows that less time was spent on later paragraphs than earlier paragraphs for the bad texts, but that there is little variation for time spent on paragraphs within a page for the good texts. As in Experiment 4, the values were arrived at by averaging across pages for the two good and the two bad texts, the time participants spent looking at each paragraph of text *if* the paragraph was visited.

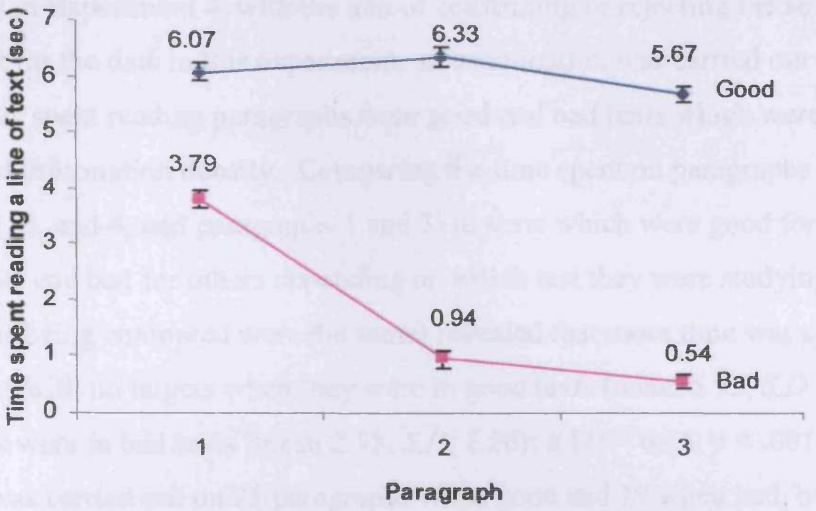


Figure 4.10 The mean time per line (with +/-S.D.) to read each paragraph for good and bad texts in Experiment 5.

Four participants out of sixteen did not visit a bad text during their reading and therefore because no comparison could be made between the time they spent on paragraphs of the good and bad texts, the times these participants spent on paragraphs of the good texts was also removed before the analysis was carried out. Inspection of Figure 4.10 indicates that more time was spent on all three paragraphs in the good texts than the bad texts, and that there were more differences across paragraphs within the bad texts than the good texts. An ANOVA with factors of text type (good and bad) and paragraph (1 to 3) confirmed this description of the data by revealing that there was a main effect of text [$F(1,11) = 508.62, p < 0.001$], a main effect of paragraph [$F(2,22) = 49.86, p < 0.001$], and an interaction between text and paragraph [$F(2,22) = 91.11, p < 0.001$]. Simple main effects analyses of good and bad texts at each paragraph reveal that for each paragraph 1 to 3 significantly more time was spent reading the good than the bad texts (smallest $F(1,11) = 153.11\ p < .001$). Simple main effects analyses within text type (good and bad) across paragraphs reveals that for the good texts there is only a significant difference between paragraph 2 and paragraph 3 ($F(2,10) = 5.03,\ p < .05$). There are no other differences between paragraphs in the good texts. For the bad texts, there were significant differences between all paragraphs ($F(2,10) = 87.25,\ p < .001$). This replicates the pattern found in Experiment 4 (see Figure 4.3, Section 4.4).

As in Experiment 4, with the aim of confirming or rejecting the search model to account for the data in this experiment, an examination was carried out on the times participants spent reading paragraphs from good and bad texts which were equivalent in terms of information density. Comparing the time spent on paragraphs (a mixture of pages 1, 3, and 4, and paragraphs 1 and 3) in texts which were good for some participants and bad for others depending on which test they were studying for (so the paragraphs being compared were the same) revealed that more time was spent on the paragraphs with no targets when they were in good texts (mean 5.95, *S.D. 1.09*) than when they were in bad texts (mean 2.78, *S.D. 2.26*); $t(42) = 6.15, p < .001$. This analysis was carried out on 25 paragraphs when good and 19 when bad, because participants did not always visit all these paragraphs when they were the bad texts. This analysis suggests that the search model cannot account for the participants' behaviour in this experiment.

An analysis was done to investigate the number of participants who looked at each of the paragraphs on a page of a bad text. Out of the 73 visits in total to pages of a bad text (this was from 12 participants because 4 participants did not visit any bad texts at all), all of these 73 visits included looking at the first paragraph and the second paragraph, and 62 of these visits included a look at the third paragraph of a bad text page. All participants visited the paragraphs in order, and only looked at later paragraphs if they had already looked at earlier paragraphs. Table 4.17 shows the frequency of participants who visited each paragraph on a page of a bad text. As in Experiment 4, a participant may have made both 2 and 3 'paragraph consulting visits' in different texts, and will therefore be counted in both of these categories above. Also, participants may re-visit texts during their reading, and therefore there is no maximum value for the number of 'paragraph-type' visits. As before, if participants made 3 paragraph visits to both of the bad texts, then they would be counted once as a participant who made this type of visit, whereas if they made a 3 paragraph visit to one bad text and a 2 paragraph visit to the other bad text then they would be represented twice in the 'number of participants who made a visit of this kind' column below.

Table 4.17: Number of visits to paragraphs on a page of bad text and the number of participants who made such 1, 2, or 3 paragraph consulting visits

Page Section	Number of visits	Number of participants who made a visit of this kind
1 paragraph only	0	0
2 paragraphs (P1,P2)	11	7
3 paragraphs (P1,P2,P3)	62	12

The same analysis was carried out for the good texts. Out of the 95 visits in total to pages of a good text (all 16 participants included), all included looking at the first paragraph, 92 looking at the second paragraph also, and 83 looking at all three paragraphs. All participants visited the paragraphs in order.

Table 4.18: Number of visits to paragraphs on a page of good text and the number of participants who made such 1, 2, or 3 paragraph consulting visits

Page Section	Number of visits	Number of participants who made a visit of this kind
1 paragraph only	3	3
2 paragraphs (P1,P2)	9	9
3 paragraphs (P1,P2,P3)	83	16

As in Experiment 4, looking at the quarters of text consulted by readers on a good page, the instances here where the reader does not look at the second or last paragraph of text occurred at the end of the reading time, suggesting again that we can assume they did not look at them because they had run out of reading time.

Gaze plots

As in Experiment 4, gaze plots were created to enable a more detailed view of the participants' scan path whilst reading. Again, each fixation is represented by a dot, and the larger the dot the longer the fixation. Two exemplary participants were chosen and their scan paths are presented below. Figures 4.11 – 4.14 show the scan path of participant 19 as they looked at the first page of each of the texts available to them. The texts were presented to participant 19 in the menu bar in the order bad, bad, good, good. They spent most of their time in the good texts and their favourite text was a good text. They visited the texts in the order of the menu bar and visited all four pages of the first bad text, two pages of the second bad text, all four pages of the

first good text, and only two pages of the last good text – but this is because they ran out of time when reading this text.

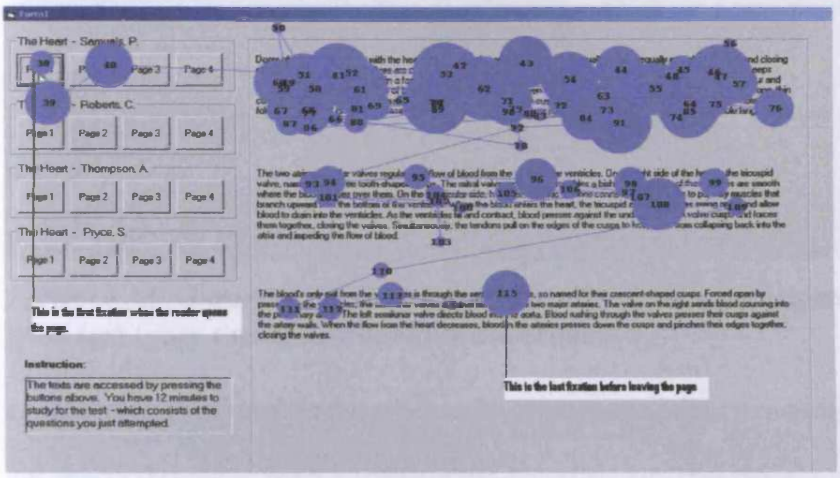


Figure 4.11: Gaze plot detailing the eye fixations when participant 19 was reading a bad text (page one of text 1) in Experiment 5.

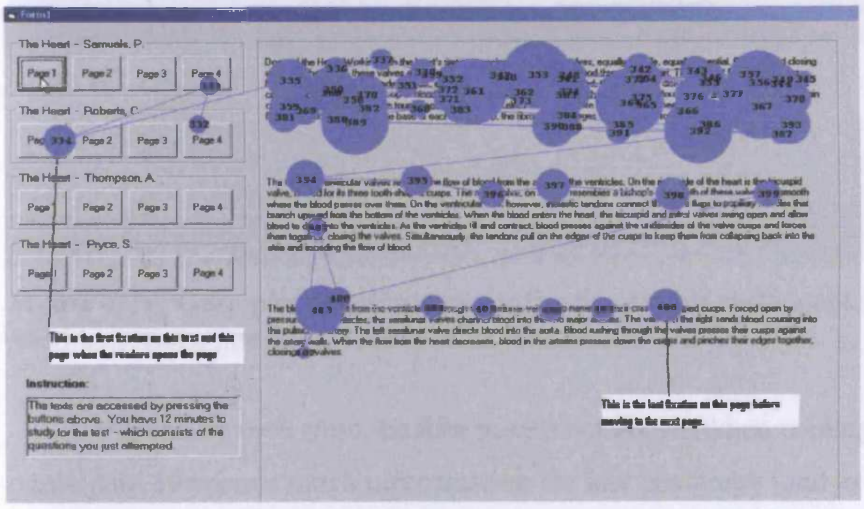


Figure 4.12: Gaze plot detailing the eye fixations when participant 19 was reading a bad text (page one of text two) in Experiment 5.

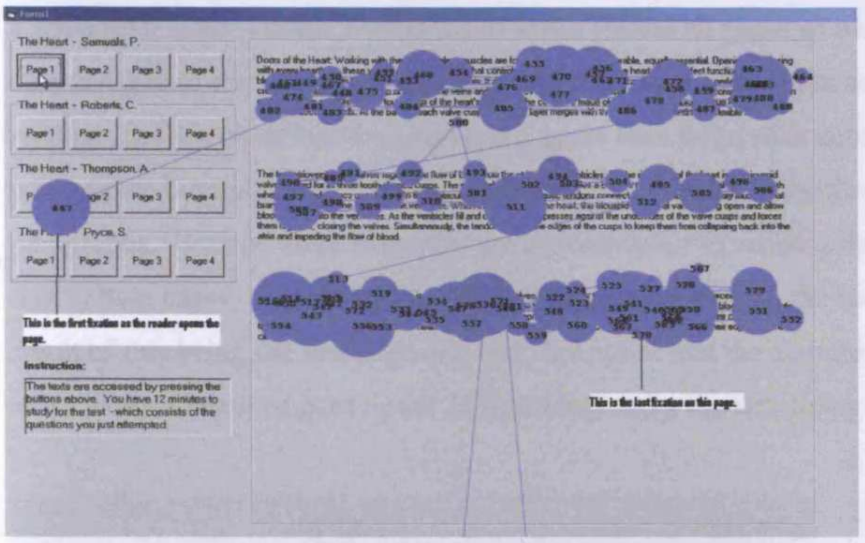


Figure 4.13: Gaze plot detailing the eye fixations when participant 19 was reading a good text (page one of text three) in Experiment 5.

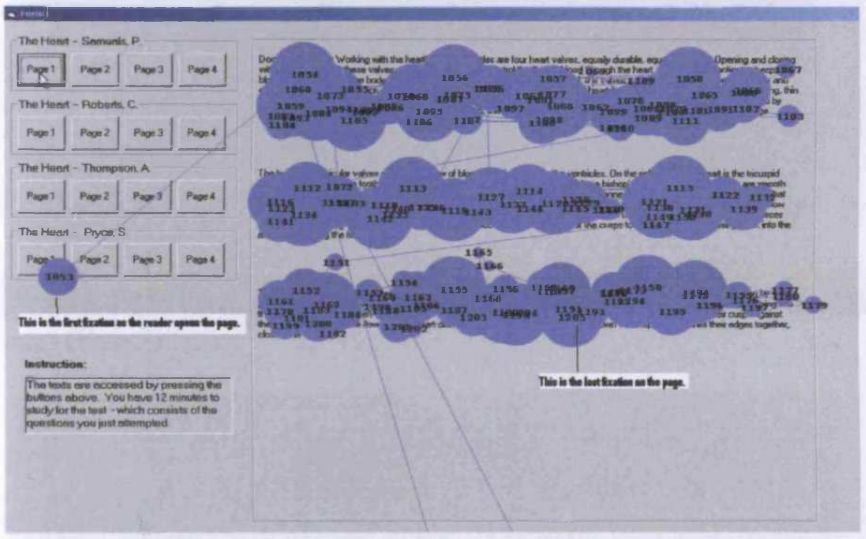


Figure 4.14: Gaze plot detailing the eye fixations when participant 19 was reading a good text (page one of text four) in Experiment 5.

As can be seen from the four gaze plots above, when reading a good text participant 19 spends much more time on the last paragraph (and to some extent the middle paragraph) than they do when reading a bad text. However, as is illustrated by the dots depicting the scan path, the reader spends a similar amount of time consulting the first paragraph of both the good and the bad texts.

Figure 4.15 shows the gaze plot for participant 21, a satisficer, reading page 4 of a good text and Figure 4.16 shows the same satisficer reading page 4 of a bad text. Participant 21 was presented with the texts in the menu bar in the order bad, good, bad, good. They spent most of their time in the good texts and their favourite text was

a good text. They visited 3 texts out of 4 and visited all pages of the first 2 texts and only 2 pages of the last text visited (but this was because they ran out of time whilst reading it). It is clear that the good text page is read fully, with equal time spent on all paragraphs, but that the bad text is skimmed over, especially for the second and third paragraphs. Both of these examples are of participant 21 reading the last page of a text in both cases, suggesting that the skimming behaviour in the bad text is not a result of this being the last page of a text, but rather that the skimming behaviour is more likely a result of participant 21 operating under the satisficing model.

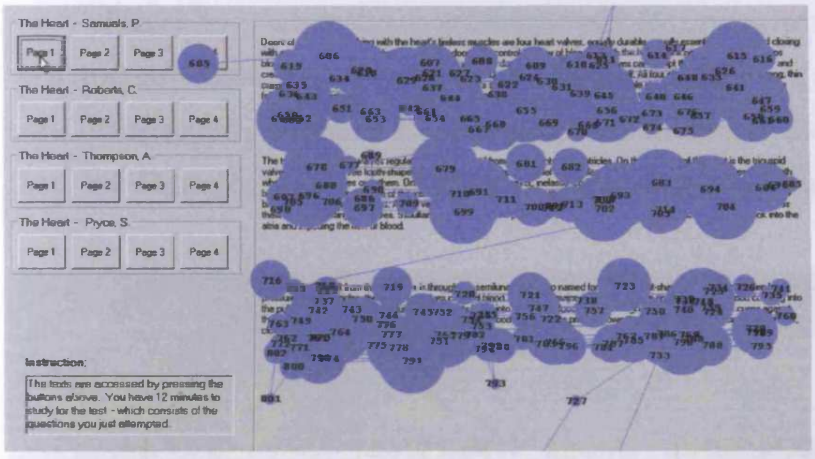


Figure 4.15: Gaze plot of participant 21, a satisficer, reading the last page of a good text (Exp 5)

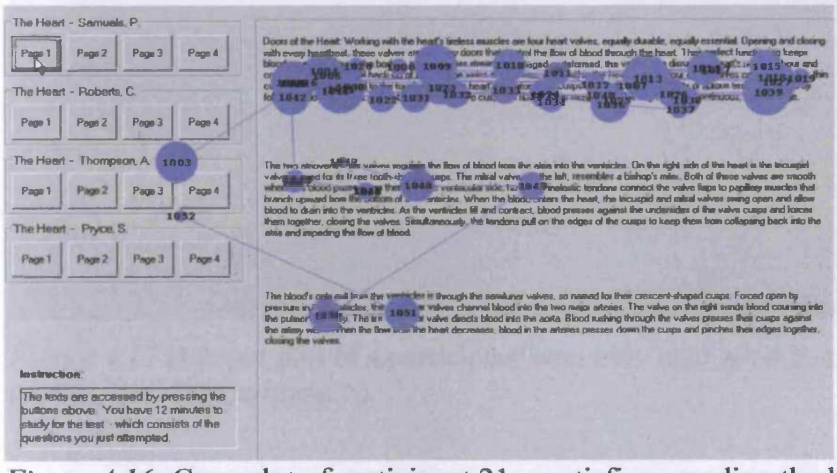


Figure 4.16: Gaze plot of participant 21, a satisficer, reading the last page of a bad text (Exp 5).

Hotspot plot

Other recordings from the eye tracking system support the argument that readers begin by reading the page normally, and then either continue to read normally if they are finding information necessary for their goal, or skim the remainder of the

page of text (either by selecting random sections in a blocked text to skim, or by skimming the first line of each paragraph) if they are not finding the information they need. One such recording is the hotspot plot, which pools across all pages read, and gives a summary of the cumulative times spent on different portions of the page during the twelve minutes reading time. Hotspot plots are being analysed in this Experiment so that they can be compared to research carried out by Neilsen (2006) who proposed that people's attention on a screen occurs in an F-shape. There therefore exist examples of hotspot plots for participants who read only good texts, but not only bad texts, as no participants read only bad texts.

Figure 4.17 below shows the hotspot plot for a participant from Experiment 5 who read only good texts. As can be seen from the intensity of colour on various parts of the screen, the reader spent as much time fixating on the last paragraph as the first. See Figure 4.19 for a scale of fixation represented by intensity of colour.

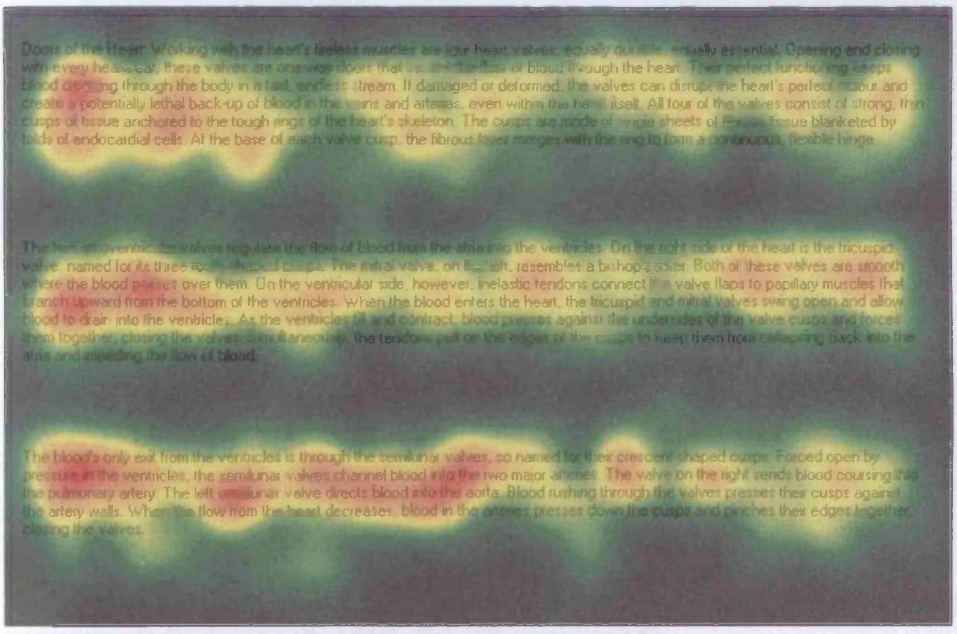


Figure 4.17 Hotspot plot of a participant who only read good texts during their reading time (Experiment 5).

On the other hand, Figure 4.18 shows a hotspot plot for a participant from Experiment 5 who read both good *and* bad texts. If Figure 4.18 is contrasted with Figure 4.17, where the participant read only good texts, it can be seen that there is less fixation on the last two paragraphs of the text than the first paragraph. This is consistent with their being read for good texts but not for bad texts.

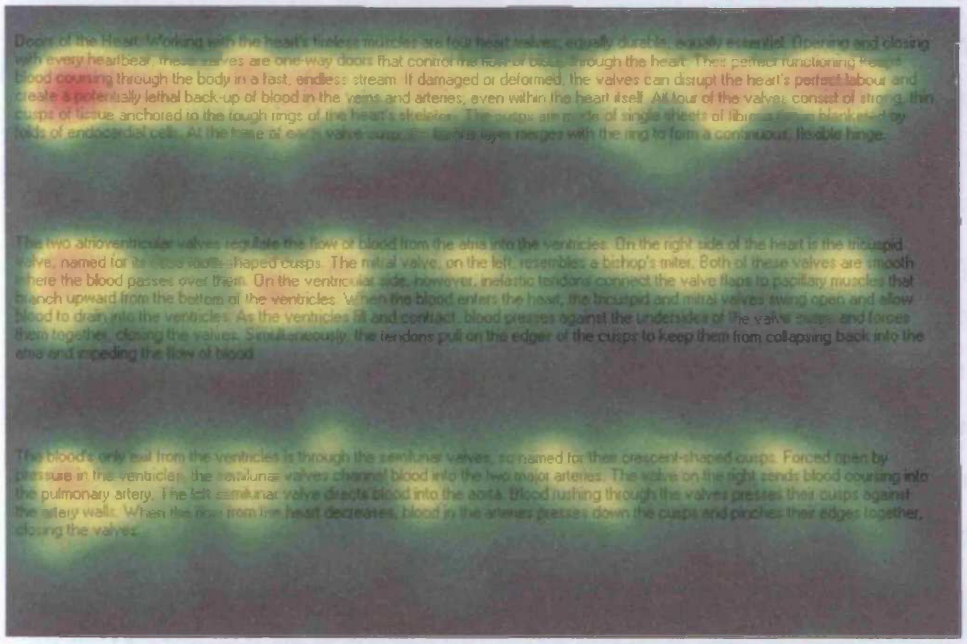


Figure 4.18 Hotspot plot of a participant who read both good and bad text throughout their reading (Experiment 5).

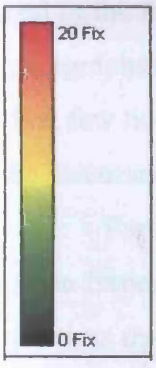


Figure 4.19 Hotspot plot fixation scale showing how the intensity of colour corresponds to the length of fixation (Experiment 5).

4.3.4 Discussion

Experiment 5 was designed in order to explore the eye tracking data in more detail when the page of text was clearly sectioned. This would also help to clarify whether participants were still satisficing and then skim-checking the rest of the page before leaving. Rather than having one single block of text per page (as in

Experiment 4 and earlier experiments), this study has shown that breaking this single block into three paragraphs does not change the time allocation strategies of readers. The browsing patterns suggest that all participants still used the satisficing strategy with participants skimming the rest of the paragraphs of bad pages before leaving the page (in line with the suggestion that they are satisficing then skim-checking).

Whilst there was no significant difference in the time spent reading paragraph 1 and 2 or paragraph 1 and 3 in a good text, a difference was found between paragraph 2 and 3. However, this could be, as in Experiment 4 where there was a difference between the last two quarters, a result of the reader becoming faster at reading as they progress down the page. In any case, since the times spent reading the paragraphs in the good texts do not decrease linearly, it can be argued that participants were not reading then skimming as they would if the text was bad.

The times spent in paragraphs in the bad texts were all significantly different. The time readers spent in the first paragraph of a bad text was much more than they spent in the last paragraph, and the difference in time decreases linearly with the paragraph order. This clearly suggests that readers were beginning to read the page as normal, but, as they found that they were not encountering the facts required for their test in the first paragraph, they were making judgements about the remaining paragraphs on that page and skim reading the rest of the page by briefly consulting the first few lines of the remaining paragraphs before moving to the next page. This will be discussed in more detail in Section 4.4.

Participants visited, on average, 3.65 pages out of 4 when reading a bad text. As in Experiment 4, this suggests that for most participants the unit of text, or patch, was less than or equal to the *page*, since readers did not seem prepared to reject an entire text on the basis of just one page. In line with the data from Experiment 4, Table 4.20 showed that participants consulted each text only once, if at all, and always in a systematic manner from top to bottom on the text menu bar. This further strengthens the argument that readers judge the relevance of a text as they read through the text, as this data shows that they do not revisit texts, and therefore are not judging texts before they read to learn.

Experiment 4 showed that participants read the first few lines of a page at a normal rate and if they decided that they were not getting the information they required from the page, then they would skim read the remainder before moving onto the next page. Experiment 5 was designed to investigate whether, if the reading

interface was designed such that three paragraphs appeared on each page, the reader would read the first paragraph and then decide to stay or leave, or look at all of the paragraphs before making a decision to stay reading or to leave the page. Readers seemed to adopt the same approach in Experiment 5. That is, that they began reading the page normally and skim read the remainder before deciding to stay or leave. The only difference with the paragraph interface was that the readers used the paragraph structure as a guide of where to skim read in order to check the rest of the page. The gaze plots show that participants looked at the beginning of a paragraph to make their decision to stay or leave, which is similar behaviour to skimming sections in a blocked text, just using the physical properties of the text as a guideline for their skimming. Since the eye tracking data indicates that readers were looking at all or most of the paragraphs of a page, albeit very briefly in the case of the bad text pages, this suggests that the unit of text which was being judged for the stay/leave decision is less than or equal to a paragraph.

An analysis of how many visits were made to each paragraph of text, and how many participants made such 1, 2, or 3 'paragraph consulting visits' was carried out in order to better understand the skimming behaviour of the majority of the participants (see Tables 4.12 and 4.13). This data strengthens the argument that readers will generally consult most of the text sections on a page before rejecting it and moving away. This suggests that the level of text at which readers are prepared to satisfice is the paragraph, telling us that satisficing behaviour involves small rather than large units of text.

In terms of the hotspot plots, which measure the dwell time of readers' eye fixations on various sections of the text, we can see how more time is spent on the latter sections of a good text than on the latter sections of a bad text. Due to the fact that the hotspot plot in Figure 4.17 includes data for both the reading of good and bad texts, it is perhaps not as reliable as other measures, such as the gaze plot, in supporting our argument for readers' skimming behaviour. The hotspot plots in Figures 4.17 and 4.18 do, however, provide us with other information about the nature of reading which cannot be deduced from the gaze plots. The hotspot plots indicate that readers fixate more on the left of the screen than the right of the screen. This supports the research into how readers view web pages in an F-shape (Nielsen, 2006) or, termed differently but relating to the same phenomena, how readers scan items in an inverted triangle shape, Google's Golden Triangle (Enquiro search solutions,

2005). Two of the three websites tested by Nielsen had content menu bars on the left of the screen. All participants in the experiments in this thesis were native speakers of English, and therefore would naturally read from left to right. However, in line with previous research (Neilsen, 2006; Enquiro, 2005), it does suggest that more salient information in a text could be placed on the left of the reading screen (discounting the area taken up with the menu bar) to increase the chances of the reader seeing this information. In his review of eye movement research over the past few decades, Rayner (1998) reports that the first eye fixation on a line of text tends to be longer than the rest, and the last fixation shorter, supporting the images found in our hotspot plots. This would also explain Neilsen's (2006) concept of the F-shaped pattern of user's attention on a page.

The data presented in the gaze plots provided a visual image of the readers' scan path which supports the argument that readers begin by reading the page normally, and then either continue to read normally if they are finding information necessary for their goal (in a good text), or skim the remainder of the page of text (perhaps either by selecting random sections in a blocked text to skim, or by skimming the paragraphs of the text) if they are not finding the information they need (in a bad text).

4.4 General Discussion

The preceding pair of eye tracking studies indicate that when reading under time pressure, with a goal of learning for a specific test, readers do not necessarily 'read' all of the text, nor do they quickly reject it and move onto another text, but rather they satisfice and then skim-check the remainder of the page before leaving the page. Rayner and Pollatsek (1989) suggest that skimming is a type of reading whereby the text is scanned without being deeply comprehended. Considering that not all of the information contained in the four texts is relevant to the task goal, it is not important that some information is missed or not comprehended by the readers using this strategy of satisficing then skim-checking before leaving. The data from Experiments 4 and 5 show that the pattern of behaviour displayed by participants does not fit directly with the satisficing model, but rather it appears that participants are satisficing and then skimming. One interpretation of the strategy being used by participants is that they are 'satsificing then skimming', which refers to the way in which participants skim text which is not deemed good enough for their learning

needs (according to whether the text answers the test questions), shown by very brief visits to patches which would not allow the reader time to read the text fully. Another interpretation of their behaviour is that they are using a satisfice then skim-check strategy, whereby they use paragraphs/sections as a patch with which they can satisfice, but then skim-check the remainder of the paragraphs/sections on the page before moving to the next page. The satisficing model would predict that people read a patch and then leap to the next patch, and it would predict that a patch is either read or not read rather than skimmed.

Both experiments show that the number of pages of a bad text visited by participants is the same whether it is the first bad text visited or the second bad text visited. This suggests that the criteria which readers were using to judge the value of a text was not changing for the second bad text visited depending on the experience of the first bad text visited. This shows that participants were treating each text as a separate unit and evaluating the text on its own merits, since they did not let their experience of a previous (bad) text effect their judgement of the current (bad) text. This also reinforces the claim that for most participants, satisficing occurs at small text units, and not on the basis of large sections of text(s).

The data in Experiments 4 and 5 not only supports the view that readers are using a satisficing strategy, but also provides evidence for the fact that readers are *not* sampling, and they are not displaying behaviour which would fit with linear reading or random sampling either. If participants were searching for the optimal text (this would be one of the good texts in this experiment) through sampling, then they would have visited all texts briefly at the beginning of their reading time and chosen the best one to exploit. The data did not support the suggestion that this kind of behaviour is being played out. Similarly, if participants were reading the texts in a linear fashion then the data would indicate that they spend the same amount of time on all texts visited (taking into account that the last text visited may be shorter because of the time limit running out), and accelerated linear reading would indicate that the participant had spent proportionally the same amount of time on each page and text available. The data does not support the idea that participants were sampling randomly either, since this would be marked by indications that they had jumped around the texts and pages in no particular order. The data from Experiments 4 and 5 clearly show a more systematic, strategic approach than these alternative strategies would elicit.

Various data from these experiments also provide evidence for the fact that readers were not following a search model of reading. The fact that participants did not always visit every page of a text indicates that they cannot be using the search model, as this model predicts that texts are scanned uniformly until target information is encountered. Also, the search model predicts that all of the text is scanned, but the gaze plots show that sometimes sections of text are skipped over completely, especially in the later paragraphs of bad texts. This would not happen if readers were following the search model, and presumably the model would also predict that the time spent on paragraphs of a bad text would not differ very much (because the participant would have scanned through and not paused very often as there were very few target facts), which is clearly not the case given Experiment 5 results. Furthermore, if participants were following the search model of scanning the texts until target information is found, it follows that, as discussed in Section 4.3.3, paragraphs in the good and bad texts which were equivalent in terms of target information density should be scanned over in the same amount of time. A paragraph of a good text which contained no target information should be scanned over just as quickly as a paragraph in a bad text which contained no target information. The data showed that more time was still spent on paragraphs in good texts which had no target information than paragraphs in bad texts which had no target information. This argues against the possibility that the search model can account for the data in Experiment 5.

It was previously discussed that the existence of scent cues (Pirolli & Card, 1995) aids the reader in finding the information which s/he desires. However, what is evident from the current eye-tracking studies is the tendency for readers to allocate time across texts when learning for a test by skimming the text and rejecting paragraphs which do not appear to contain the information required. It appears that readers are searching by rejecting, rather than searching by following cues, and therefore after deciding to reject a certain section of text, they simply move onto the next section of text. In this way, design for skimmability need not only be a matter of clear headings and other scent like cues (Neilsen 1997), but also a matter of making the various patches of a document easily perceived and then allowing the moves from one patch to another to be readily made (Reader & Payne, 2007).

Through the use of eye tracking, Experiments 4 and 5 provide evidence of how readers reject a page of text if it is not meeting their threshold level of

satisfaction in terms of information gain. If it is known that readers read the first paragraph, or section of a text normally, and then use the paragraph structures to skim read the remainder of the page, then this could inform text designers. It would make sense to put salient information in the first paragraph/section of a text, and then start subsequent paragraphs with important or essential information, since it seems as though it is the content at the start of a paragraph which is responsible for holding the readers interest, or not. Indeed, it has been suggested in previous research that the first and last sentences of a paragraph are indicative of its content (Masson, 1982). This type of research has implications for online text designers, whose interest may focus on the sections of a text to which readers initially direct their attention. It would be these areas of initial interest where essential or primary information should be placed.

However, it must be acknowledged that there are certain limitations with regards using eye tracking as a measurement of where on the page participants are reading. In terms of how useful eye trackers are at informing about what readers are doing when reading, they can tell where the user is looking on the screen, and by tracking the density and frequency of user's eye fixations, can determine whether the user is reading the text or scanning it. They can also tell us how long a user has been looking at certain sections of a web page and whether the user is looking for a specific item or not (this is measured by pupil diameter: it is wider when users are unsure of the word for which they are searching); Schroeder, 1998: pupil diameter was not measured in these experiments. However, although eye tracking can tell us where a user has directed their eyes, there is no guarantee that the user has actually *seen* what they are looking at, or read it. Similarly, it cannot guarantee that a user did not see something just because they did not fixate their eyes upon it, as peripheral vision plays a role which is not recorded, due to the fact that only central vision (the part of the visual field which discriminates fine detail) is tracked by the eye tracker (Schroeder, 1998). For example, in his studies Schroeder (1998) found that users rarely looked directly at the scroll bar at the right of the screen. Instead, their eyes gazed to the left of the scroll bar, which indicates that they did not look directly at it even when using it, leading to the assumption that peripheral vision must play a role. Eye tracking cannot tell us *why* a user is looking at a particular section of text or page, either (Schroeder, 1998). Although these limitations are not thought to be too

restricting for the purposes of the current experiments, it is necessary to recognise their existence.

Although it has been established thus far that when faced with too many sources to consult in the time available people will adopt a satisficing strategy in order to adaptively allocate their time, it is not certain whether this tendency to satisfice is limited to the information environment in which information seekers find themselves. In my final experiment I will examine how participants' time allocation strategies are affected by their knowledge about the properties of the texts available to them. It asks whether or not people will change their strategy when they are aware that there is one information source which is much better than all the rest, which may be more difficult to find using a satisficing strategy.

CHAPTER 5

Experiment 6 – Manipulation of text quality awareness

5.1 Introduction

The experiments in Chapters 2, 3 and 4 have all suggested that when faced with too many texts to read in the time frame available, readers have a strong tendency to adopt the satisficing strategy in an attempt to allocate their time to the information source which will best help them to complete the experimental task. Indeed, even in Experiment 1 where some participants were *not* under time pressure to study, their time allocation behaviour still reflected the satisficing strategy. Further to the discovery of the tendency for participants to use the satisficing strategy, it has been shown that this strategy is adaptive in the sense that most readers managed to allocate most of their time to the information sources (texts) which contained more rather than less of the information they needed to complete the post-reading test (Experiments 2 - 5).

Although it has emerged that the behaviour shown by most participants under the experimental conditions studied here, is indication of the satisficing strategy, it is not known whether the general tendency to behave in this manner has been due to the experimental set-up or to the specific instructions given to the participants. The tendency to satisfice rather than sample (or indeed use any other kind of strategy, such as linear reading, random searching, or a search model) when allocating time across the texts could be due to the fact that the situations to which readers have been exposed have been situations for which the satisficing strategy would be the better strategy to adopt. When some or all of the texts are good enough, and there are no differences between these good enough texts, then the satisficing strategy is more adaptive than the sampling strategy because more of the readers' time can be spent reading instead of judging. Or, the preference for satisficing could simply reflect the readers' general tendency for one strategy over another. As mentioned earlier, this is not to suggest that satisficing and sampling are the only two time allocation strategies that may exist. Rather, these are the two strategies to which the behaviour observed in previous experiments best fits. Other strategies, such as linear reading or using a

search model, are possibilities, even though this behaviour was not evident in the population samples for the experiments in this thesis.

In order to ascertain whether it is the experimental set-up which has encouraged the readers to use the satisficing strategy over any other strategy, a situation was created for Experiment 6 in which satisficing would *not* be the best strategy for the reader to employ, in terms of information gained. In order to create this situation one text was much better than all the rest. That way, assuming an intermediate level threshold in terms of information gain, the satisficing strategy may result in a reader settling on a 'good enough' text and spending most or all of their time reading that, not knowing about or finding the 'best' text. For Experiment 6, a situation was therefore created whereby some texts were 'good enough' in that they answered some of the questions required for the post-reading test, one text was the 'best' because it answered all of the questions, and one text was 'bad' because it answered only one of the questions. Again, the labels 'best', 'good enough' and 'bad' given to the texts are not indicative of their linguistic composition or readability, and only indicate how useful or relevant the texts are for the experimental task. Under these circumstances, an alternative strategy, such as sampling, may be more beneficial to the reader because they would sample all of the available texts before reading to learn, and in doing so would locate the 'best' text and choose to read that one, assuming their ability to accurately judge the texts. The main aim of this study was to contrast the time allocation behaviour of participants who were told about the existence of this 'best' text (the informed condition) with the behaviour of those participants who were not told that a 'best' text existed (the un-informed condition). The experiment would thus give some indication of whether manipulating participants' beliefs about the texts would influence the strategy they would use to allocate their time across the texts. It was predicted that participants in the informed condition would sample more than those in the un-informed condition, who were predicted to behave as participants in previous experiments, thus satisficing.

There has been no evidence in previous experiments that people who satisfice have better learning scores than readers using any other time allocation strategy (e.g., sampling) because the number of samplers has been too small for a comparison to be made. A lack of difference in learning scores could, however, be due to the nature of the task and the quality of the available texts, or indeed the possibility that participants have problems remembering the questions from the pre-test. If one text is

much better, then readers may find this text more easily and quickly by using a sampling strategy rather than the satisficing strategy, and therefore samplers may have a higher learning score (because they have spent more time reading the best text). Experiment 6 will provide some evidence for or against this idea. It was predicted that Experiment 6 would show that if readers use a satisficing strategy when there is one text that is much better than the rest (one information source that is perfect for their needs and several that are 'good enough'), this may not result in the best learning score (unless the best text is presented first, in which case the participant would most likely read it first).

5.2 Method

Participants

Twenty-four participants, all of whom were Cardiff University Psychology undergraduates, received course credit for their participation. There were 22 females and 2 males, and ages ranged from 18 to 28, mean 20.04, *S.D. 2.12*. Participants were alternately assigned to one of the two conditions.

Design and Materials

Experiment 6 investigated both between- and within-participant factors. The dependent variables were a) the browsing behaviour of the participants, as measured by the time they spent in each text and the number of visits made to each page/text, and b) the learning score (calculated by comparing the pre-reading test score and the post-reading test score). The independent variables were a) the text quality (good enough, bad, or best), and b) the two conditions in the study, informed about the existence of a 'best' text, and un-informed about the 'best' text.

Participants were asked to learn for a test on the topic of the Milky Way. The test comprised sixteen questions, each worth one point for a correct answer. Two example questions are below (See Appendix H for the full test):

Question: Who was the first person to claim that the Milky Way consisted of distant stars?
ANSWER: Democritus

Question: What blocks our view of most of the Galaxy?
ANSWER: interstellar dust

Four instructional texts on the Milky Way were used in this study. This topic was chosen as it was thought unlikely that participants would have much knowledge about the Milky Way, and equally that it was not too difficult that participants would not be able to learn about it through their reading of the texts. The four texts were constructed by taking information from various websites and altering the content of each text to ensure that they were different enough in relevance (in terms of how many test questions they answered). One text was 'bad' as it only answered one question, which was located on page 3. This was to fit in with the natural flow of the text, and to avoid having it located on page 1. Two of the texts were classed as 'good enough' as they answered seven questions out of sixteen (both texts answered the same seven questions as each other), two on pages 1, 2, and 3 and one answer on page 4 for each text. One text was 'best' as it answered all 16 of the questions spread over the four pages. Seven answers were on page 1, two each on pages 2 and 3, and five answers on page 4. The reason why the distribution of these facts was not even over the four pages of the text was that it seemed important to maintain a natural flow in the text, and moving facts to make the distribution even would have disrupted the coherence of the text. All texts were between 1259 and 1459 words in length, and consisted of four pages. The text was written in natural paragraphs, with the beginning of each paragraph slightly indented. The properties of the text font and size, and the screen size were exactly the same as in previous experiments (see Section 2.2.2 for details). Figure 5.1 shows the design of the reading interface.

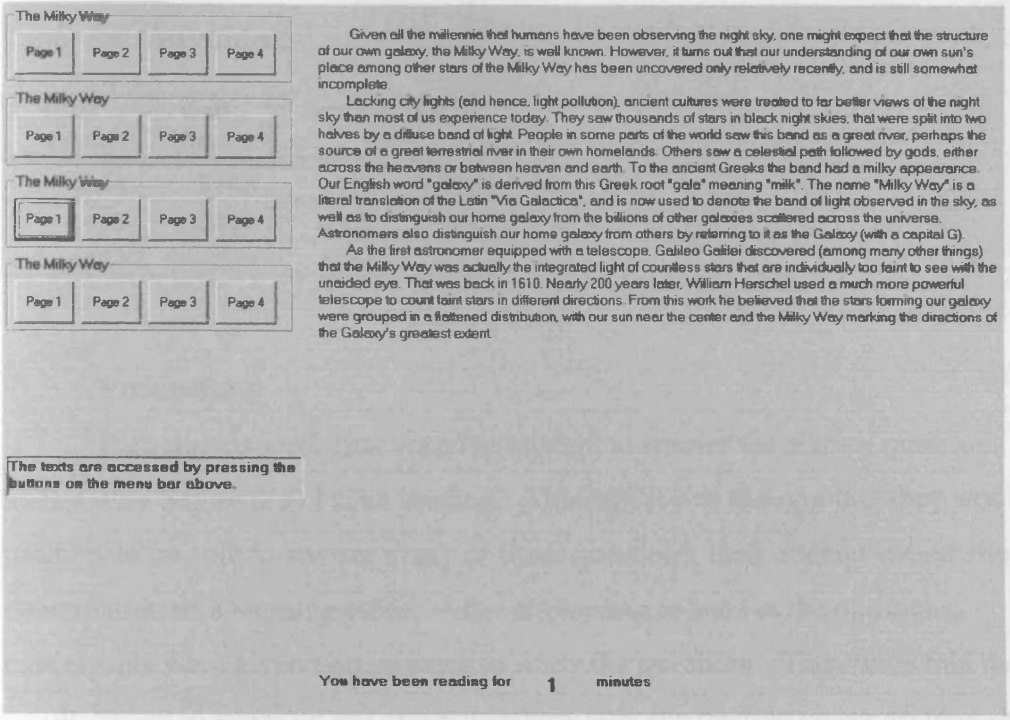


Figure 5.1: The reading interface for Experiment 6

The order in which the texts were available on the menu bar was varied. There were twelve orders. Out of the twenty-four possible orders of four texts, the twelve orders where text D always appeared before text C were selected. Since these two texts (C and D) were the 'good enough' texts, and there were no differences in quality between these texts, it was decided that provided that the order of the best, the bad and the two good enough texts was counter-balanced, then no order effects should occur. The best text was included as a first text on the menu bar for 25% of the participants despite the fact that if participants began reading the first text then sampling would not be useful. This was because the participants may begin reading any of the four texts first, and previous experiments have shown that sometimes participants did begin reading the text in the first position on the menu bar (e.g., in Experiments 2, 3, 4 and 5) and sometimes they did not (e.g., in Experiment 1). The text orders used are shown in Table 5.1, each order being used once in the informed and once in the un-informed conditions. There was an indication at the bottom of the screen telling participants for how long they had been reading.

Table 5.1: Text orders

Best first	Bad first	Good enough first	
1243	2143	4123	4231
1423	2413	4132	4312
1432	2431	4213	4321

Procedure

Participants were first asked to attempt to answer the sixteen questions on the Milky Way before they began reading. Although it was thought that they would be unlikely to be able to answer many of these questions, their attempt would allow computation of a learning score. After attempting to answer the questions, participants were given two minutes to study the questions. They were told that they would be tested again on the same questions after the reading phase, so they should read with a view to learning for this test.

Once the two minutes study time had passed, participants went on to the reading stage of the experiment. Participants were given twelve minutes to read as many or as few of the texts as they wished, and they were advised that they could allocate their time across the texts in any way they felt best supported their learning. Those participants in the 'informed' condition were told at the beginning of the experiment that while all of the texts may be relevant to the test, there was one text which was much better than the rest. They were not told about the presence of the bad text because the intention was for them to allocate time to the best text, and under the satisficing rule they would reject a bad text without needing to be warned about it, whereas they may have settled on a good enough text if it met their minimum criteria in terms of information gain. Participants in the 'un-informed' condition were not told anything about the relevance of the texts to the test questions. The texts were accessed via a button menu system at the left of the reading screen to enable participants to visit any page of any text at any time.

5.3 Results

This section will be divided into three sections: adaptivity, time allocation strategy, and learning score. The main interest in this experiment was to identify whether the manipulation of instruction had any effect on the time allocation strategy

of the participants, and therefore this section will focus on the differences between conditions (informed and un-informed) for participants classified as samplers or satisficers, rather than on the descriptive data examined in previous experiments.

Adaptivity

In order to identify whether participants were being adaptive in their allocation of time across the four texts, analyses were carried out to examine the time they spent reading the various texts (best, good enough, and bad). If participants were allocating time adaptively, then the data would show that they spent more time in the best text. Table 5.2 presents data for each condition concerning the amount of time participants spent reading certain texts.

Table 5.2: Total time (sec) spent on texts for each condition

Condition		Total time on best text	Total time on bad text	Total time on a good enough text	mean time per page	% time in favourite text
INFORMED	Mean	399.18	84.44	113.61	46.35	64.59
	S.D.	*224.33*	*104.35*	*80.62*	*29.00*	*22.77*
UNINFORMED	Mean	276.39	100.21	167.96	31.30	49.28
	S.D.	*175.42*	*71.84*	*81.08*	*10.99*	*21.01*

As can be seen from Table 5.2, participants spent more time reading the best text in both conditions and the least time reading the bad text. An ANOVA with factors of text (best, good enough (time spent on both good enough texts was divided by two to get an average), and bad) and condition (informed and un-informed) confirmed this description of the data by showing that there was a main effect of text ($F(2,44) = 15.28, p < .001$), no main effect of condition ($F(2,44) = 1.97$), and no interaction between text and condition ($F(1,22) = 2.48$). Pairwise comparisons revealed that the differences between the best and the bad texts and between the best and the good enough texts were both significantly different ($F(2,21) = 11.22, p < .01$) but the difference in time spent reading the bad and the good enough texts was not significantly different.

The following table shows the times participants spent on the various texts according to not only which condition they were in but also according to what

strategy they were categorised as using. The samplers cannot be compared because there were no samplers in the un-informed condition, but a comparison of the time spent in each text was analysed for participants classified as satisficers. The table suggests that more time was spent in the best text for both conditions, but that there are no differences between conditions. An ANOVA with factors of text (best, bad, good enough) and condition (informed, un-informed) confirmed this by revealing that there was a main effect of text ($F(2,26) = 7.41, p < .01$), no main effect of condition ($F(1,13) < 1$), and no interaction ($F(2,26) = 2.26, p = .13$). An analysis was also carried out to compare the times spent on the texts within the informed condition between the samplers and the satisficers. The table suggests that there were no differences between samplers and satisficers. An ANOVA with factors of text and strategy (samplers, satisficers) confirmed this description by revealing a main effect of text ($F(2,18) = 10.17, p < .001$), no main effect of strategy ($F(1,9) < 1$) and no interaction ($F(2,18) < 1$).

Table 5.3: Time (sec) spent on text according to condition and strategy used

		Mean total time *(S.D.)* spent on:		
Condition	**Strategy**	**Best**	**Bad**	**Good enough**
informed	**samplers**	428.42	66.09	217.37
		174.47	*66.89*	*136.88*
	satisficers	441.34	53.53	215.59
		226.94	*59.31*	*197.77*
un-informed	**samplers**	0.00	0.00	0.00
		0.00	0.00	0.00
	satisficers	267.26	110.09	335.41
		175.42	*79.84*	*167.84*

An analysis was also conducted to compare the average length of a visit to the best and the bad texts. The data shows that the average length of a visit to the best text was longer than the average length of a visit to the bad text (see Table 5.4). An ANOVA with factors of text and condition confirmed this by revealing that there was a main effect of text ($F(1,22) = 9.48, p < .01$), no main effect of condition ($F(1,22) < 1$), and no interaction between text and condition ($F(1,22) = 1.4$).

Table 5.4: Average length of visits to best and bad texts

Text	Informed		Uninformed	
	Mean time *(s.d.)*	**Mean frq. of visits**	**Mean time** *(s.d.)*	**Mean frq. of visits**
Best	**225.81** *(167.60)*	**1.83** *(1.03)*	**167.87** *(183.18)*	**2.17** *(0.94)*
Bad	**46.03** *(56.59)*	**1.67** *(1.23)*	**88.01** *(77.08)*	**1.25** *(0.62)*

The number of participants who spent most of their time in the best, good enough, and bad texts was not significantly different across conditions ($\chi 2 = 2.15$, df = 2): see Table 5.5.

Table 5.5 Number of participants who spent most of their time in each of the text types according to condition and strategy used.

Spent most time in:	Samplers		Satisficers		Residuals		Total
	Informed	Un-informed	Informed	Un-informed	Informed	Un-informed	
Best text (1)	3	0	4	4	0	1	12
Good enough texts (2)	2	0	2	5	0	2	11
Bad text (1)	0	0	0	0	1	0	1
Total	5	0	6	9	1	3	24

Closer analysis of the seven participants in the un-informed condition who spent most of their time in the good enough texts, showed that six of these participants spent most of their time in the first text on the menu bar, which happened to be a good enough text.

In order to determine how much time was spent on the best text early on in the reading phase, an analysis was done to look at the time participants spent reading the best text in the first third of the experimental time. This time frame was chosen as the first 240 seconds as it is the time during which they must visit all available texts at least once in order to be classified as displaying sampling behaviour. Some participants were presented with the best text in the first position on the menu bar, and so these participants were excluded from this analysis. Data on the remaining participants revealed that the time readers spent in the best text during the first third of the experiment was higher for those in the informed condition (mean = 76.08, *SD* =

89.71) than for those in the un-informed condition (mean = 2.4, *SD* = *7.*): $t(16) = 2.45, p < 0.05$.

Similarly, looking at the length of the very first text visit provides an indication of whether the participants were allocating time adaptively across the texts. Those participants in the un-informed condition spent significantly longer on their first text visit than did participants in the informed condition ($t(22) = 2.00, p < 0.05$): see Figure 5.2.

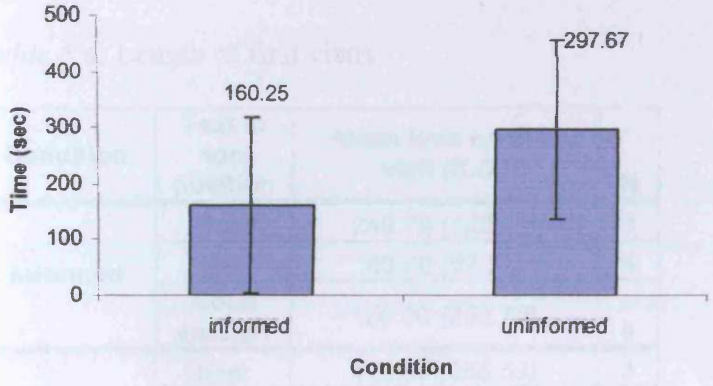


Figure 5.2 The time in seconds (with +/- S.D.) participants spent on their very first text visit for informed and un-informed conditions for Experiment 6.

Indeed, these times could be affected by which text appeared first on the menu bar, but an analysis of condition and text type appearing first (best, bad, good enough) revealed that participants in the un-informed condition spent longer on their first text visit regardless of whether this text was the best text (un-informed mean 419.0 (*S.D. 256.6*), informed mean 240.8 (*S.D. 123.3*)), the bad text (un-informed mean 184.7 (*S.D. 75.4*), informed mean 60.0 (*S.D. 67.6*)), or a good enough text (un-informed mean 293.5 (*S.D. 108.4*), informed mean 180.0 (*S.D. 222.7*)); An ANOVA with factors of text order (best, good enough, bad) and condition (informed, un-informed) revealed that there was a main effect of condition, $F(1,18) = 4.81, p < .05$, no effect of text order $F(1,18) = 3.26$, and no interaction, $F < 1$.

Ten participants in the informed condition began by reading the text in the top position on the menu bar. The mean length of first visits for each text read first is presented in Table 5.6. The two participants in the informed condition who had a

good enough text in the top position, but did not begin by reading the text in the top position, read the 'best' and the 'bad' texts first, which accounts for the different N values in the table. Since there were two good enough texts and only one each of the best and the bad, then in order to make the frequency of text first proportionate to the number of texts of that kind available, then the good enough text appeared in first place twice as many times as the best and the bad text first. It is interesting to consider the time participants spent on their first visit to the different text types because if they spent more time on a first visit to the 'best' text than any other text, then this would imply that they are behaving in an adaptive manner.

Table 5.6: Length of first visits

Condition	Text in top position	Mean time on first visit *(S.D.)*	N
Informed	Best	240.75 *(123.29)*	4
	Bad	60.00 *(67.57)*	4
	Good enough	180.00 *(222.70)*	4
Uninformed	Best	419.00 *(256.58)*	3
	Bad	184.67 *(75.38)*	3
	Good enough	293.50 *(108.38)*	6

The data in Table 5.6 indicates that more time was spent on the first visit to the best text than the first visit to the bad text. A 2(condition) by 3(text) ANOVA confirmed this description of the data by revealing a main effect of text, $F(2,10) = 5.2$, $p < .05$, but no main effect of condition, $F(1,5) = 2.4$, and no interaction between condition and text type, $F < 1$.

Time Allocation Strategy

Overall, fifteen participants displayed a satisficing strategy, five displayed a sampling strategy, and four displayed neither and were therefore classified as residuals. Table 5.7 shows the number of satisficers, samplers and residuals in each condition.

Table 5.7: Time Allocation Strategy

Condition	Satisficers	Samplers	Residuals
Informed	6	5	1
Uninformed	9	0	3

A Fisher-Yates exact text revealed that the number of samplers found in the informed condition was significantly higher than the number of samplers in the un-informed condition ($p < .05$). If those participants in the informed condition who had the best text at the top position in the menu bar are eliminated, then there are four satisficers, four samplers, and one residual in the informed condition. In the un-informed condition, eliminating the participants who had the best text in the top position, there are seven satisficers, no samplers, and two residuals. The difference between the number of satisficers and samplers *within* the informed condition is not significantly different. In the un-informed condition there were no samplers at all, even when participants with the best text first were eliminated.

A closer look at the participants who were classed as residuals revealed that the three in the un-informed condition satisficed with three out of the four texts they visited, and the residual in the informed condition visited three texts and only satisficed with one of them. Thus, the three residuals in the un-informed condition were satisficing most of the time. The one residual in the informed condition satisficed with one text and almost with the other two, but the earlier visits to these texts were slightly shorter than later visits (an early visit of 2 seconds for one text and an early visit which was 47 seconds shorter than a later visit to that text). Although they did not fit directly into the satisficing category, their behaviour was more towards a satisficing behaviour that it was towards a sampling, linear, or random search behaviour, as their visits to texts and pages was still systematic. Four out of the five samplers spent most of their time in the best text. The other sampler spent most of their time in a good enough text (time in good enough text 40.1% and time in best text 36.6%).

Looking at the number of pages and texts visited in each condition also gives an indication of the kind of strategies the participants were using to allocate their time across the texts. Table 5.8 shows the number of text and page visits made, and the number of participants who visited all pages/texts available according to which

condition they were in and which time allocation strategy they were classified as using.

Table 5.8: Number of pages and texts visited by condition and strategy used

Condition	Strategy		Number of text visits	Number of page visits	Number of unique texts visited	Number of unique pages visited	Number of pts who visited all 16 pages	Number of pts who visited 1 2 3 4 texts:			
INFORMED	Sampler	Mean	8.2	22.2	12.2	4	1	0	0	0	5
		S.D.	*1.79*	*5.54*	*3.7*	*0*					
	Satisficer	Mean	5	19.5	11.17	3.17	3	0	2	1	3
		S.D.	*3.35*	*11.02*	*5.38*	*0.98*					
	All pts (inc residuals)	Mean	6.33	19.83	3.5	11.42	4	0	2	2	8
		S.D.	*2.99*	*8.82*	*0.8*	*4.36*					
UN-INFORMED	Sampler	Mean									
		S.D.									
	Satisficer	Mean	6.5	25.89	14.11	3.56	6	1	0	1	7
		S.D.	*3.88*	*9.06*	*4.01*	*1.01*					
	All pts (inc residuals)	Mean	6.92	25.33	3.67	14.5	8	1	0	1	10
		S.D.	*3.55*	*8.45*	*0.89*	*3.5*					

There was no significant difference between conditions in the number of page visits ($t(22) = 1.56$), text visits ($t(22) = 0.43$), unique pages visited ($t(22) = 1.91$), unique texts visited ($t(22) = 0.48$), mean time spent per page visit ($t(22) = 1.68$) or the percentage of time readers spent in their favourite text ($t(22) = 1.71$).

Learning scores

Table 5.9 shows the mean learning score for both the informed and the un-informed conditions. There were no significant differences between the pre-scores ($t(22) = 0.86$), or the post-scores ($t(22) = 0.64$), or the number of points improvement between pre- and post-scores ($t(22) = 0.43$) between the informed and the un-informed conditions or between the learning scores ($t(22) = 0.61$) for each condition. However, there was a significant improvement from pre- (mean 0.38, *SD 0.95*) to post-score (mean 8.52, *SD 3.64*) across conditions ($t(23) = 11.37$, $p < .001$).

Table 5.9: Mean (*S.D.*) learning scores for both conditions (max score = 16)

Condition	pre test score	post test score	no. points improved from pre to post	Learning score
Informed	0.54 (*1.29*)	9.00 (*4.99*)	8.46 (*4.78*)	0.55 (*0.31*)
Uninformed	0.21 (*0.40*)	8.04 (*1.53*)	7.83 (*1.64*)	0.50 (*0.10*)

The fact that the post-test score is similar for the informed and the un-informed conditions is due to the fact that even participants in the un-informed condition read the best text, and therefore learned for the test. However, this is not a problem because what is of interest here is the time allocation behaviour, and what strategies the participants in the different conditions used in order to study for the test. Since one of the experimental questions was to see whether the samplers in the informed condition learned more than the satisficers and the one residual by sampling and finding the best text, an analysis compared the learning scores between these two groups of readers within the informed condition. The pre-scores and the post-scores between the samplers (mean pre-score 0.3, *S.D. 0.45,* mean post-score 10.7, *S.D. 3.*90) and the others (mean pre-score 0.7, *S.D. 1.*7, mean post-score 7.8, *S.D. 5.*6) were not significantly different (pre-scores between samplers and others, t(10) = -0.53, and post-scores between samplers and others, t(10) = 0.99). In terms of the learning score, no significant difference was found between the samplers (mean 0.67, *S.D. 0.64*) and the others (mean 0.48, *S.D. 0.25*) in the informed condition (t(10) = 1.00).

In the un-informed condition, there was no significant difference between the learning scores (t(10) = 2.00) or the points improvement (t(10) = 2.11) between readers who spent most of their time in the best text (learning score mean 0.44, *S.D. 0.07*, and points improved mean 6.8, *S.D. 1.25*) and readers who spent most of their time in the good enough texts (learning score mean 0.54, *S.D. 0.10*, and points improved mean 8.57, *S.D. 1.54*).

In the informed condition, there was no significant difference in the learning scores (t(10) = 2.02) or the points improvement (t(10) = 1.87) between readers who spent most of their time in the best text (learning score mean 0.69, *S.D. 0.25*, and points improvement mean 10.43, *S.D. 3.96*) and those who spent most of their time in the good enough texts (learning score mean 0.36, *S.D. 0.31*, and points improvement mean 5.70, *S.D. 4.79*).

There was no significant difference found in the learning scores between the informed and the un-informed conditions for readers who spent most of their time in the best text ($t(10) = 2.18$) or for readers who spent most of their time in the good enough texts ($t(10) = 1.43$). No comparison could be made for participants who spent most time in the bad text between conditions because only 1 participant in the informed condition spent most of their time in the bad text.

However, a correlation between learning score and time spent in best text revealed that participants in the informed condition *did* learn more the longer they spent reading the best text ($r = 0.66$, n = 12, $p < 0.05$): see Figure 5.3. However, there was no significant correlation between the time spent in the best text and the learning score for participants in the un-informed condition ($r = -0.09$, n = 12, *ns*): See Figure 5.4. A comparison of these two correlations revealed that the difference was approaching significance (using a Fisher Z-transform, $Z = 1.87$, $p = .06$).

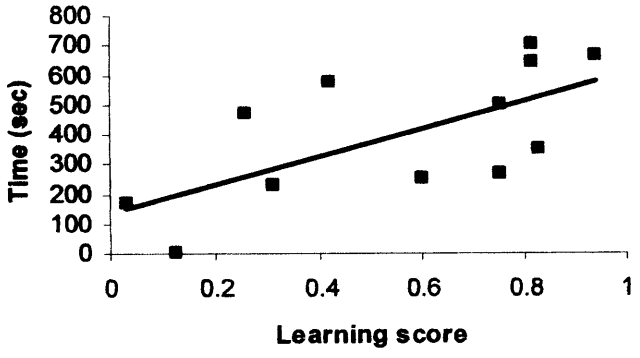


Figure 5.3 The relationship between times spent reading the 'best' text and learning score for the informed condition in Experiment 6.

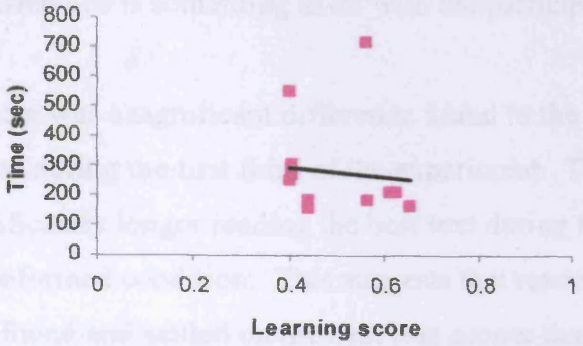


Figure 5.4 The relationship between times spent reading the 'best' text and learning score for the un-informed condition in Experiment 6 (no significant correlation).

5.4 Discussion

Experiment 6 investigated whether readers would change the way in which they allocated time across available texts if they were aware that there was one text out of the four available which was much better than the others in terms of relevance/usefulness to the test for which they were studying. It was hypothesised that, if readers were aware that a 'best' text existed, they would use a sampling strategy to find this optimal text, rather than satisficing, which runs the risk of the reader spending more time in good enough texts and not having the time to find the optimal text. The data suggests that only *some* of the participants who were aware of the 'best' text (those in the informed condition) used a sampling strategy, the rest displaying satisficing behaviour like those in the un-informed condition.

It is reasonable to suppose that those readers who spent most of their time in the best text would have higher learning scores than readers who spent most of their time reading the good enough texts, but this effect was not found. In fact, there were no differences in learning scores depending on the condition, the time spent in certain texts, or the time allocation strategy used. It was predicted that those participants who used a sampling strategy would find the best text more readily, and therefore have a higher learning score than those who did not. However, for those participants in the informed condition, the longer they spent reading the best text the better their learning score, which would be expected since the items which needed to be learned for the test were all present in the best text. This effect was not mirrored for participants in the un-informed condition. The precise reason for this is not known, but it is possible

that this difference is something to do with the participants' knowledge about the text qualities.

There was a significant difference found in the amount of time readers spent in the best text during the first third of the experiment. Those in the informed condition spent significantly longer reading the best text during this time frame than did readers in the un-informed condition. This suggests that readers in the informed condition may have found and settled on the best text sooner than readers in the un-informed condition. This fits with the fact that the participants in the informed condition were aware of the best text and therefore may have been allocating time to texts on the basis that this 'best' text was one of the four available texts. They may have done this by setting a higher threshold level. However, participants in the un-informed condition were not aware of the best text and therefore may only have been allocating time to texts which met their threshold level of information gain, which could conceivably be one of the good enough texts. This implies that the instructions given prior to the experiment had an effect on their time allocation strategy.

The fact that participants in the un-informed condition spent significantly longer viewing the first text they visited than those in the informed condition also suggests that participants may have been influenced by the instructions about text quality. Participants in the informed group may not have spent as much time on their first text visit because they were aware of the fact that one text out of the four was much better than the rest, and therefore were maybe not prepared to spend as long reading the first text they visited without first finding out if it was the best text or not. Those in the un-informed condition were not aware of the differing quality of the texts and therefore had no reason not to spend time reading the first text they arrived at if it was equal to or above their pre-set threshold of information gain.

Results showed that significantly more readers in the informed condition used a sampling strategy than in the un-informed condition. This suggests that some participants were aware that sampling would be the better strategy in order to find the best text for their learning needs under these circumstances. Advising readers of the existence of a best text did persuade some of them to use a sampling strategy (in the sense that it changed their time allocation strategy), whereas it was not considered in the un-informed condition, as there were no participants using a sampling strategy. However, still only five out of twelve of the participants in the informed condition displayed sampling behaviour, so although it is significantly more than the un-

informed condition, it is still less than half of the participants in that group. One reason for the low number of samplers in the informed condition could be the fact that participants do not feel confident in judging the quality of the entire text on the basis of a rapid visit. This idea has been discussed earlier to potentially explain the infrequency of sampling behaviour.

One of the original research questions was whether satisficing was popular due to the situation in which the readers found themselves, or due to the readers adopting this strategy despite the characteristics of the information environment. Results from this experiment suggest that some readers still satisficed even when they knew that this kind of time allocation may not necessarily be the most beneficial for them to find the information they needed to learn for the post-reading test. This therefore indicates that the high proportion of readers whose behaviour reflected the satisficing strategy was unlikely to be the result of the information environment in which the reader found themselves.

The prevalence of satisficing in Experiment 6, and in previous experiments, is surprising to the extent that as a time allocation strategy, satisficing almost seems counterintuitive. When presented with a number of options it makes more sense to sample what is on offer before choosing which one is most appropriate. It could be argued that the time constraints placed on participants may have had an effect on this, but Experiment 1 showed a propensity for satisficing even when participants were not under time pressure. Since there were little or no time costs involved with sampling in the set-up of these experiments, the question as to why participants do not sample more is paramount.

Sampling is relevant to the issue of time allocation across multiple texts because the information needs of a reader are likely to change after reading any relevant text. Sampling is only viable if text quality for the task can be reliably assessed relatively quickly considering the time available to the reader. The infrequency of sampling could therefore be a result of readers not being prepared to judge an entire text on the basis of a rapid visit to one section of it. Results from Experiments 4 and 5 suggest that even when rejecting texts, readers consult all or most of the pages of a text, and all or most of the paragraphs or sections on a page. Therefore, it could be argued that not all readers did sample when aware that there was a best text because they did not feel that they were able to reliably judge a text from a rapid visit to it. However, data from Experiment 6 showed that sampling *was*

possible with these materials under these experimental conditions, since nearly half of the participants in the informed condition employed a sampling strategy, even though the time they spent reading certain texts and their learning scores were not affected by their time allocation strategy. The fact that these readers sampled suggests that when they were aware of the situation concerning the quality of the texts available to them, that they were more likely to use a sampling strategy to find the best text than when they were unaware of the text variation. The fact that less than half of the total readers in the informed condition employed a sampling strategy suggests that although sampling is usable as a time allocation strategy, satisficing still presents as an attractive/adaptive strategy for the user to employ. There appears to be no universal preference for one strategy over the other in the informed condition.

In conclusion, the data from this experiment suggests that although manipulating the instructions given to participants (and therefore the expectations the participants had about the available texts before they began studying) did have some effect on the way in which they allocated their time across the texts, it can be assumed from the tendency of participants to adopt the satisficing strategy to allocate their time, that the most widely used study time allocation strategy, under these conditions, appears to be one of satisficing.

CHAPTER 6

6.1 General Discussion

6.1.1 Objectives

The aim of the research in this thesis was to explore the factors that affect how readers allocate time across multiple texts and to identify the strategies they use to adaptively allocate their time. The behaviour of participants when faced with the problem of too many texts to read in the time available was observed and then models of time allocation were applied to the data to explain the tendencies.

The questions laid out in Chapter 1 which the experiments in this thesis aimed to address were: a) whether readers were able to adaptively allocate their time across multiple texts in limited time (that is, whether they were able to allocate more time to the texts from which they would most benefit), b) what strategies they used to achieve this, and c) whether the strategies employed were dependent on the layout of text in the information interface.

This thesis has therefore investigated some of the strategies readers use when allocating time across multiple texts in order to learn for a test (writing an essay and learning for a test in Experiment 1, and just learning for a test in Experiments 2 to 6) when they have limited time in which to study and little or no prior knowledge about the likely quality or usefulness of the available texts. This chapter will briefly summarise the major findings of each experiment and how they contribute to the development of the satisficing model, discuss the limitations and further developments of this model, and then the overall conclusions that can be drawn from this series of experiments with relation to time allocation theories, and the wider implications of such findings in terms of applications to the real world.

6.1.2 Summary of Findings

Experiments 1 to 6 show certain patterns of behaviour which have been analysed in order to explain the strategies people use when allocating time across multiple texts in a limited time. This section will present some of the major findings from across the series of experiments.

Experiments 2, 3, 4, and 5 all show that more time was spent overall on the 'good' texts than on the 'bad' texts, which suggests that participants were able to judge text quality and actively allocate attention to texts judged better, according to the task requirements. This finding is important because it indicates that readers are adaptive in their allocation of time across written texts. The manner in which the texts were consulted was also relatively consistent across the experiments. On the whole, participants did not tend to revisit texts until all the other texts had been visited at least once first. The first visit to a text for most participants (around 75% in most experiments) was the longest visit they ever made to that text during their reading. Experiment 1 revealed that participants behaved in the same manner regardless of the time they had in which to study (15 minutes or 30-45 minutes) and the number of texts they had available to consult (4 or 8). Experiment 1 also showed that most time was spent reading the easy text rather the intermediate texts or most difficult text.

The eye tracking experiments (Experiments 4 and 5) showed detailed images of the way in which participants allocated their time when reading within a page of text. The data showed that all or most of the sections, or paragraphs, on a page were viewed by the reader before they made the decision to leave the page and move onto the next one. For those texts classed as 'bad', the visits made to the later sections/paragraphs on a page were very brief, but generally, participants still did not abandon the page before at least skimming these sections. For the texts classed as good, the duration of visits to later sections/paragraphs on a page were no different from visits to earlier sections/paragraphs on the page. This finding is also valuable because it helps to refine the model(s) which can account for this time allocation behaviour, and allow a more detailed explanation of the processes involved when readers are under time pressure to learn information from written on-line texts.

Experiment 6 examined participants' behaviour when the amount of information regarding the text quality was manipulated. When participants were told about the quality of the available texts (i.e., that one was much better than the rest for containing the information they needed for their test) they were more likely to briefly visit all available texts before settling on a chosen text to read than when they were not told about the quality of the texts before the reading phase. However, despite this behavioural change depending on the instructions given, still only half of the participants who were told behaved in this way. The other half still did not revisit texts and the first visit they made to any text was the longest visit they ever made to

that text. This is the behaviour observed in the participants who were not told about the text quality.

6.1.3 Theoretical Implications

This section will discuss each experiment in turn with respect to the theories being used to explain the pattern of data found. It will explain how each experiment adds to the complexity of the model which is needed to explain the pattern of results. It will summarise the interpretation of the data put forward and the explanations used to support these conclusions.

Experiment 1 examined how changing the number of texts available and the time readers had to study the texts affected the strategies readers used to allocate time across texts. Readers were not sensitive to these variations, as the browsing patterns classified according to one of the three alternatives set out (i.e., satisficing, sampling, or residual) suggested that their time allocation strategy during reading was, for the majority, in line with the satisficing model (as explained in Chapter 1). In Experiment 1, readers were "reading with a view to answering more questions *like* the ones they had already attempted" (these were the instructions given to participants), so they were not aware of the specific facts for which they would need to answer the questions in the post-reading test. The fact that neither text number nor study time affected the way participants allocated their time across texts, leads to the suggestion that time allocation strategy is not an artefact of text number or time pressure. It is hypothesised that the strategy being used in Experiment 1 is the satisficing strategy. It is clear that participants were not allocating time randomly across the texts, or reading linearly, because the time spent across texts differed significantly, and participants did show systematic behaviour throughout their reading. Similarly, the search model does not fit the data because participants were not aware of the facts which they needed to learn for the post-reading test, and therefore could not have been applying a search model for target information. Of course, students may be able to prioritise information in a non-fictional text and identify what information is important and what is not. However, since all of the texts were on the same topic and participants were just required to learn information, it is difficult to apply a model

based on a search mechanism to a process which requires the recognition of salient information from background information.

Readers visited most pages of a text, suggesting that the unit of text, or patch, being evaluated for the stay/leave decision was less than the entire document, although at this stage it was not yet clear exactly at what level they were making their stay/leave judgements. If the patch was the entire document, then readers would leave the text after only one visit to it, but since they visited most pages of a text, at this stage this suggests that the patch is a page (or less) as once they have finished reading one page or rejecting one page, they jumped to the next page (patch). Browsing was also exhaustive at the level of the text. Experiment 1 showed no effect of study time or number of texts on the learning score, and comparison between learning scores for participants displaying sampling behaviour versus those displaying behaviour classed as satisficing was not possible due to the small number of participants who sampled. The learning scores were also quite low across the series of experiments, and it is thought that this may be due to participants not being motivated enough to learn.

In order to more accurately test whether readers were adaptive in the sense that they allocated more time to texts from which they would learn the most for the post-reading test, Experiments 2 and 3 manipulated four texts so that they differed in direct relevance/usefulness to the test. Two of them were relevant to the post-reading test and the other two were not, but they were all of a similar difficulty rating. In these experiments participants were also exposed to the post-reading test prior to reading so that they knew exactly the information for which they needed to learn for the post-reading test. Although Experiments 2 and 3 changed the task into more of a search task than a learning task, and this may alter participants' behaviour to the extent that it shifts away from sampling, the fact that participants' behaviour did not differ from behaviour in Experiment 1 indicates that there were no obvious effects of changing the experimental task.

The fact that participants spent more time on the good texts in both Experiments 2 and 3 indicate that a simple model based on linear reading cannot account for the time allocation behaviour observed. Similarly, the systematic way in which texts were consulted and read also discounts a random search model. Therefore a more specific model is necessary to explain the pattern of data found in Experiments 2 and 3. It is possible that participants were following a search model,

whereby they rapidly scan the texts for target information (answers to the test questions) and stop to learn that information when it is encountered. However, there are issues with this model which make it implausible (e.g., the fact that it assumes people can recognise meaningful units of text without properly reading the text), which were discussed in Section 3.4. One finding which argues against the possibility of the search model being able to account for participants' behaviour is the fact that participants spent more time reading paragraphs/sections of text in the good texts which contained no 'target' information than paragraphs in the bad texts which contained no target information. If participants were using a search model to allocate their time across the texts, then the data would show a similar time being spent on paragraphs with no target information in the good texts *and* the bad texts. The fact that this is not found argues against the search model. Also, there is data in Experiments 2 and 3 that the search model fails to explain, for example, the fact that some participants left texts before they had looked at every page (this is also discussed in Section 3.4). It seems therefore, that the satisficing model is the model that fits best with the data observed in these experiments.

When looking at whether participants' behaviour was consistent with either of the two strategies by which they were characterised throughout the thesis, the browsing patterns of participants in Experiments 2 and 3 were more consistent with the satisficing strategy than with the sampling strategy. Also, readers were able to be adaptive in terms of to which texts they allocated most time (participants spent more time on the relevant than the irrelevant texts). These findings provide some support for the idea that readers are adaptive in their choice of which texts to read in the sense that readers were generally able to adaptively allocate more time to materials from which they would most benefit. However, adaptivity was not found to be related in any obvious way to the time allocation strategy used by the reader, so it cannot be claimed that this adaptivity is a product of using a particular strategy. Indeed, adaptivity was evident whichever strategy was used. Experiment 3 showed that adaptivity did, however, have an effect on the amount the reader learned during the study time. Those participants who spent more time in the good texts had higher learning scores than those participants who spent more of their time in the bad texts. This fits with what was predicted.

The data in Experiments 1 to 3 can therefore be explained using the satisficing model. Experiments 4 and 5 (eye tracking experiments) examined not only

the time allocation strategies of the readers (i.e., whether people were satisficing or sampling), but *how* they were satisficing – at what level of text were they making their judgements?

The data from Experiments 4 and 5 showed that the time participants spent reading the first paragraph of a bad text was significantly higher than the time spent reading the last paragraph, whereas for the good texts this difference was not so noticeable and there was no linear decrease in time spent on paragraphs as there was for the bad texts. However, as the gaze plots showed, 85% of readers in Experiment 5 still consulted most of the paragraphs on a page before leaving the page to read elsewhere, even if it was only by scanning the first line or two of later paragraphs. Similarly, in Experiment 4, which used blocked text, 70% of readers consulted all four quarters of the page, albeit briefly for the later quarters of the bad texts.

The search model discussed above, which involved a procedural rule of searching and pausing when encountering target information, does not account for the pattern of data regarding time spent on paragraphs found in Experiments 4 and 5. If participants were using a search model, then they would allocate time across sections of text according to how much target information each contained, and therefore would allocate just as much time on paragraphs at the top of the page in a bad text as paragraphs at the bottom of the page. Instead, another model is necessary to explain why participants spent less time on later paragraphs of bad texts but the same amount of time on later paragraphs of good texts. It is clear that judgements were being made about text patches on the basis of adjacent patches, and that participants were reading while employing some kind of judgement criterion. If they begin reading the page with a set criterion which is applied to the first paragraph, then their experience of this section of the page may influence their predictions about the remaining paragraphs, to the extent that if the first paragraph is bad then they may adjust the criterion level for later paragraphs. This can be explained using the satisficing model, since it states that a reader will leave a patch of text when it falls below threshold, thus causing participants to leave paragraphs of bad texts that are later in the page earlier than ones at the top of the page because they would have, according to the idea of an adjustable judgement criterion, altered their threshold level based on the first (bad) paragraph.

Unlike the data in Experiments 1 – 3, the data shown in the gaze plots in Experiments 4 and 5 cannot be fully explained by the satisficing model. The satisficing model cannot account for why participants spent less time on the later

paragraphs of the bad texts than they did on the earlier paragraphs; it could only account for their behaviour if they had either spent the same time on the first and last paragraphs, or if they had not visited the last paragraph at all. This is because according to the satisficing model, participants would leave a patch when information gain had dropped below a certain threshold, and jump to the next patch. Participants would therefore either read a patch or not read a patch. There are a few reasons that might explain why participants spend less time on later paragraphs of bad texts. The cost of skimming a patch is very low compared with switching pages or documents, and so participants may as well skim the remainder of the paragraphs on the page before leaving as well as satisficing. Participants could also be treating the page as a patch and conservatively checking the later paragraphs on a page before leaving, just in case there is information there that they may miss (although they have already decided to leave that page). Since the gaze plots show behaviour which does not quite fit with the simple satisficing model, it is necessary to develop the model further, perhaps one in which participants satisfice then skim in order to double check the rest of the paragraphs on the page before leaving.

Participants treated each text as separate, and the fact that they visited most pages of a text (although some participants did leave a text before having consulted every page), even a bad text, suggests that the unit of text which was being evaluated was less than or equal to a page. This refers to the size of text which readers treated as a patch. If participants left the text after visiting only one page, then this suggests that they would have been treating the whole document as the patch, whereas the fact that they visited nearly every page of a document suggests that the size of text which they were treating as a patch must have been a page (or less). This is based on the assumption that readers will leave a patch before they reach the end if it is not meeting their desired level of information gain, after which they will move systematically to the next patch. Thus, at this stage, a page is a patch because readers jump from one page to the next after having read or rejected it. If readers jumped straight to the next document after having read or rejected the first page of one document then the patch would be the entire document because readers would not have consulted any of the other pages of the first document. But here, readers jump to the next page, so it follows that the patch must be the page (or less).

The finding that readers who were classed as satisficers did not tend to revisit texts also indicates that they were judging texts as they read them (integrating

judgement and learning as suggested by the satisficing model) and not judging them and then choosing which ones to learn from (separating judgement from learning as suggested by a sampling strategy). The difference in the layout of the reading interface in Experiments 4 and 5 did not appear to change the time allocation strategies of the readers. These eye tracking studies therefore suggest that satisficing behaviour involves small, rather than large, units of text. Furthermore, given that the rapid visits evident in Experiments 1 to 3 are not likely to be long enough for participants to actually read the text (assuming an average reading speed of 250 words per minute; Bailey, 2000), coupled with the eye tracking data collected in Experiments 4 and 5, this implies that rejected texts were skimmed, rather than read fully or rejected on the basis of first page judgements.

Finally, Experiment 6 attempted to shift participants away from the satisficing strategy by creating a situation in which satisficing would *not* be the most beneficial strategy to adopt. To create this, texts were manipulated such that two were good enough, one was bad and one was the 'best' in terms of the number of test questions they answered. Half of the participants were told about the existence of the best text (informed condition) and the other half were not told anything about the quality of the texts (un-informed condition). There were more samplers in the informed than the un-informed condition, although within the informed condition the number of samplers and satisficers was more or less equal. Although the instruction that there was a best text did encourage more participants to sample, it was still not found to be the majority strategy in these conditions, as the browsing patterns still suggested that people were satisficing. Possible reasons for this were discussed in Section 5.4. One of these reasons relates to the idea that perhaps readers did not sample because they were unwilling to judge the value of a text on the basis of the first page, or they felt that they could not reliably judge the text quickly enough for sampling to be adaptive.

The time participants spent reading the best text during the first third of the study time in Experiment 6 also supports the claim that the instructions given to participants prior to the experiment had an effect on the time allocation strategy they employed. Those in the informed condition spent significantly more time reading the best text during the first third of the study time than did participants in the un-informed condition. In terms of the learning scores obtained in Experiment 6, there was no difference found between the scores of participants across the two experimental conditions. It had been predicted that a sampling strategy would enable

the reader to find the 'best' text more quickly and therefore allow them to spend more time on it, thus resulting in a higher learning score. This was not found to be the case. A possible reason for this could be that the readers who satisficed were still able to find the 'best' text to exploit. Experiment 6 therefore indicated that the high proportion of participants who appeared to be using a satisficing strategy did so not because the information environment encouraged such a strategy (because in Experiment 6 the information environment encouraged a sampling strategy for the informed condition), but because it was the more popular strategy of the two.

This finding is interesting because it suggests that satisficing appears to be the strategy readers use to adaptively allocate their time across texts under these conditions, and although it did not result in a better learning score in the experiments in this thesis, it does still indicate that using the satisficing strategy does not restrict the reader in terms of their information gain. This implies that satisficing is an adaptive strategy readers have developed to deal with the issue of having too much information to consult and not enough time in which to consult it all.

6.1.4 Limitations and further developments

Although the browsing pattern evident in the experiments in this thesis tend to suggest that satisficing is the time allocation strategy used by most people under the experimental conditions used, it must be noted that the satisficing model is quite flexible as an account of time allocation behaviour, and this aspect of the theory may invite criticism. As Reader and Payne (2007) pointed out, the satisficing model presumes an arbitrary and not necessarily constant threshold of satisfaction with the information that is being gained, or the amount the reader is learning during the reading/study phase. The level of this threshold will have effects on the browsing behaviour. If the threshold is very low, then this may lead to the reader studying a single text, whereas if the level was set very high then this may lead to rapid switching between texts which may appear to be behaviour so similar to sampling that it cannot be differentiated. A moderate satisficing threshold is assumed, from the prediction of satisficing that all first visits to texts would be the longest visits to those texts (Reader & Payne, 2007).

Another aspect of the satisificing strategy that is flexible is the unit of text which is being evaluated by the reader for the stay/leave decision. On the one hand, if the unit of evaluation is the entire document, then behaviour would show that readers leave texts after visiting the first page only, since they would judge the rest of the document on the basis of that visit to the first page, and if the document was the patch-unit then the reader would jump to the next document after one visit to one page of the first document (because the satisficing model assumes that once the reader has read or rejected a patch of text then they will move systematically onto the next patch). On the other hand, if the unit of text being evaluated is small (e.g., a paragraph), then a satisificing strategy would result in a kind of skimming behaviour, in which all or most of the pages of a document are visited, even if for a very short period of time. This type of skimming behaviour is evident in the experiments in this thesis. Also, the proposal that evaluations are made over small units of text is consistent with the lack of sampling behaviour.

An assumption of the satisficing model is that written text is divided into 'patches'. The skimmer's choice, in a similar fashion to a foraging animal, is when they should leave a patch, at which point, according to the model, they move to the next patch. This model is supported by the data in this thesis, and experiments such as the eye tracking experiments were carried out to try to refine this model by asking, what is a patch? If it is assumed that patches are visible, conventional units of text, then a document, a page, or a paragraph could be a patch. As discussed earlier, if the unit, or patch, of text were the document, then the data should show readers leaving documents before they reach the end as they would only visit one page and then jump to the next document. This is because the unit of assessment's best prediction is that all of that unit should be visited. If the unit is visited, then how soon it is left will depend on its individual quality. If the patch-unit is left after only a brief visit this implies that it did *not* contain the information the reader needed to learn for the test, and if the whole patch-unit was visited then this implies that either it *did* contain the information the reader needed to learn, or that the information gain was at a certain level that the reader thought it profitable to continue searching there, as they chose to read that unit rather than moving onto another.

Similarly for other patch-units, if the unit was the page, then the data should show readers leaving pages before they reach the end, and if the unit was the paragraph, the data should show readers leaving paragraphs before they reach the end

of those paragraphs. Therefore, if nearly all document visits include visits to *all* of the pages of the document, then it follows that the patch-unit must be a page, or less. In the same vein, if nearly all page visits include visits to *all* of the paragraphs/sections on the page, then it follows that the patch-unit must be a paragraph/section. If a paragraph is a patch, then once a paragraph has been read or rejected, the reader will jump to the next paragraph, as is supported by the eye tracking data. However, as discussed in Section 6.1.3, the gaze plots show that readers seem to be satisficing and then skimming later paragraphs or sections of the text as a kind of conservative check on the paragraphs/sections before leaving the page. This data suggests that the simple satisficing model needs to be extended to one which includes both satisficing *and* skimming as participants tend to read and reject unsuitable patches using both of these techniques. It is therefore suggested that readers are not only satisficing, but they are 'satisficing then skimming' when allocating their limited time across multiple texts. This satisficing then skimming model claims to account for the pattern of data observed (i.e., the skimming of small text units), which is why it is argued, in the absence of a more appropriate or accurate model, that this model can account for people's behaviour when allocating time across multiple texts under the conditions used in this thesis.

However, it is possible that the distinction being made between satisficing and sampling may not be as clear as was first thought. If participants are satisficing with small units of text (e.g. paragraphs), then it could also be suggested that participants were sampling each of the paragraphs and only reading them if they were judged to be equal to or above their threshold level of information gain. Although for the purposes of this research sampling has been defined as the participant visiting each text to judge their usefulness prior to studying them, in reality sampling may not necessarily require this to occur. Instead, sampling could be manifested as the reader visiting small units of text and only reading those which meet their needs. In this sense, sampling behaviour would not be too dissimilar to satisficing behaviour with small text units.

If the patch-unit of text being judged for the stay/leave page decision is indeed a page section, or paragraph (as suggested by the fact that the data in Experiments 4 and 5 show that participants consulted all paragraphs or sections of a page), then it would be expected that all of the units should be visited (as they generally are), and that the time it takes readers to leave each unit would depend on the individual quality

and richness of that unit. Therefore, it would be predicted on the basis of the simple satisficing model that the time participants spend on each of the paragraphs or page sections of a good text should be similar, and also for a bad text the time they spend on sections should be similar (because if the quality of the text does not change as one progresses down the page, then there is no reason for the reader to spend any more or less time on later sections/paragraphs). However, the data from Experiments 4 and 5 indicated otherwise. The time spent on quarter sections of a page of a bad text in Experiment 4 decreased as the reader went down the page such that the time they spent on the last section was less than the time spent on the first section. This pattern would suggest, according to the simple satisficing model, that the quality of sections later on the page were poorer than the quality of sections at the beginning of the page, which, given the construction of the texts, is known not to be the case. This also happened in Experiment 5, where the time spent on the last paragraph on a page of a bad text was less than the time spent on the first paragraph. The time spent on paragraphs and sections of a good text did not change according to their position on the page.

Despite the fact that this does not tally with what the satisficing model would predict based on the paragraph or text section being the patch, it does make sense that people will be sensitive to patches at different levels of hierarchy simultaneously, because such dependencies do inevitably exist in the real world. This means that readers may expect good paragraphs to be near other good paragraphs and in the same document as other good paragraphs (also for bad paragraphs). Thus, in their satisficing, readers may be treating paragraphs (or text sections) as the unit of text to be judged, but on the basis of the first paragraph of the page, they may be making a priori estimates of all subsequent paragraphs on that page. Then, when they begin the next page, the process starts again. This suggests that the satisficing model has two levels of patch-units; one which is treated as the primary patch (paragraph/text section) and another which is treated as a secondary patch which contains the smaller primary patches (the page). In theoretical terms this indicates that readers are working to a dynamic criterion within a page of text, and the criterion which they are using to make their judgements later on in the page is different to the one used to make judgements at the beginning of the page. The fact that when they begin a new page they revert back to their original criterion level also suggests that this dynamic criterion resets itself over pages, which makes sense assuming that people expect

interesting information to be at the top of the page. A further explanation could arise from there being an interval between the pages of the text which requires the reader to press a button to access the next page. Maybe the fact that a decision has to be made at the end of a page to jump either to the next page or to the next document causes the reader to reset their criterion level and treat the next page as a new piece of text. If the four pages of the text were accessed by a scrolling mechanism on the screen, then perhaps the criterion level would not be reset for subsequent pages, as the scrolling action would lack the interval which causes the reader to psychologically re-evaluate their expectations of the text.

These experiments build on the original model of satisficing laid out by Reader and Payne (2007) by further specifying the units of text readers are treating as patches. The eye tracking experiments have shown that readers use small rather than large units of text as patches, and that they use these patches as a way of moving through the text in such a way that most of their time is allocated to the texts from which they will most benefit. It must be pointed out that the eye tracking data is not entirely supportive of the simple satisficing model, as it shows that readers often make brief visits to last paragraphs/sections of bad texts when they should, according to the simple satisficing model, either spend the same amount of time as earlier paragraphs/secions or not time at all. However, some suggestions have been put forward to account for this behaviour, including the notion that readers are satisficing at the level of the page and skim checking the rest of the page before moving onto the next page, or satisficing at the level of the paragraph/section but making a priori estimates of the likely usefulness of later paragraphs/sections on the basis of the earlier ones.

6.1.5 Links with other theories

This section will discuss the results found in the experiments in this thesis in relation to previous research in adjacent areas. Specifically, it will examine the similarities and inconsistencies between the pattern of data found here with information foraging theory, and studies carried out in the area of study time allocation.

The research in this thesis combines methods from comprehension and text selection, computational language analysis (use of LSA in Experiment 1),

cognitive/learning theory (information foraging, time allocation), and eye tracking methodology in attempting to identify how readers allocate their time across multiple texts when learning for a pre-seen test. By combining these methods and theoretical approaches the outcome provides a more robust and multi-faceted account of the way in which readers approach the problem of adaptively allocating time across multiple texts.

Overall, the findings in this thesis fit with what information foraging theory would predict in the sense that readers will exploit the patches (texts) from which they will gain the most information by adapting their search strategies to maximise this opportunity. As Pirolli and Card (1999) point out, human behaviour is often not as rational as optimization models would suggest, and perfect information and infinite computational resources rarely exist. It is on this basis that they believe that a more realistic account of human behaviour involves bounded rationality or choice making based on satisficing. Pirolli and Card (1999) propose that the most optimal information forager will be one who is able to solve the problem of maximising the rate of information gained per unit cost taking into account the task's environmental constraints (such as costs of finding and accessing different information sources). Considering the time constraints present in the experiments in this thesis and the differing quality of the various information sources (texts) available to the participants, it can be concluded that these participants, according to the pattern of data found, 'foraged' by using a satisficing strategy, consistent with what Pirolli and Card propose is more accurate as an account of human foraging behaviour (i.e., not as rational as optimization).

The nature of information sources is such that they are often composed of a 'patchy' structure, and the information forager is required to move from one patch to another (i.e., in the case of these experiments from one paragraph, page, or document to another) in order to find their information. The need to decide upon how much time to spend on within-patch foraging tasks and on between-patch tasks is overcome, according to Pirolli and Card (1999), through the use of enrichment and scent following activities. These are moulding the environment to fit the strategies, and finding information by following cues, which represent the imperfect perception of costs of certain information source paths. What the findings in these experiments have shown is that readers have a tendency to search by rejection, so in a sense, search according to the lack of cues available. The pattern of

findings apparent in the experiments in this thesis suggest that readers find the most valuable information for their needs by rejecting sections of the information source (text) which are not providing much information gain (below a certain pre-set threshold), and moving systematically onto the next section. Furthermore, it has been shown in Experiments 4 and 5 that readers judge small units of text and once read or rejected, they automatically move onto the next small section (e.g., paragraph) of text to evaluate.

It is possible that this is what would be predicted by the information foraging model in the absence of scent cues, although this has not been shown or suggested previously. Information foraging theory proposes that information scent is used to make an assessment about the usefulness and importance of the information source, and that these assessments influence the decisions of the reader in what sources to exploit. In this sense, the scent model is dynamic, since as the state of the information forager changes as they go through their search, they must make decisions based on these imperfect proximal cues (these could be citations, abstracts, text snippets etc. that represent documents). If these cues are strong, then the forager will make the correct choices about which source to exploit for their needs, but if there is no scent, Pirolli and Card suggest that this will result in the forager performing random searches. The results in this thesis indicate that readers also make decisions about the usefulness and importance of information sources by judging the text by *rejecting* sections of an information source which lack scent cues indicating relevance or suitability. This is not to suggest that readers do not follow scent cues, but it illustrates the ways in which a forager may seek out desired information in the absence of these scent cues.

In terms of previous research into study time allocation, the findings from Experiment 1 that readers spent more time on the easy than the difficult texts are in agreement with the claim made by Son and Metcalfe (2000) that people will allocate more time to the easy (or judged easy) items/texts than to the difficult (or judged difficult) texts when they are learning for a test *under time pressure.* Most extant models of study time allocation claim that more study time is allocated to difficult items/texts, but these studies allow ample time for studying. Experiment 1 showed that more time was spent on the easiest text than the most difficult text and most participants favoured the easiest text than one of the intermediate texts, with which there was calculated to be most overlap between the participants' existing knowledge

and the level of the text themselves (i.e., the zone of learnability). According to theories based on the zone of learnability (e.g., Wolfe, 1998; Metcalfe, 2002) most time should have been spent on the intermediate level texts, due to these being the texts with the most overlap in terms of the participants' knowledge and the content of the texts. Another reason why more time should have been spent on the intermediate level texts, B and C, in Experiment 1 was the fact that these texts unintentionally answered more of the test questions than the other two texts. The behaviour of participants in Experiment 1 did not reflect this tendency, as most time was spent on the easiest text. However, there are other aspects of reading apart from mere information gain that may influence which texts the reader wishes to read (e.g., enjoyment, interest). Participants in Experiment 1 may have read the easy text because it was more enjoyable than reading the difficult text.

Despite this confound with Experiment 1 data, this conclusion about the fact that the participants did not spend most of their time in texts overlapping with their predicted zone of learnability still stands, as these were the texts with most answers anyway. However, as mentioned above, study time allocation may also be influenced by factors other than the perceived difficulty of the study items/texts. Factors such as what the participant is interested in or what they are motivated to learn might also have an effect on how much time they allocate to which items/texts, and this could account for the inconsistency between predictions made by theory and the pattern of data in Experiment 1.

Recent research and applications in Information Foraging Theory

In the past decade there has been much research carried out based on the idea of information foraging and information scent as drivers for how web users navigate around the pages of a website. This kind of research is ongoing, but this section will briefly mention some of the main developments that have been made in this area.

Recent applications of Information Foraging Theory include the development of a rational analysis of web use, which led to the construction of a cognitive model of web navigation called Scent-based Navigation and Information Foraging in the ACT architecture (SNIF-ACT); Pirolli & Fu, 2003; Pirolli, Fu, Chi, & Farahat, 2005. SNIF-ACT was developed to simulate users as they perform unfamiliar information-seeking tasks on the World Wide Web. The model is based on the concept of information scent, which characterises how users judge the utility of different courses

of actions they may make, based on local cues such as link images/text (Pirolli & Fu, 2003). This model then predicts that users who are doing unfamiliar tasks would choose links that have high information scent, and that users will leave a website when the information scent of that site drops below a certain threshold. The authors claim that this model is able to make useful predictions about complex interactions between users and the WWW. This would help in improving usability of websites, and also develop systems to improve user-WWW interaction which would help people find, make sense of, and use information to enhance solutions to everyday problems including health, careers, finance, etc. (Pirolli & Fu, 2003).

Other research using information foraging theory as a basis for usability issues includes work by Nielsen (2003), who claims that information foraging is a useful tool in analyzing online media because humans like to get maximum benefit for minimum effort. He also claims that humans, as informavores, need to feel that the scent they are following (information scent: cues related to their desired outcome) is getting stronger, otherwise they will leave and look elsewhere. The most obvious design implication of the suggestion that users follow scent when searching for information is to make sure that links and descriptions of categories are explicit enough to show users what is at that destination. If there are several navigation options, then it is easier for the user if they can see which trail leads to their desired destination and which trails lead elsewhere (Nielsen, 2003). Nielsen also suggests that feedback is necessary to keep users on the right path. Highlighting how patch-leaving models translate to website usage, Nielsen also discusses how sites should be designed to deal with users tendencies to leave patches when information gain falls. He suggests that the decision of when to leave a patch is somewhat dependent on how easy it is to reach the next patch. In order to keep users in a site, Neilsen recommends designers make their site worthy of the users attention by providing good information that is easy to find, and also make it easy for users to find more good information in the site (making sites 'sticky'). Information foraging predicts that the easier it is for users to find other good patches, the sooner they will leave a patch, and therefore Neilsen claims that search engines like Google encourage users to spend less time in any *one* site (because Google makes it is so easy to find another good site). He also claims that broadband connections encourage information snacking because people make shorter visits more often than with the old dial-up system. Following a patch-leaving model, these things do nothing but encourage shorter site visits.

When searching for information on the internet users are constantly re-evaluating the costs of accessing each information site and the profitability of each information site, so that they can maximise their intake of valuable information. Information foraging theory states that people will adapt their strategies and restructure their environment in order to maximise this information gain (Galletta, Henry, McCoy, & Polak, 2006). Because we use information scent to evaluate the path to information sources, it follows that sites with stronger scent (i.e., more explicit proximal cues) will give users more idea about distal content (where they are going). Users will prefer to visit sites which offer low-cost information access (i.e., sites that can be accessed with minimum effort and as quickly as possible) over sites which are high-cost (Galletta et al., 2006). Todd and Benbasat (1999) claim that users prefer low cost strategies that maximise our 'hit rate' on sites, suggesting that they would prefer to pursue sites which have clear information scent and provide clear information about distal content.

Other recent studies into how users make decisions about which pages to visit within a website has been carried out by Katz and Byrne (2003). They argue that web surfers searching for information within a website have the task of deciding whether to allocate time to the search function in the site (a search strategy) or to the menu (a browse strategy). They propose that information scent has implications for the user's cost-benefit analysis of that site. If a site has poor information scent, then this may encourage the user to use a search strategy rather than a browsing strategy, as it would be less costly in terms of time and effort. The authors claim that the user's decision of whether to search or browse is influenced by the design and structure of the website, specifically the information architecture in terms of labelling and menu structure, as well as the user's inclination.

6.1.6 Applications

The research in this thesis has suggested that skimming a document may not necessarily involve the reader deliberately searching for specific sections of the text, as Pirolli and Card (1999) have termed 'scent following', but rather that skimming could arise from readers rejecting patches (sections of text) and jumping to the next patch just because it is next. In this way, a 'skimmable' document would be one in

which the patches of a text are easily perceived by the reader and where the moves from one patch to the next can be readily made (i.e., a paragraph). Although there is no concrete evidence to suggest this from previous studies, the data from Experiments 4 and 5 in this thesis imply that readers do indeed 'skim' paragraphs of the text and reject those sections which are below their threshold level for information gain, because irrelevant sections of text are left earlier than relevant sections of text. Therefore, on the basis of this research, writers of on-line texts could be advised to put salient information at the top of the page, and then other information which they want people to read should be placed at the beginning of each subsequent paragraph, or section of page, since it is these areas of the page which are responsible for holding the readers' attention.

This design of on-line materials is already used, and journalists base the content of their writing on an inverted pyramid, placing the salient information at the top, i.e., the base of the pyramid (Ricketson, 2004). They work on the premise that the body of the text should merely amplify what the reader already knows is coming, as they write in such a way that they tell the reader what they want to know as quickly as possible. Also, the hotspot plots in Experiments 4 and 5 suggest that important information may be better placed at the left and at the top of the screen, as this is where most attention was directed. This is also illustrated by Google's Golden Triangle, the area shaped like a triangle at the top of the search results page which shows most eye-tracking activity. Eye Tools Eye Tracking research (2006) conducted a study whereby fifty participants were given five different scenarios that required the use of a search engine (Google was used each time). Results revealed that 100% of people looked at the listings in the top three positions, 85% of people looked at the fourth listing, 60% looked at the fifth listing, and only 50% looked at the sixth. Research also showed that 72% of users clicked on the first link. This research confirms the importance of the location of information on a page. Nielsen (2006) also showed that readers' attention on a web page appeared to be an F-Shaped pattern. However, the ability of a reader to skim a text may also depend on factors other than the layout of the patches of a text or how perceivable they may be. Factors such as the reader's reading level, their knowledge about the topic, the length of the text, and the cohesion of the text may also influence to what extent the reader is able to skim in their search for information.

Another potential design implication which arises from this research centres on the manipulation in Experiment 1 of the difficulty of the available texts, rather than the relevance of the texts. Experiment 1 showed that readers had a preference for the easiest text, therefore suggesting that they were sensitive to the level of difficulty of the texts as they read. The most obvious implication of this finding is the fact that people prefer to read easy texts, and so this may advise writers of on-line texts to keep their texts as simple as possible. The hyper-text structure on-line gives users an opportunity to see what the text/document is going to be like in terms of difficulty and style by providing the user with short summaries, headings, or text snippets as an indication of what to expect in the text/document. Such guides not only reveal the content of the text but also other qualities that are relevant to the utility of the text, such as difficulty. This research reinforces the need to allow readers to experience the style and vocabulary of the text as well as the topic before they choose to read, since the preference for the easy text in Experiment 1 suggests that readers would choose simple, easy to read texts over complex ones, especially when under time pressure. It has also been shown, by Reader and Payne (2007), that readers will consult aids such as outlines when they are available in order to sample the style and difficulty of the texts. The findings of Experiment 1 may therefore recommend to website and digital library designers that they should perhaps include summaries, headings, et cetera.

A practical application from this research may relate to pedagogical issues in selective reading. If it is the case that those readers who are adaptive in their allocation of time to texts (in the sense that they allocate more time to task relevant texts) learn more than readers who are not successful in their time allocation, then it may be beneficial to teach people how to read selectively. Most of the participants in the experiments in this thesis were able to allocate more time to the relevant/good texts, although there were some who allocated more time to the non relevant/bad texts. However, since the participants who took part in these experiments were university undergraduates, then this may account for their ability to choose appropriate texts, due to their developed study skills. It may be useful, however, to teach such selective reading skills earlier on in education to ensure that people are best equipped to deal with the increasing sources of information when they are allocating time across texts under time pressure to learn for a specific goal.

6.1.7 Future Research

Relating to the experiments in this thesis, it would be beneficial to investigate what happens to readers' time allocation behaviour when there are obvious scent cues present in the texts. The texts used in this thesis did not contain overt scent cues, and perhaps if these were added in the form of sub-headings, citations, etc. that represent the documents, then maybe the time allocation behaviour of the participants would present a somewhat different pattern. It is possible that if scent cues were available, then readers would use these to guide their reading rather than guiding their reading by rejecting sections of text in favour of the next. Exactly how the browsing behaviour may change is difficult to determine, but it is possible that the presence of scent cues may encourage the reader to use more of a sampling style strategy than a satisficing strategy.

There were no major learning effects found in the experiments in this thesis (except for Experiment 6 where the more time readers spent reading the 'best' text, the higher their learning score was) and it is possible that this is due to a lack of motivation for the participants to learn. In a real-life situation, a reader would have more motivation to learn the information for which they were searching, as this would be their reason for searching for it. However, in these experiments, participants may have lacked this motivation to learn, as their performance in the post-reading test had no bearing on anything and whether they did well or not was of no importance to them as they got paid or received course credit regardless of their test score. Therefore, a future experiment to increase participant motivation could involve giving participants money as a reward for their efforts, for example, a set amount of money for each question answered correctly. This may then show more of a learning effect, as participants would be motivated to learn (as there is a monetary reward). If this did show a learning effect, this learning could be explored with regards to which texts participants spent most time reading and the time allocation strategy they used to find these texts.

In order to compare this research with other research in study time allocation, it would be useful to explore whether the results of the experiments in this thesis would have been the same if the task had been different, for example if the task had been for participants to find and highlight facts rather than to find and remember facts for a test. It is possible that simple fact searching, and reading to understand, may

engender different time allocation strategies as the former requires the participant to simply remember and the later requires more cognitive effort for the participants to understand the information.

More generally, it may be useful to consider how foraging theory can be developed when other kinds of reading are involved, not just reading to learn for a test. When people are searching for information there are other issues which become important, other than mere information gain, for example, enjoyment or interest. Whether foraging theory accounts for reading when other parameters are involved is not clear from this research. A possible future experiment to explore how foraging theory could explain the strategies people use when they are carrying out different tasks, or when they are just reading for enjoyment may involve observing readers' behaviour in more natural settings when information gain is not the sole objective of reading.

Another idea for future research in the area of time allocation across multiple texts is to investigate whether people behave in a similar manner when allocating time across websites that contain graphics and illustrations, rather than just plain texts. The presence of graphical illustrations may have an impact on where the reader directs their attention on the page, and therefore may affect the way in which they allocate their time across the page. It would be practical to identify if the principles which can be applied to the design of online texts according to the results of the research in this thesis could also be applied to websites which contain more visual stimuli than plain text. If the tendency to judge text relevance on the basis of small units of text still applies when the text is fragmented due to figures or images, then this could potentially inform website designers as to which aspects of their site are used for relevance judgements. Eye tracking would be the most informative method to use to investigate such tendencies.

Another extension to the research in this thesis would be to observe browsing behaviour when there are many more texts than were used in Experiment 1 (four and eight texts), and to establish whether there is a limit to the number of texts with which readers are prepared to satisfice or sample. It seems sensible to assume that the more texts that are available the more likely it is that people will satisfice because if there are a large number of texts then sampling would become impossible. It would also be interesting to see whether people are systematic in their allocation of time across texts when there are more pages per text than used in this thesis. Research from Google

(Enquiro Search Solutions, 2005) shows that when doing a Google search only 50% of people moved onto the second page of the results, suggesting perhaps that extra pages would become redundant if there were too many. However, the research in this thesis shows that readers consulted all pages of a text, perhaps highlighting the differences between time allocation behaviour using an internet search engine and time allocation behaviour using on-line documents.

In terms of the time available for readers to allocate to texts, it may be informative to investigate whether people still use a satisficing strategy when they are searching for information for a specific purpose over a much longer period of time. All of the experiments in this thesis observed time allocation behaviour in a very limited amount of time with a well defined task goal. In the real world, it may occur that internet users faced with the task of allocating time across information sources may have several days to search for some required information, and have access to hundreds of information sources from which they can extract that information. Faced with plenty of time and resources, would people still use a satisficing strategy? This is outside the scope of the current research, which was interested in how people allocate time to limited sources when under time pressure, but it may be beneficial to examine whether the satisficing model would account for how people allocate time when time and resources are not so constrained.

The research in this thesis has made a start in investigating one of the primary problems which faces information foraging theory by identifying and examining the mechanisms behind the way in which people allocate their time and attention to the most useful of the information sources available to them. By extending this research to further explore the time, resource, and opportunity costs which exist for information foragers, a more thorough understanding of how people allocate time across texts will aid designers of information sources in making their site or item more attractive to the information forager. Knowing the strategies people use when allocating time across texts would also add to the already existing knowledge about how to make web-sites or information sources more productive and 'user-friendly', increasing revenue for the website developers and increasing satisfaction for the user.

References

Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology, 96*, 124-129.

Bailey, B. (2000). Human Interaction Speeds. *Human Factors International, User Interface Design newsletter, Aug 2000.*

Basmajian, J. V. (1982). *Primary Anatomy.* Baltimore: Williams & Wilkins.

Burnie, D. (1995). *Human Body, Concise Encyclopedia.* London: Dorling Kindersley..

Byrne, M., Anderson, J., Douglass, S., & Matessa, M. (1999). Eye-tracking the visual search of click-down menus. Paper presented at the CHI 99, Pittsburgh, PA, USA.

Byrne, M., John, B., Wehrle, N., & Crow, D. (1999). The tangled web we wove: a taxonomy of www use. Paper presented at the CHI 99, Pittsburgh, PA, USA.

Byron, M. (2004). *Satisficing and Maximising: Moral Theorists on Practical Reason*: Cambridge University Press.

Charnov, E.L. (1976). Optimal foraging: the marginal value theorem. *Theoretical Population Biology*, 9, 129-136.

Collier's Encyclopedia no. 11. (1992). NY: MacMillan Educational Co.

Davis, G. P., & Park, E. (1981). *The Heart: The living pump.* Washington, DC: News Books.

Dennett, D. (1991). *Consciousness Explained.* Boston: Little, Brown.

DiClemente, D., F., & Hantula, D., A. (2003). Optimal Foraging Online: Increasing Sensitivity to Delay. *Psychology and Marketing, 20(9),* 785-809.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker & J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theroy and practice* (pp. 249-276). Mahwah, NJ: Erlbaum.

Egan, D. E. (1988). Individual differences in human-computer interaction. In M. Helander (Ed.), *Handbook of Human-computer Interaction (pp. 541-568).* Amsterdam: North-Holland.

Enquiro Search Solutions (2005). Enquiro Eye Tracking Report I: Google, at

http://www.enquiro.com/research/eyetrackingreport.asp

EyeTools (2006). Eyetools, Enquiro, and Did-it uncover Search's Golden Triangle, at http://www.eyetools.com/inpage/research_google_eyetracking_heatmap.htm

Foltz, P. W. (1996). Latent Semantic Analysis for text-based research. *Behaviour Research Methods, Instruments and Computers, 28(2)*, 197-202.

Galletta, D., Henry, R., McCoy, S., Polak, P. (2006). The interactive effects of wensite delay, breadth, and familiarity. By Website Optimization, at http://www.websiteoptimization.com (date accessed 01/07).

Goldberg, J., Stimson, M., Lewenstein, M., Scott, N., & Wichansky, A. (2002). Eye Tracking in Web Search tasks: *Design Implications. Paper presented at the ETRA 02*, New Orleans, Louisiana, USA.

Goodrich, M. A., Stirling, W. C., & Frost, R. L. (1998). A Theory of Satisficing Decisions and Control. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 28(6)*

Gould, S. E. E. (1953). *Pathology of the heart.* Springfield, IL: Thomas.

Green, R. F. (1984). Stopping rules for optimal foragers. *The American Naturalist, 123(1),* 30-43.

Herbers, J. (1981). Time resources and laziness in animals. *Oecologia, 49,* 252-262.

Intelligent Essay Assessor. (1998-2005). Pearson Knowledge Technologies. Available: http://www.knowledge-technologies.com. (Date accessed 08/04)

Iwasa, Y., Higashi M., & Yamamura N. (1981). Prey distribution as a factor determining the choice of optimal foraging strategy. *American Nature,* 117, 710-723

Katz, M. A. & Byrne, M. D. (2003). The effect of scent and breadth on use of site-specific search on e-commerce websites. *ACM Trans. Human-Computer Interaction,* 10(3), 198-220.

Katzer, J. (1987). User Studies, information science, and communication. *Canadian Journal of Information Science, 12 (3/4),* 15-30

Krebs, J. R., Kacelnik, A., & Taylor, P. (1978). Test of optimal sampling by foraging great tits. *Nature, 275*, 27-31.

Laham, D. (last updated 1998). Latent Semantic Analysis at CU Boulder. Available: http://lsa.colorado.edu/. (Date accessed 09/04)

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25(2&3),* 259-284.

Marchionini, G. (1995). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science (1986-1998), 40(1),* 54.

Masson, M. E. (1982). Cognitive processes in skimming stories. *Journal of Experimental Psychology: Learning, Memory and Cognition., 8(5),* 400-417.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics, 17,* 578-586.

McNair, J. N. (1982). Optimal giving-up times and the marginal value theorem. *American Nature,* 119, 511-529.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131(3),* 349-363.

Miller, G. A. (1983). Informavores. In F. Machlup & U. Mansfield (Eds.), *The study of information: Interdisciplinary messages* (pp. 111-113), New York: Wiley.

Mohrman, D. E., & Heller, L. J. (2003). *Cardiovascular Physiology.* USA: McGraw-Hill.

Neilsen, J. (1997). Alert Box: Be Succint! Writing for the Web. Available: http://www.useit.com/alertbox/9703b.html. (Date accessed 12/04)

Neilsen, J. (2003). Alert Box: Information Foraging: Why Google Makes People Leave Your Site Faster. Available: http://www.useit.com/alertbox/20030630.html (Date accessed 01/05)

Neilsen, J. (2006). Alert Box: F-Shaped pattern for reading web content. Available: http://www.useit.com/alertbox/reading_pattern.html (Date accessed 06/06)

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: knowing about knowing (*pp. pp 1-25). Cambridge, MA: MIT Press.

O'Connor, B. (1993). Browsing: A framework for seeking functional information. *Knowledge: Creation, Diffusion, Utilization, 15(2),* 211-232.

Pirolli, P., & Card, S. (1995). Information Foraging in Information Access Environments. Paper presented at the *SIGCHI conference on Human factors in computing systems* Denver, Colorado, United States.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106(4),* 643-

675.

Pirolli, P., Fu, W., Chi, E. & Farahat, A. (2005.) Information scent and web navigation: Theory, models, and automated usability evaluation. *Human Computer Interaction International 2005; 2005 July 22-27*; Las Vegas; NV; USA

Pirolii, P. & Fu, W.(2003). SNIF-ACT: a model of information foraging on the World Wide Web. *9th International Conference on User Modeling; 2003 June 22-26*; Johnstown; PA.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124(3),* 372-422.

Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading.* New Jersey, USA: Prentice-Hall Inc.

Reader, W., & Payne, S. (2007). Allocating Time Across Multiple Texts. *Journal of Human Computer Interaction, 22(3).*

Rehder, B., Schreiner, M. E., Wolfe, M. B., Laham, D., Landauer, T., & Kintsch, W. (1998). Using Latent Semantic Analysis to assess knowledge: some technical considerations. *Discourse Processes, 25(2&3),* 337-354.

Rice, R. E., McCreadie, M. M., & Chang, S.-J. (2001). *Accessing and Browsing Information and Communication. Cambridge,* MA: MIT Press.

Ricketson, M. (2004). *Writing Feature Stories: How to Research and Write Newspaper and Magazine Articles,* Sydney: Allen & Unwin.

Roberts, N. (1982). A search for information man. *Social Science Information Studies, 2,* 93-104.

Russell, M. (2005). Using eye-tracking data to understand first impressions of a website. *Usability News, published by the Software Usability Research Lab (SURL) at Wichita University,* at http://www.usabilitynews.org. (vol 7.1) (date accessed 01/07).

Sandstrom, P. E. (1994). An optimal foraging approach to information seeking and use. *Library Quarterly, 64(4),* 414-449.

Schroeder, W. (1998). Testing web sites with eye-tracking. available at http://www.uie.com/articles/eye_tracking. (Date accessed 04/05)

Silverstein, A., & Silverstein, V. B. (1983). *Heartbeats; Your body, your heart.* New York: Lippincott.

Simon, H. A. (1955). A behavioural model of rational choice. *Quarterly Journal of Economics, 69(1)*, 99 - 118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138.

Simon, H. A. (1976). *Administrative behaviour: A study of decision-making processes in administrative organization.* 3rd Edition. New York: Free Press

Simon, H. A. (1990). Invariants of human behaviour. *Annual Review of Psychology, 41*, 1 - 19.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory and Cognition., 26(1),* 204-221.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging Theory.* Princeton, NJ: Princeton University Press.

Texas Heart Institute. Available: www.tmc.edu/thi/anatomy.html. (Date accessed 02/04)

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory and Cognition., 25(4),* 1024-1037.

Todd, P. & I. Benbasat., (1999), Evaluating the impact of DSS, cognitive effort, and incentives on strategy selection, *Inform. Systems Res.* 10(4) 356-374.

Vygotsky, L. S. (1987). *The collected works of L. S. Vygotsky: Vol 1 Problems of general psychology.* New York: Plenum Press.

Ward, D. & Blaustein, J. (1992). The role of satisficing in foraging theory. *Oikos, 63(2),* 312-317

Williams, G. C. (1992). *Natural Selection: Domains, levels, and challenges.* New York: Oxford University Press.

Winter, S. G. (2000). The satisficing principle in capability learning. *Strategic Management Journal, 21(10-11),* 981-996.

Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: matching readers and texts by Latent Semantic Analysis. *Discourse Processes, 25(2&3),* 309-3

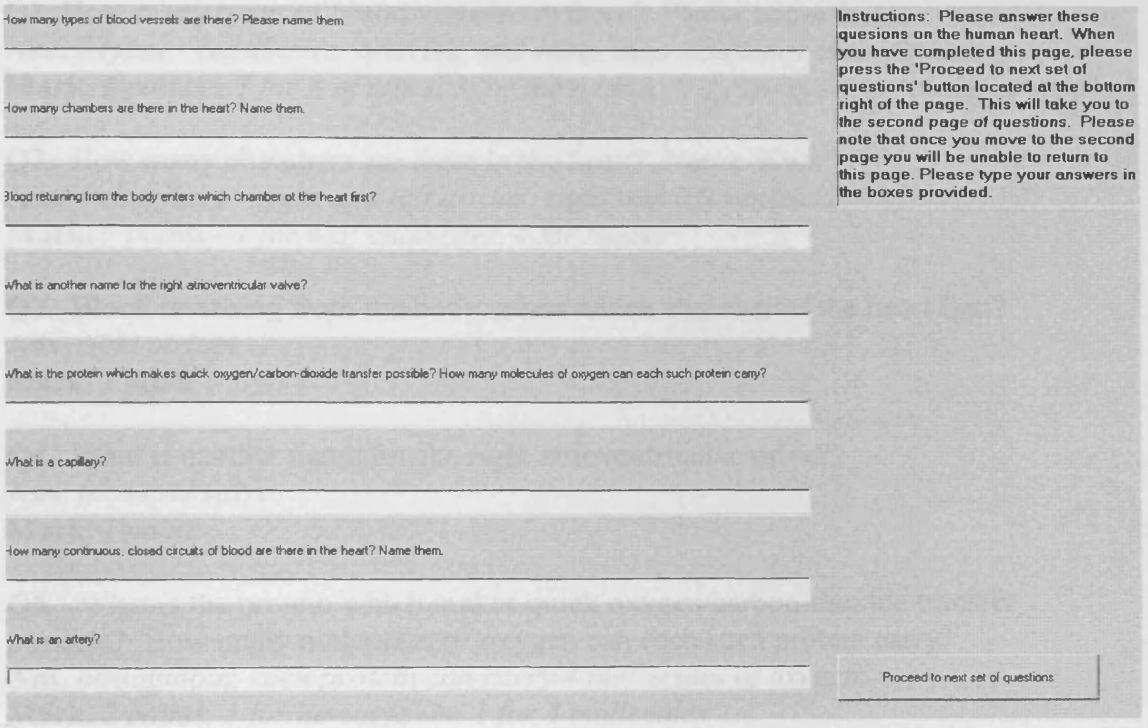
APPENDIX A

Latent Semantic Analysis

LSA is a way of extracting and representing the contextual-usage of words through the use of statistical computations which are applied to a large corpus of text (Landauer, Foltz, & Laham, 1998). LSA infers relations of contextual usage of words in given text passages by taking 'raw text parsed into words defined as unique character strings and separated into meaningful passages or samples such as sentences or paragraphs' (Landauer et al., 1998, p 263). LSA allows us to approximate human judgements of the similarity of meaning between words due to the fact that the measures which are produced by LSA of word-word, word-passage, and passage-passage relations are so well correlated with a variety of human cognitive phenomena comprising semantic similarity. The similarities predicted by LSA infer much deeper relations than simple co-occurrences, or correlations in word usage (hence the term 'latent semantic') and therefore are more accurate predictors of human judgements and performance than surface level contingencies (Landauer et al., 1998). Since LSA allows us to compare the semantic similarity between pieces of textual information, it is of great value in text based research. One of the main advantages of using semantics as a comparison between texts is that such a comparison would not depend on surface features, such as the choice of words used (Foltz, 1996). LSA is built on the assumption that there is some latent structure in the pattern of word usage across different texts, and that this latent structure can be estimated using statistical techniques (Foltz, 1996). LSA analyses the associations between words and 'produces a representation in which words that are used in similar contexts will be more semantically associated' (Foltz, 1996, p 3).

APPENDIX B

Layout of on-screen test (page one shown here).



APPENDIX C
Heart Test 1

Q1. How many types of blood vessels are there? Please name them.
Ans. *5 types - capillaries, arteries, and veins. also arterioles and veinules.*
Mark. *6 points - 1 for # of types, 3 for the names (6 if also get arterioles and venules)*

Q2. How many chambers are there in the heart? Name them.
Ans. *4 chambers - right and left atrium, right and left ventricle*
Mark. *5 points - 1 for # of chambers, 4 for names*

Q3. Blood returning from the body enters which chamber of the heart first?
Ans. *right atrium*
Mark. *1 point*

Q4. What is another name for the right atrioventricular valve?
Ans. *tricuspid valve*
Mark. *1 point*

Q5. What is the protein which makes quick oxygen/carbon-dioxide transfer possible? How many molecules of oxygen can each such protein carry?
Ans. *hemoglobin. each protein can carry 4 mollecules of oxygen.*
Mark. *2 points, 1 for hemoglobin, 1 for 4 mollecules*

Q6. What is a capillary?
Ans. *type of blood vessel that is involved in oxygen / carbon dioxide exchange*
Mark. *2 points - 1 for blood vessel, 1 for gas exchange. no extra point for smallest.*

Q7. How many continuous, closed circuits of blood are there from the heart? Name them.
Ans. *2 - pulmonary and systemic (or lung / body or something with that idea)*
Mark. *3 points - 1 for 2 closed circuits, 2 for the names*

Q8. What is an artery?
Ans. *blood vessel that carries blood away from the heart (or that carries oxygenatd blood, even though that's not techinically correct.*
Mark. *2 points - 1 for blood vessel, one for function*

Q9. What is an atrium?
Ans. *It is a chamber of the heart that is located at the top of the heart. there are two of them and they serve as the receiving chambers for blood.*
Mark. *2 points - 1 for heart chamber, 1 for location or receiving chamber*

Q10. Where does the blood entering the left atrium come from?
Ans. *pulmonary vein (or lungs)*
Mark. *1 point*

Q11. What is a ventricle?
Ans. *chamber of the heart that is located at the bottom of the heart. there are two of them and they serve as the dispensing or exiting chambers.*
Mark. *2 points - 1 for heart chamber, 1 for location or exit chamber.*

Q12. What is another name for the left atrioventricular valve?
Ans. *bicuspid or mitral valve*
Mark. *1 point*

Q13. Where does blood entering the left ventricle come from? The right ventricle?
Ans. *blood entering the left ventricle comes from left atrium. blood from right ventricle comes from right atrium.*
Mark. *2 points - 1 for left, 1 for right*

Q14. The pacemaker is the common term for what specific part of the heart? Where is it located?
Ans. *sinoatrial (or SA) node. it is located in the wall of the right atrium*
Mark. *2 points - 1 for name, 1 for location*

Q15. Which side of the heart is larger? Why?
Ans. *left side is larger because it has to pump blood to the entire body, as opposed to just the heart.*
Mark. *2 points - 1 for left side, 1 for pumping to body*

Q16. What is unusual about the pulmonary veins?
Ans. *pulmonary veins carry oxygenated blood*
Mark. *1 point*

Q17. What are the names of the main veins which carry blood back to the heart from the body? How many such veins are there? From what part of the body does each such vein return blood?
Ans. *superior and inferior vena cava. superior comes from the upper part of the body (head/ arms, etc), inferior from the lower body.*
Mark. *5 points - 2 for names, 1 for "two of them", 2 for where they return from*

APPENDIX D

Proportion of time spent on texts A, B, C, and D in the four text condition of Experiment 1.

Text	Condition		
	Timed	Un-timed	Overall mean (sd)
A	39.79 *(18.18)*	26.09 *(13.86)*	32.94 *(17.35)*
B	23.35 *(15.83)*	27.21 *(8.38)*	25.28 *(12.60)*
C	25.55 *(15.51)*	29.65 *(9.79)*	27.60 *(12.91)*
D	11.33 *(9.34)*	17.11 *(8.75)*	14.22 *(9.37)*

APPENDIX E

Time spent in texts A, B, C, and D for timed and un-timed according to the low/high pre-test score in the four text condition (Experiment 1).

Text	Time	Pre-test	Mean	*S.D.*	N
A (easy)	**Timed**	low	370.13	*120.85*	10
		high	319.35	*235.69*	5
	Un-timed	low	665.93	*283.05*	6
		high	351.39	*133.84*	9
B (intermediate)	**Timed**	low	180.7	*142.2*	10
		high	262	*137.2*	5
	Un-timed	low	424.7	*134.4*	6
		high	542.4	*139.2*	9
C (intermediate)	**Timed**	low	245.5	*152.8*	10
		high	190.3	*111*	5
	Un-timed	low	494.5	*130.9*	6
		high	575.7	*209.4*	9
D (hard)	**Timed**	low	91.01	*86.25*	10
		high	119.31	*80.99*	5
	Un-timed	low	276.37	*178.18*	6
		high	342.23	*160.61*	9

APPENDIX F

Time spent in texts A, B, C, and D for timed and un-timed according to the low/high pre-test score in the eight text condition (Experiment 1)

Text	Time	Pre-test	Mean	*S.D.*	N
A8	**Timed**	low	244.96	*173.17*	7
		high	68.71	*64.64*	8
	Un-timed	low	205.36	*212.17*	9
		high	227.48	*180.6*	6
B8	**Timed**	low	100.78	*94.98*	7
		high	150.09	*127.63*	8
	Un-timed	low	156.07	*167.81*	9
		high	221.87	*210.18*	6
C8	**Timed**	low	31.12	*30.59*	7
		high	180.64	*127.91*	8
	Un-timed	low	247.59	*352.3*	9
		high	388.15	*350.6*	6
D8	**Timed**	low	41.23	*42.88*	7
		high	105.78	*127.6*	8
	Un-timed	low	158.5	*180.39*	9
		high	52.34	*74.42*	6

APPENDIX G

Heart Test 2

Q1. What type of tissue are the heart's valves made of?
Ans. *Thin cusps of fibrous tissue anchored to the heart's skeleton*
Mark. *1 point*

Q2. What is the purpose of arterial pressure being kept higher than venous pressure?
Ans. *Arterial pressure is kept higher than venous pressure by the pumping action of the heart which allows blood to flow through all the organs*
Mark. 2 points: 1 point for pumping action of heart, and 1 point for allowing blood to flow to organs

Q3. Through which valves does the blood exit the ventricles?
Ans. *through the semilunar valves*
Mark. *1 point*

Q4. How many layers does the myocardial muscle of each ventricle have?
Ans. *three*
Mark. *1 point*

Q5. From the ventricles, blood is channelled into two main arteries; what are they called?
Ans. *the pulmonary artery and the aorta*
Mark. *2 points: 1 point for pulmonary artery, 1 point for aorta*

Q6. How is the apex of the heart moved towards the mitral valve?
Ans. *left ventricular contraction causes a decrease in diameter and a shortening of the axis between the base and the apex of the heart.*
Mark. *2 points: 1 for decrease diameter, 1 for shorten axis*

Q7. How much blood does the right atrium hold and how thick is it's wall? How much blood does the right ventricle hold and how thick is it's wall?
Ans. *a) it can hold 3.5 tablespoons of blood, and it's wall is less than an eighth of an inch thick*
b) it can hold more than a quarter of a cup of blood, and it's wall is a quarter of an inch thick
Mark. *4 points: 1 for how much blood, 1 for thickness of wall (for both a and b)*

Q8. What happens during the phase of the cardiac cycle called systole? And during diastole?
Ans. *a) the ventricular muscle cells are contracting*
b) the ventricular muscle cells relax, pressure in the ventricle falls below that of the atrium, the AV valve opens and the ventricle fills with blood
Mark. *2 points: 1 for a, and 1 for b.*

Q9. How many pulmonary veins empty into the main cavity of the left atrium?
Ans. *Four*
Mark. *1 point*

Q10. Cardiac function can be influenced by neural inputs from which divisions of the autonomic nervous systems?
Ans. *the sympathetic and the parasympathetic divisions*
Mark. *2 points: 1 for sympathetic and 1 for parasympathetic*

Q11. Are the papillary muscles stronger in the right or the left ventricle?
Ans. *the left*
Mark. *1 point*

Q12. Capillaries wrap themselves around pockets of lung tissue called what?
Ans. *alveoli*
Mark. *1 point*

Q13. How many of these air sacs does the lung contain and what is the surface area?
Ans. *750 million sacs and surface area of more than 750 sq feet*
Mark. *2 points: 1 for number and 1 for area*

Q14. What releases norepinephrine? What releases acetylcholine?
Ans. *the sympathetic nerves and the parasympathetic nerves*
Mark. *2 points: 1 for sympathetic and 1 for parasympathetic*

Q15. How big is the inside diameter of the aorta?
Ans. *25 mm*
Mark. *1 point*

Q16. What do all the vessels have in common?
Ans. *They are all lined with a contiguous single layer of endothelial cells*
Mark. *1 point*

Q17. What colour is the blood when it travels back through the heart to the lungs?
Ans. *a dull purple colour*
Mark. *1 point*

APPENDIX H

Milky Way Test

Bonus questions in **bold.**
Each question is worth 1 point.

Q1. What does 'Via Galactica' mean?
Ans: *the milky way*

Q2. Who was the first person to claim that the Milky Way consisted of distant stars?
Ans: *Democritus*

Q3. Who discovered that the Milky Way was integrated light of countless stars which are too faint to be seen individually?
Ans: *Galileo Galilei*

Q4. Approximately how many stars is it thought the Galaxy consists of?
Ans: *approx 100 billion stars*

Q5. How long ago was the Sun and planetary system formed?
Ans: *4560 million years ago*

Q6. What blocks our view of most of the Galaxy?
Ans: interstellar dust

Q7. What did astronomers begin to suspect about the Milky Way in the 1980s?
Ans: that the milky way is a barred spiral rather than an ordinary *spiral*

Q8. What is the distance between the Sun and the galactic centre?
Ans: approx 27,700 light years

Q9. What did people in Eastern Asia believe about the hazy band of stars in the Milky Way?
Ans: people believed that the hazy bank of stars was the 'silvery rover' of heaven

Q10. What are spiral nebulae?
Ans: *enormous and distant collections of stars, similar to our galaxy*

Q11. What can astronomers learn about stars from the Milky Way?
AnsS: *can learn about the birth, life and death of stars*

Q12. What is the active site for formation of stars called?
Ans: *nebulae (regions of dense interstellar gas)*

Q13. What is the name of the smallest star ever observed?
Ans: brown dwarf orbiting a cool red star Gliese 229 (called Gliese 229B)

Q14. What is Eta Carinae?
Ans: a supermassive star of gas and dust clouds. One of the brightest stars I the southern sky.

Q15. What is Akashaganga?
Ans: is the Indian name for the milky way galaxy, which means Ganges River of the Sky

Q16. How old is the universe currently considered to be?
Ans: about 13.6 billion years