An investigation of unconscious precognition in the visual attention system

David W Smith

Doctor of Philosophy
Department of Psychology
University of Edinburgh
2012

ABSTRACT

Precognition can be defined as an anomalous correlation between current cognitive activity and a future event. Using behavioural and physiological measures, a number of previous studies have reported evidence for unconscious precognition during a variety of task conditions. The current thesis presents five experiments that were designed to test for unconscious precognition in the visual attention system while participants were engaged in a short term visual memory task. Each trial consisted of a *study* and *test* phase. In the study phase, participants were required to memorise an array of four stimuli while their eye movements were recorded. After a brief retention interval, a probe stimulus was presented for a yes/no recognition test. Two conditions were employed and were randomly determined. In the *old* condition, the probe was a stimulus viewed during study, termed the *target*. In the *new* condition, the probe was a novel stimulus. Experiments tested for the presence of precognition by examining whether there was a difference in the degree to which visual attention was allocated to items during the study phase of old and new trials. Two further studies were also carried out involving simulations that aimed to establish the extent to which a previously described artefact, termed the expectation bias, may impact on the results.

Experiment 1 suggested that participants spent more time attending to target stimuli in old compared to new trials, a result that appeared to provide evidence for precognition. However, the data was considered unreliable due to inadequate randomisation. An exact replication of Experiment 1 was carried out in Experiment 2 with adequate randomisation, but failed to find evidence for precognition. Experiment 3A was a further attempt to replicate the preliminary results of Experiment 1 using more extensive randomisation procedures while Experiment 3B explored the potential role of the probe stimulus in generating a precognitive effect. However, no support for the precognitive hypothesis was found in either experiment. A fully balanced design was employed in Experiment 4 in order to control for potential confounds such as position and saliency effects. The results supported the precognitive hypothesis and suggested that less attention was allocated to targets in the old condition. An exploratory analysis also examined the relationship between several standardised stimulus variables and the apparent precognitive effect observed in Experiment 4. The results revealed a suggestive relationship between the size of the effect and item ratings of familiarity and visual complexity. Simulations of an expectation bias in Experiments 5A and 5B together with post-hoc examination of the data from the current series of experiments suggest that this artefact is not a plausible explanation for the observed effects. The thesis ends with a discussion of several methodological issues that may impact on both the interpretation of positive results and the conclusions that may be reached from this body of data as a whole. Finally, suggestions for further work are made.

Acknowledgements

I would like to thank Caroline Watt for her guidance and patience as well as members of the KPU for their valuable feedback during my time as a PhD student.

I would also like to acknowledge the financial support of the Denwyn Dobby Scholarship fund for making this research possible.

A special thanks to my parents and my wife, Cindy Coleman, for giving me encouragement and supporting my decision to take the road less travelled.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(David Smith)

Table of Contents

Chapter 1 - General Introduction

Extra-sensory Experiences

Broadly speaking, *extra-sensory experiences* appear to convey information that could not ordinarily have been obtained. It is perhaps easiest to illustrate this central characteristic by way of an example. The following was taken from the spontaneous case collection of Louisa E. Rhine (1961):

> One summer several industrialists went for a fishing trip into the wilds of Canada. Among them was the district manager of a sheet and tin plate company. For about two weeks they had been in the deep woods, cut off from all news sources.
>
> The night before they were to return home, the district manager had a dream, so clear, so vivid, he could not sleep afterward. In it, he writes, "one of our locomotive cranes that was unloading a car of scrap iron, together with the car, was on the track near the back of a river alongside the water tower which served the locomotives. For some unaccountable reason, as the huge magnet swung around with a heavy load of scrap, it suddenly toppled over the river bank. The operator, whom I called by name, jumped clear of the crane and landed below it as it came bounding, tumbling and bouncing down the river bank, and he finally disappeared from view as the crane came to rest twenty feet below at the water's edge. I particularly noted the number of the crane and the number and positions of the railroad cars, and was able to tell how the crane operator was dressed. Furthermore, I noticed the approximate damage done to the crane. I did not know, however, what had finally happened to the operator. He had disappeared under or behind the crane after it had come to rest. In other words, I was observing the accident from somewhere in or across the river.
>
> "Upon my return to the mill the following day, the first man I met was the master mechanic. He told me to come with him to inspect the crane of my dream, to talk with the operator who had emerged from the accident without a scratch. The operator explained his lack of injury by the fact that the crane had fallen over in front of him as he made his last jump and as it made its last bounce. The record showed the smallest detail to be as I had dreamed it, with one exception. The exception was that the accident had happened two hours after the dream". (pp. 43-44)

The experience just described might be categorised as 'extra-sensory' because it appears to convey information about a future event that could not ordinarily have been obtained. Here, the dream is assumed to have foretold an apparently unforeseeable event. Of course, experiences that take place outside of controlled laboratory conditions are prone to error and the extra-sensory *appearance* of such reports may reflect processes that can be described by current scientific knowledge. Accordingly, one approach to the scientific study of such experiences, termed *anomalistic psychology* (Zusne & Jones, 1982), proposes that the action of various cognitive and perceptual biases can give rise to an experience that only appears to violate our common sense notions of causality. This approach has produced a body of evidence suggesting that biases in memory, reasoning, probability judgement and perception can contribute to the formation of a belief that a paranormal event has taken place when, in fact, it has not (Blagrove, French & Jones, 2006; French, 2003; French & Wilson, 2006; Gilovich, Vallone, & Tversky, 1985; Pronin, Lin, & Ross, 2002). An alternative approach is to ask whether the paranormal appearance of some extra-sensory experiences reflects an underlying process that is, in fact, paranormal and therefore not yet fully described by science. This central question is addressed by *parapsychology*, a field that attempts to test for *psi* - the putative anomalous process underlying such reports. It should be noted that the hypotheses tested by anomalistic psychology and parapsychology are not mutually exclusive; it is possible that a proportion of reported extra-sensory experiences are due to perceptual and cognitive biases while a remainder are genuinely anomalous. Scientific study of these experiences, and their underlying causes, can therefore proceed in a way that involves both fields working in parallel.

Extra-sensory experiences have been classified in various ways. One method has been to group them according to phenomenological criteria. For example, Rhine (1961) grouped experiences into four major categories according to the ‘form’ of the reported experience, which she labelled as *intuitive impressions*, *hallucinations*, *realistic dreams* and *unrealistic dreams*. Intuitive impressions were characterised as feelings of familiarity, a sudden ‘just knowing’, conspicuous emotions or compulsive behaviour, all of which seemed to provide information about some external event by apparently extra-sensory means. Typically, intuitive impressions were not experienced with any accompanying sensory imagery that would have provided more contextual detail. On the other hand, experiences placed in the category of hallucinations mostly conveyed information in the form of sensory perceptions. Here, simple hallucinatory sensations were reported in a variety of modalities but more complex hallucinations were sometimes reported involving so-called ‘apparitions’ of relatives recently deceased or in physical crisis some distance away from the percipient at the time of the experience. Realistic dreams were characterised by detailed and seemingly accurate information about apparently inaccessible events while unrealistic dreams were interpreted as symbolic representations of such events. The dream report provided at the start of this chapter is an example of an extra-sensory experience categorised as a realistic dream. Naturally, unrealistic dreams would be relatively more prone to misinterpretation, confirmation bias and other cognitive biases that artificially create correspondences when none are present, and Rhine (1961) readily acknowledged this problem. However, the aim of Rhine's classification scheme was not one of authenticating the paranormal nature of such reports. Rather, it was an attempt to derive testable hypotheses about the processes that may underlie such reports based on trends that may have emerged from analysis of their content and character.

A more popular approach to the classification of extra-sensory experiences has been to group them into subtypes based on the circumstances during which the information appears to have been obtained. Thalbourne (2003, 2004) summarises this classification scheme, defining each subtype according to an ostensible correlation between the information conveyed by the experience and a *target* event that appears to be inaccessible to the percipient. The experience is labelled as *telepathic* when the target event is assumed to be mental in nature. For example, the following case was categorised as a telepathic experience in Radin (1997a):

> In the middle of the night, out of a deep sleep, Fred suddenly jerked upright into a sitting position. He clutched his chest, gasping for breath. His wife, abruptly awakened by her husband's sudden movement, anxiously asked, "What's wrong?". A few moments later, when Fred was able to breathe normally again, he told his wife he was alright, but he had a feeling that something terrible had happened. They glanced at the clock: 2:05 a.m.
>
> Fifteen minutes later, as they settled back to sleep, the phone rang. Fred's father was on the line. "I have bad news", he said. "Your mother just had a heart attack. We were sleeping when she suddenly sat bolt upright, clutched her chest, and…she passed away." Fred was shocked. "When did this happen?" he asked. "About fifteen minutes ago, just after 2:00 a.m.," replied his father. (p. 24)

Here, since the experience appeared to convey information about the thoughts, feelings and sensations of another person, the report is labelled as telepathy. However, when the target event is assumed to be physical, the experience is labelled as *clairvoyance*. For example, Rhine (1961) tells of the case involving Emmanuel Swedenborg, an eighteenth century Swedish scientist and mystic, who in 1759 was able to describe the details of a fire that had suddenly broken out near his home some three hundred miles away. Regardless of the authenticity of such reports, it is assumed that since the information conveyed by the experience appears to have originated from the physical environment rather than from the mind of another person, the experience is labelled as clairvoyance[1].

---

[1] It should be noted, however, that classifying the target event as being either 'mental 'or 'physical' does not imply an ontological distinction. For example, under physicalism, 'mental' target events may

Temporal criteria determine whether the experience is labelled as *retrocognitive* (the target event occurred before the experience), *contemporaneous* (the target event occurred at the same time as the experience), or *precognitive* (the target event occurred after the experience). For example, the dream report provided at the start of this chapter could be categorised as precognitive because the corresponding real-world event was reported to have occurred after the dream had been experienced. Terms can also be combined to categorise a particular experience according to both the presumed substantive nature of the target event and its temporal relationship to the experience. For example, an experience labelled as precognitive telepathy is one that ostensibly correlates with a mental target event arising after the experience took place. However, in practice it is often unclear whether the information conveyed by an extra-sensory experience originated from the mind of another person, a physical object or event, or a combination of both. For example consider the following case from Gurney, Myers and Podmore (1886):

> I was walking along in a country lane at A, the place where my parents then resided. I was reading geometry as I walked along...when in a moment I saw a bedroom known as the White Room in my home, and upon the floor lay my mother, to all appearance dead. The vision must have remained some minutes, during which time my real surroundings appeared to pale and die out; but as the vision faded, actual surroundings came back, at first dimly, and then clearly. I could not doubt that what I had seen was real, so instead of going home, I went at once to the house of our medical man and found him at home. He at once set out with me for my home, on the way putting questions I could not answer, as my mother was to all appearance well when I left home. I led the doctor straight to the White Room, where we found my mother actually lying as in my vision. This was true even to minute details. She had been seized suddenly by an attack of the heart, and would soon have breathed her last but for the doctor's timely advent. I shall get my mother and father to read and sign this. (p. 194)

In this case, are we to categorise the experience as a case of clairvoyance or telepathy? On the one hand, it appears as if the percipient had direct extra-sensory access to information about the physical circumstances surrounding the medical emergency, in which case the report might be labelled as clairvoyance. On the other hand, one might also suppose that the percipient had access to information about the thoughts, feelings and perceptions of the mother who must have been aware of the physical circumstances surrounding her medical emergency, in which case the experience may be categorised as telepathy.

---

be ontologically equivalent to 'physical' target events. However, under dualism, 'mental' and 'physical' target events may be ontologically distinct.

The use of temporal criteria is less problematic in this respect since it is relatively straightforward to determine when a target event occurred relative to the experience, if sufficient information is available. However, commonsense notions of causality are particularly challenged when the target event occurs *after* the experience; as far as a parapsychological approach is concerned, precognition implies that people are able to anticipate the future without recourse to inference or deduction. Many parapsychological studies have attempted to test for precognition which can be defined as an anomalous correlation between current cognitive activity and a future event. In this endeavour to reproduce the phenomenon in a laboratory environment, where processes of inference and deduction can be controlled, the evidential problems inherent to spontaneous case reports can be avoided.

Outline of the thesis

The current thesis adopts a parapsychological approach to the investigation of ostensible precognition by asking whether anomalous correlations can be observed between current cognitive activity and the future presentation of experimental stimuli. Chapter 2 will present a selective review of the range of methodologies used to test for precognition and anomalous anticipatory effects. The experimental paradigm used in the current thesis will be introduced in Chapter 3 along with a description of the main hypothesis under test. Five experiments designed to test the precognitive hypothesis using the current paradigm will then be reported. Chapter 4 will present two studies involving simulations of a methodological artifact that has previously been suggested as an explanation for a number of precognitive effects and it will be examined whether such an artifact can potentially explain the results reported in the current thesis. The thesis will end with a general discussion of the experimental results and some methodological problems that may impact on their interpretation. A number of suggestions for further work will also be made.

Chapter 2 - Laboratory studies of precognition

Introduction

In the laboratory, a test for precognition usually involves measuring some aspect of a participant's behaviour in the present and observing whether this behaviour in some way correlates with the presentation of target stimuli in the future. Crucially, the experimental design must prevent the participant from obtaining any reliable information about upcoming stimuli by normal means, such as by inference or deduction. This is normally achieved by randomising the appropriate experimental variables. Numerous studies have been performed that follow this basic procedure using various response measures and experimental manipulations. The current section will describe these studies in detail. In doing so, the goal will be to provide a representative overview of the range of methodologies used in experimental tests of precognition. It will then be shown how the experimental paradigm used in the current thesis builds upon the methodology of previous research.

Before reviewing specific studies in the relevant literature, it may be instructive to define two types of experiment. On the one hand, an experiment may explicitly require participants to use precognition to obtain information about target stimuli, either by a strategy specified in the experimental instructions or by one of their own choosing. This may be classed as a study of *intentional precognition*. On the other hand, an experiment may ask participants to engage in a task that does not explicitly require them to use precognition. This may be classed as a study of *unintentional precognition*. Here, participants may or may not be informed that the experiment is investigating precognition. Both types of experiment may use behavioural responses, physiological recording of nervous system activity, or a combination of these measures as dependent variables. However, it should be noted that drawing a distinction between tests of intentional and unintentional precognition is not meant to imply that any effects observed in each class of experiment necessarily have different underlying causes. In addition, some studies do not fall neatly into one particular class of experiment. For example, in some tests of unintentional precognition, participants are informed about the nature of the experiment but are asked to play a passive role and refrain from using any intentional strategies. However, since participants are aware that the experiment is investigating precognition, they may adopt intentional strategies nonetheless. Unfortunately, few studies of this kind report steps taken to evaluate whether such strategies took place. Examples of experiments testing for precognitive effects that fall into each of these categories will now be discussed.

Studies of intentional precognition using behavioural measures

One of the simplest methods for investigating precognition in the laboratory has been the use of a *forced choice* design which, by definition, restricts a participant's response to a fixed set of alternatives. Forced choice designs testing for intentional precognition require participants to predict which of several alternative stimuli will be randomly selected as a *target* in the future. After a decision is made, participants are then typically presented with the target stimulus as feedback or they are given information about the accuracy of their response. Forced choice designs therefore allow responses to be statistically compared to the null hypothesis of obtaining a particular overall score by chance alone. Thus, if a forced choice experiment obtains an overall score that deviates significantly from theoretical chance expectation (while at the same time preventing relevant information about target stimuli from being obtained by normal means) this is regarded as evidence in support of psi – the anomalous process hypothesised to underlie the above chance effect.

For example, Rhine (1938) reports a series of forced choice precognition experiments carried out with a number of selected subjects using a deck of 25 cards printed with standard geometric symbols, termed Zener cards (Rhine, 1934; also see figure 1). Each deck contained 5 cards of each symbol and in each experimental run a subject was tasked with guessing the specific order of symbols that would result from a shuffled and dealt deck. Moreover, guesses were recorded before the deck was shuffled and dealt. A hit was recorded if a guess coincided with the actual symbol at a particular sequential position in the run. Therefore, 5 hits per run were expected by chance alone. In this series of experiments, an average of 5.14 hits were obtained in a total of 4523 runs. The observed number of hits can then be compared to the null hypothesis; that is, the expected score based on mean chance expectation of 5 hits per run. From this, the probability of obtaining the observed results by chance alone can be derived. The associated 1-tailed p-value for this series was 0.000003 which would result in rejection of the null hypothesis and might be interpreted as evidence for the precognitive hypothesis. However, one could argue that, in such studies, the card shuffling technique used to randomise the order of target stimuli was inadequate. If sufficient patterning exists in the sequence of target stimuli, this may increase the likelihood of a spurious match between responses and targets, leading to a false positive. For this reason, most contemporary precognition studies employ automated randomisation techniques such as the use of software-based pseudo-random number generators (pseudo-RNGs) that produce an output relatively free of bias.

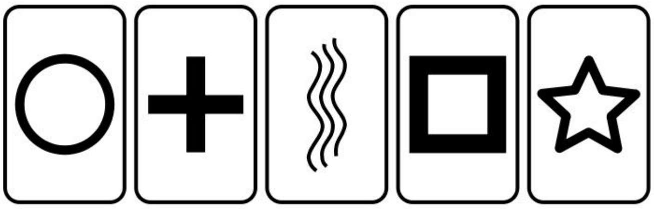


Figure 1. Symbols contained in a standard deck of Zener cards used in many forced choice studies. In early studies, 25 cards were used in a single deck with 5 cards of each symbol.

Honorton and Ferrari (1989) report a meta-analysis of 309 forced choice precognition studies conducted between 1935 and 1987. Although all studies included in this meta-analysis were of a forced choice design, the authors note that "the studies use a variety of methodologies, ranging from guessing ESP cards and other card symbols to automated random number generator experiments" (Honorton & Ferrari, 1989, p. 283). Unfortunately, the meta-analysis does not provide any information on the number of studies that were testing for intentional as opposed to unintentional precognition. However, it will be assumed that the majority of studies were intentional tests of extra-sensory ability since the forced choice technique was introduced for this purpose. The meta-analysis gave an overall combined z-score of 11.41, suggesting that the results were extremely unlikely to be due to chance alone. The authors also assessed the influence of a potential file drawer of null studies on their main finding using Rosenthal's 'fail-safe N' technique (Rosenthal, 1984). They concluded that an average of 46 studies with null results for every reported study in their meta-analysis would be required to reduce the main finding to chance level. However, Scargle (2000) argues that the fail-safe N technique can overestimate the number of studies in a file drawer required to nullify a particular finding because it assumes that the distribution of study effect sizes in the file drawer is unbiased. This is a problem because, by definition, a file drawer will be biased to contain fewer positive studies than expected by chance, a factor not taken into consideration in the fail-safe N calculation.

The overall effect size reported by Honorton and Ferrari (1989) was very small (r = 0.02). However, several interesting relationships emerged from the data. Firstly, effect sizes varied according to the degree of feedback given to participants. The largest effect sizes resulted from studies that provided trial-by-trial feedback. Studies that provided feedback after each run of trials produced the next largest effect size, followed by studies where the feedback was delayed (obtained mainly by postal delivery). Studies that provided no feedback to participants produced non-significant results with the smallest mean effect size. This suggests that feedback may be an important factor in studies testing for precognition and studies that provide trial-by-trial feedback to participants may stand the best chance of obtaining a positive result. It should be noted that this trend cannot be explained by sensory leakage since targets were selected after participants' responses were recorded. However, trial-by-trial feedback does allow for the possibility that participants were learning to respond to non-random structure contained in the sequence of trials as a result of poor randomisation. Secondly, effect sizes tended to increase as the time interval between the response and target selection was reduced. The largest effect size came from studies where this interval was on the order of milliseconds, whereas intervals of over a month produced a near zero effect size. However, as the authors note, this relationship was confounded with degree of feedback. Nonetheless, similar negative correlations have been reported in analyses of experiential data where the number of reported precognitive experiences decreased as the time interval between the experience and perceived target event increased (Green, 1960; Orme, 1974). However, in cases of reported experiences occurring outside the controlled conditions of the laboratory, this may have been due to normal cognitive factors such as forgetting rather than representing any underlying characteristic of precognition.

An alternative method of testing for intentional precognition is the *free response* design. Here, in contrast to forced choice designs where there are a fixed number of discrete responses that can be made for a given trial, the participant's response is not constrained by knowledge about the stimulus set. Instead, stimuli are randomly selected from a much larger pool of which the participant has little knowledge or expectation. For example, a large and diverse set of photographs collected by the experimenter may serve as the pool from which target stimuli are selected at random. Stimuli in free response experiments are typically visual such as realistic photographs or art prints (Krippner, Honorton & Ullman, 1971) but dynamic targets such as video clips are also used (Bem & Honorton, 1994). Precognitive free response designs usually proceed as follows. Firstly, a *receiver* is encouraged to describe their subjective impressions about a *target* stimulus (for example, a realistic picture) that will be randomly selected from a much larger pool in the future. After this period is over, the receiver's responses are sent to an independent *judge* who compares them against a set of four stimuli (randomly selected from the same stimulus pool). The judge must then decide which stimulus is the closest match to the receiver's mental impressions. After a first place match has been recorded, one of the four stimuli is randomly selected as the target. If the judge's selection and the target selection are the same, the trial is a 'hit'. The final stage usually involves presenting the target stimulus to the receiver as *feedback*. According to 'observational theory' (see Houtkooper, 2002, for a review), this is a necessary step for obtaining precognitive effects. Providing the receiver with feedback of the target stimulus also seems intuitively analogous to how precognitive experiences are reported to occur. That is, it is rare to find a case where the percipient of an experience does not at some point become aware of the external 'target' event that they have apparently precognised. During the entire free response procedure, care must be taken to ensure that the receiver, judge and experimenters do not know the identity of the target stimulus before the judge has selected their first place match. In practice, this is easily achieved by randomly selecting the target after completion of the judging phase. Results from free response experiments are statistically assessed in a similar way to forced choice experiments by comparing the observed hit rate with what would be expected by chance.

One of the first formal series of free response precognition experiments was conducted at Maimonides Medical Centre, New York, and involved a selected participant, Malcolm Bessent, attempting to dream about an audio-visual display that would be presented to him the next morning (Krippner, Honorton & Ullman, 1971, 1972). In a typical session, Bessent would be sequestered in a sleep laboratory and monitored for signs of REM sleep which has been shown to be associated with dreaming. Once a sufficient period of REM sleep had been observed, he was woken up and asked to report what he had been dreaming about. Waking Bessent fresh from the dream state in this way meant that he could report the details of his dream with greater clarity. He was then allowed to go back to sleep but was wakened a number of times through the night, each time making a record of his dream impressions. The next morning, an experimenter would randomly select a target 'theme' comprising of a visual and audio display from a pre-arranged pool which was then presented to Bessent as feedback. His dream transcripts were also sent to several independent judges who blindly evaluated them against the target display along with a number of randomly selected decoy displays and ranked each one according to how similar they appeared to the transcripts. The precognitive hypothesis was that Bessent's dream transcripts would be matched to the target more often than expected by chance. Two series of eight dream precognition trials were conducted with Bessent as receiver. Child (1985) analysed the data from each series in terms of binary hits (i.e., when the target was ranked higher than the median rank for all stimuli in the judging set) and found that both series obtained significant results (series 1, $p < 0.05$, $r = 0.73$; series 2, $p < 0.05$, $r = 0.65$ (effect sizes reported in Sherwood & Roe, 2003)) which suggested that Bessent was dreaming about events that he would experience in the future without using inference or deduction. On reviewing the Maimonides studies, which included another 11 experimental series investigating dream telepathy, Child (1985) could identify no major procedural flaws that may have accounted for the positive results. However, Child did point out that the original data had been handed to a number of independent statisticians who "must also have influenced the choice of procedures and measures" (p. 1223) and also raised concerns that the judging of each transcript against each target in the series was not carried out independently and therefore may have violated the independence assumption of the statistical tests.

Because dream studies proved to be relatively time consuming, parapsychologists turned to another technique, known as the 'ganzfeld' procedure, which is aimed at inducing a dream-like state of consciousness in a relatively short period of time while the receiver is awake and able to report their ongoing mental impressions. In this procedure, first used in parapsychology studies in the 1970s (Braud, Wood & Braud, 1975; Honorton & Harper, 1974), the receiver undergoes a period of sensory deprivation and relaxation whereby each of the visual, auditory and tactile senses are exposed to a uniform and unstructured input. For example, a uniform visual field is created by placing halved ping-pong balls over the receiver's eyes. Similarly, an unstructured auditory input such as white noise is played through headphones and the receiver sits in a comfortable reclining chair in order to reduce unwanted tactile sensations. It has been suggested that this sensory deprivation and relaxation technique facilitates psi by reducing external and internal sources of noise that may otherwise mask a relatively weak psi signal thereby increasing the signal to noise ratio, an idea that has become known as the noise reduction model of psi functioning (Honorton, 1977). After a short period of time in the ganzfeld state, simple perceptual disturbances are usually reported such as zigzagged lines in the visual field, but after more prolonged exposure (up to ten minutes) more complex, pseudo-hallucinatory images are experienced (Wackermann, Pütz & Allefeld, 2000). At this stage, the receiver is instructed to report their subjective impressions which are recorded and later sent to independent experimenters for judging against the target and several decoys as described previously. In a precognitive ganzfeld design, a target stimulus is randomly selected once the mentation session is over and the receiver is then either presented with the target as feedback while an independent experimenter performs the judging procedure or the receiver performs the judging procedure themselves while blind to the target. In telepathy designs there is another participant, called the sender, sequestered in an adjacent experimental room concentrating on the target stimulus and attempting to telepathically influence the receiver's mentation. In a clairvoyance design, the target picture is displayed in a remote location without any sender present. In both the telepathy and clairvoyance designs, the receiver or an independent experimenter performs the judging.

Several meta-analyses of the ganzfeld database exist (Bem & Honorton, 1994; Bem, Palmer & Broughton, 2001; Honorton, 1985; Hyman, 1985; Milton & Wiseman, 1999; Storm, Tressoldi & Di Risio, 2010a). Five out of six meta-analyses report an overall significant above chance effect. However, no consensus has emerged on whether these results can be interpreted as anomalous cognition rather than being due to flaws in the experimental methodology or problems relating to meta-analytical procedures (Hyman, 2010; Storm, Tressoldi & Di Risio, 2010b). The most recent meta-analysis (Storm et al., 2010a) was aimed at testing the noise reduction model and included both ganzfeld and non-ganzfeld experiments. A total of 59 free response studies conducted between 1997 and 2008 were included of which 10 were explicitly designed to test for precognition. Three groups of studies were formed. A *ganzfeld* group included studies that used the ganzfeld procedure to place receivers in a state of mild sensory deprivation, whereas studies that placed receivers in other states that could be regarded as 'noise reductive' such as dreaming, hypnosis, relaxation and meditation were classified as *non-ganzfeld noise-reduction* studies. Finally, a *non-ganzfeld, no noise- reduction* group was created from studies in which receivers were in a normal state of wakefulness. This meta-analysis found that the mean effect size for both the ganzfeld and non-ganzfeld noise-reduction groups was significantly greater than the effect size for the non-ganzfeld, no noise- reduction group which was at chance. This finding lends support to the noise reduction model of psi suggesting that participants were more able to attend to psi signals in the absence of irrelevant sources of noise (Honorton, 1977).

One particular theoretical problem should be mentioned here in relation to free response precognition studies. Although the target stimulus is typically selected at random after the receiver has reported their mental impressions, some studies also involve the receiver acting as the judge. The judging procedure involves direct observation of the target and decoy stimuli in order to rank each image against the receiver's mental impressions. Observational theories of psi (Houtkooper, 2002) suggest that if the receiver directly observes the target and decoy stimuli during the judging phase, this may confound the correct choice of target because information about the target *and* decoys may be obtained precognitively. In effect, the judging procedure can be conceptualised as two stages of 'observation'; initial inspection of the entire judging set and subsequent knowledge of the target via feedback. Observational theory suggests that examination of the judging set before the actual target is revealed provides the receiver with precognitive information about the content of the target *and* decoy images, while explicit knowledge of target identity after the judging phase is complete further biases the receiver's mentation towards features relevant to the target. Therefore, although the receiver's mentation is expected to contain relatively more precognitive information relevant to the target, it is also expected to contain some precognitive elements relevant to decoy images. These predictions from observational theory suggest that free response precognition studies should use independent judges so that the receiver can be presented with feedback of the target stimulus in isolation in order to increase the signal to noise ratio. Indeed, as noted by Bierman (2010), one of the most successful ganzfeld studies involved this very procedure in which receivers viewed the target after their mentation had been recorded and attempted to influence their subjective impressions 'retro-actively' while the judging procedure was performed blind by an independent experimenter (Wezelman, Gerding & Verhoeven, 1997). Using this procedure, Wezelman et al. obtained a 44% hit rate where 25% would be expected by chance.

In one prospective study designed to test the idea that the receiver, acting as the judge, can obtain information about decoys in addition to the target, Bierman (1988) ran 16 ganzfeld trials in which the receiver carried out the judging procedure. In addition, a second independent judge was presented with the stimulus set containing the target and a randomly selected control set of stimuli that did not contain the target and was asked to blindly decide which stimulus set they thought, as a whole, matched the receiver's mentation. After a decision had been made, the control set was removed by an experimenter and the same independent judge was asked to blindly decide which stimulus was the target from the target set.[2] The results showed, firstly, that the receiver tended to correctly identify the target with an overall hit rate of 37.5%. Secondly, the independent judge correctly identified the stimulus set containing the target 62.5% of the time compared to the control stimulus set. Lastly, on occasions where the independent judge ranked the actual target as third or fourth place (i.e., a binary miss), they correctly identified the target set 83% of the time. Bierman (1988) interprets these findings as evidence suggesting that the receiver obtained precognitive information about the decoy stimuli as well as the target stimuli, reasoning that on occasions where the independent judge failed to correctly identify the target, they nonetheless correctly identified the correct target set above chance, presumably by matching information from the receiver's transcript with the appropriate decoy images. However, Bierman (1988) points out that although all effects were in the predicted direction, they failed to reach statistical significance and suggests that low statistical power may have been a reason for this. Further studies should therefore attempt to replicate this suggestive effect using more participants. Alternatively, past studies may be re-examined using fresh independent judges attempting to blindly identify the original target set against a randomly selected control set using the receiver's original transcripts. However, since Bierman's hypothesis is that the receiver obtains information about the *decoys*, in addition to the target, it may be better to remove the actual target from the target set in such an analysis in order to ascertain whether correct identification of the target set is really due to information relevant to decoy stimuli or whether it is entirely due to the target stimulus.

Observational theories also imply that free response experiments using telepathy and clairvoyance designs, in which the receiver also performs the judging procedure, can be interpreted in terms of precognition since the judging involves observation of the target. For example, although only approximately one sixth of the ganzfeld studies included in the Storm et al. (2010a) meta-analysis were explicitly defined as precognition studies, it is possible that precognition was operating in the majority of studies intended to investigate telepathy and clairvoyance. Similar issues could be raised with interpretation of the Maimonides dream telepathy studies (Ullman & Krippner, 1989). One may wonder, therefore, whether re-examination of previous data would reveal a larger effect for studies that only provided receivers with feedback of the target compared to studies in which the receiver also acted as the judge.

---

[2] In this study, the experimenter who interacted with the independent judge was not blind to the target set or the target stimulus within the target set, therefore some degree of sensory leakage was possible, as acknowledged by Bierman (1988).

Studies of intentional precognition using physiological measures

As well as behavioural measures, studies have used indirect and direct measures of neural activity in conjunction with explicit precognition tasks. Warren, McDonough and Don (1992a, 1992b) used event related brain potentials (ERPs) as the dependent variable in a test of precognition using a selected subject, Malcolm Bessent, who had obtained above chance results on previous forced choice (Honorton, 1987) and free response studies (Ullman & Krippner, 1989). In each trial, Bessent was successively presented with four stimuli on a computer screen. During this period, his brain's electrical response to each stimulus was recorded with EEG. The computer then displayed all four stimuli together and Bessent was required to guess which one would be randomly selected as the target image. After Bessent had made his guess, the actual target was presented again as feedback. The results showed that Bessent scored at chance on the forced choice task but analysis of the ERP data showed that the evoked potentials associated with the initial presentation of target stimuli were of greater negative amplitude than those associated with decoy stimuli. This effect was characterised as a negative slow wave appearing between 150 to 500 ms after stimulus presentation with the effect being widely distributed over the range of scalp recording sites, suggesting that Bessent's brain was selectively responding to presentation of the target image before he was made aware of which stimulus was the actual target. Since the hit rate on the forced choice task was at chance, Warren et al. suggest that this was an unconscious discrimination between the target and decoy images that did not affect his decision making processes.

In a similar study, participants were successively presented with four pictorial stimuli while their heart rate was monitored (Sartori, Massaccesi, Martinelli & Tressoldi, 2004). The computer then displayed all four stimuli together and participants were required to guess which one would be presented to them as the target. Two conditions were used; in the clairvoyant condition the target picture was covertly selected, at random, at the start of the trial before any physiological responses were recorded, whereas in the precognitive condition the target was randomly selected after the participant had viewed all four stimuli and responded to the forced choice task. It was hypothesised that, in both the clairvoyant and precognition conditions, the mean heart rate while viewing targets would be different compared to decoys. This was based on evidence that autonomic nervous system activity can act as a marker of unconscious processing (Bechara, Damasio, Tranel, & Damasio, 1997).

The results showed that mean heart rate significantly increased while viewing target images compared to decoys in both the clairvoyant and precognitive conditions ($p < 0.001$ and $p < 0.05$ respectively). However, behavioural data showed no above chance guessing suggesting that the observed increase in mean heart rate was the result of unconscious precognitive processes. In two further studies, this effect was replicated (Tressoldi, Martinelli, Massaccesi, & Sartori, 2005) using a design essentially identical to that of Sartori et al. (2004); mean heart rate was significantly greater for targets compared to non-targets ($p < 0.05$ and $p < 0.01$ for studies 1 and 2 respectively). The p-values from all three studies were reported as two-tailed therefore the significant results cannot be attributable to inappropriately tailed tests. However, the reported effect sizes were all very small with no result exceeding a Cohen's d effect size measure of 0.06. An effect size this small may raise some concern that the results were due to some unidentified methodological flaw or artifact.[3]

In another forced choice experiment, Moulton and Kosslyn (2008) measured brain activity using fMRI while participants viewed emotional and neutral stimuli. Two stimuli were consecutively presented and participants were required to guess which one would be randomly selected as the target in the future (the experimental design also included a 'sender' viewing the actual target in a remote location thereby allowing for the possibility of telepathy and clairvoyance). It was hypothesised that brain activity associated with viewing targets-to-be would be different from decoys. However, the results of this study were null; behavioural data showed that participants did not score above chance and no differential brain activation for targets compared to decoys was found. Moulton and Kosslyn (2008) interpret their findings as "the strongest evidence yet obtained against the existence of paranormal mental phenomena" (p. 182), although the authors also note the logical difficulties in affirming the null hypothesis. However, one could argue that since their behavioural data indicated that participants were unable to guess the target stimuli more than expected by chance, this would suggest that psi was not operating on this particular occasion

---

[3] It should also be noted that the author of the current thesis noticed some discrepancies between the sign of the test statistics reported in Tressoldi et al. (2005) and the direction of the effects illustrated in Figure 2 of the same publication; the t-value and effect size for Experiment 1 were reported as negative in the main body of text yet the heart rate for targets was greater than for non-targets in Figure 2. In Experiment 2, the t-value and effect size were both positive and the heart rate for targets was also greater than for non-targets as illustrated in Figure 2. Again, in Sartori (2004), the t-value reported for the clairvoyant condition was positive whereas it was negative for the precognitive condition, yet the heart rate was reported as greater for targets compared to non-targets in each condition. It will be assumed that these discrepancies were due to typographical errors and that the heart rate was indeed higher for targets than for non-targets in every study.

and it is therefore unsurprising that no underlying neurophysiological psi effect was found. Nevertheless, a number of previous studies have indeed obtained evidence for underlying neurophysiological psi effects in the absence of behavioural psi effects (Sartori et al., 2004, Tressoldi et al., 2005; Warren et al., 1992a, 1992b) and therefore this argument is not particularly convincing. It might also be reasonable to suppose that the noisy environment of an fMRI scanner is not conducive to psi functioning, but again, positive results have been obtained using fMRI in at least one other study (Bierman & Scholte, 2002) although there were noticeable differences in the design and hypothesis of interest in each case. Perhaps the most reasonable explanation for the null results, aside from the non-existence of psi, is that the study was underpowered, having run only 19 participants. Indeed, a contrast involving the data from one participant did reveal significant differences in the BOLD response between targets and decoys in a number of brain areas. However, Moulton and Kosslyn (2008) account for this result in terms of differences in the visual and semantic properties of target and decoy stimuli presented to that particular participant due to the counterbalanced design that was used (stimulus properties were counterbalanced across the experiment but not within individual participant sessions).

Studies of non-intentional precognition using behavioural measures

Intentional precognition experiments involve participants attempting to predict the future by paranormal means. However, the majority of reported precognitive experiences seem to happen spontaneously and without the conscious intent of the individual (Stokes, 1997). Therefore, various studies have used methods designed to measure non-intentional precognition. This can be done in various ways, such as allowing participants simply to play a passive role in the experiment or by ensuring that participants do not know the experiment is about precognition.

Luke, Delanoy and Sherwood (2008) describe a non-intentional precognition experiment in which participants were presented with four fractal images on a given trial and asked to decide which one they preferred. Each participant completed ten trials in this manner. The computer then randomly and covertly selected one image as a ‘target’ for each trial. If a participant’s preference matched the randomly selected target, this was designated a hit.

Furthermore, if a participant obtained more hits than chance expectation over the ten trials, that participant went on to perform a secondary task involving looking at erotic images. However, if the participant scored below chance expectation they went on to perform a boring vigilance task involving paying attention to strings of digits. Moreover, the length of time allocated to performing the pleasant and unpleasant secondary tasks was contingent on a given participant's level of performance in the covert precognition task in that a greater number of hits (or misses) resulted in a longer duration for the pleasant (or unpleasant) task. The experiment was classed as a study of non-intentional precognition because participants were unaware that they had to use any psychic ability in the preference task. Based on a model proposing that psi functions to serve the current needs of an individual (Stanford, 1974), it was predicted that participants would choose the target images more often than chance expectation, thus leading to a favourable outcome. This prediction was confirmed; the results showed that the target fractals were preferred an average of 2.85 times when 2.5 was expected by chance, a result that was statistically significant. Participants in this study were also classified as erotically reactive or un-reactive based on whether they scored above or below the mean on an erotic reactivity scale originally constructed by Bem (2003). In a planned analysis, participants who were classed as erotically reactive obtained more hits in the precognition task than erotically un-reactive participants. Together, these results suggest that an anomalous process was biasing decisions in the preference task in an unconscious but goal-directed manner in such a way that participants were rewarded with a pleasant or favourable outcome consistent with their particular disposition. Although the authors do not discuss potential normal explanations for these results, the randomisation of target fractals and their relative on-screen position was checked and found to be within chance expectation. Indeed, since trial-by-trial feedback was not provided, this would eliminate the possibility of implicit or explicit learning of any small bias that may have been present in the position of target fractals.

In a further two replication attempts by Luke, Roe and Davison (2008), participants were again presented with a preference task in the same way as in Luke, Delanoy and Sherwood (2008). However, there were some procedural differences compared to the previous study. In experiment 1, participants were randomly allocated to either a contingent or a non-contingent condition. The contingent condition involved a pleasant or unpleasant secondary task, the event of which was contingent upon performance in the preceding preference task as before. However, the pleasant secondary task did not involve viewing erotic pictures as in the previous study, but involved rating cartoon images for humorousness.[4] The unpleasant task in the contingent condition was comprised of paying attention to strings of digits as in the previous study. In the non-contingent condition, the experiment simply ended after the preference task, i.e., participants were not given a secondary task to perform. The results, when collapsed across both conditions, showed that participants chose the target fractals more often than chance expectation. Surprisingly however, when conditions were considered separately, the non-contingent condition resulted in a significant above chance hit rate whereas the contingent condition was marginally non-significant and in the predicted direction, although the difference between these conditions was not significant. This result was somewhat inconsistent with the experimental hypothesis since the non-contingent condition was not explicitly associated with any reward and was therefore expected to generate fewer hits than the contingent condition. In discussing this finding, Luke, Roe and Davison (2008) speculate that an eagerness to perform well at the task may have been enough to provide an element of reward for participants allocated to the non-contingent condition. Nonetheless, the non-contingent condition was abandoned in experiment 2 in favour of ensuring that the contrast between the pleasant and unpleasant conditions was valid by administering a post-session questionnaire to participants. Experiment 2 followed the same basic design as experiment 1 in that the pleasant secondary task consisted of judging the humorousness of cartoons whereas the unpleasant task was the same vigilance task as previously described. Again, the results showed that target fractals were chosen more often than expected by chance, leading to a greater propensity towards pleasant secondary task outcomes as predicted. Unfortunately, information about the adequateness of the randomisation output was not provided in Luke, Roe and Davison (2008). Therefore this potential explanation for the results cannot be addressed.

Precognitive effects have also been reported using reaction time as the dependent variable. In a series of studies, Klintman (1983, 1984) reported an apparent psi effect using a variation of the Stroop test. In each trial, participants were initially presented with a coloured rectangle and were required to name the colour as fast as possible. This was followed by the name of a colour-word and they were then required to indicate whether the name of the colour-word matched the colour of the previous rectangle. Based on a large body of previous research (MacLeod, 1991) the standard Stroop hypothesis predicts that reaction times to the colour-name would be slower when the colour name mismatched the colour of the rectangle. This

---

[4] This was due to ethical considerations since participants were recruited by convenience from a public exhibition on superstition at the Northampton Museum and Art Gallery.

was indeed the case but Klintman also found that when this mismatch occurred, reaction times to name the colour of the rectangle were slower even though the match and mismatch conditions were randomly determined. This was interpreted as a “time-reversed interference” effect, in that the effect of the mismatch condition was able to reach back in time and influence the time taken to name the coloured rectangle. A replication was attempted by Camfferman (1987) who obtained positive results but concluded that the effect may be explained by differences in participants’ level of alertness since the reaction time to name the coloured rectangle and the reaction time to name the colour word were not independent. Another replication attempt was carried out by Radin and May (2001) who obtained positive results. However, their method of analysis differed in many ways to Klintman (1983, 1984) and Camfferman (1987). Savva and French (2002) ran three studies attempting to replicate Klintman’s original findings. However, when they applied both the analyses used by Klintman (1983, 1984) and Camfferman (1987) to the data from their first study, results were obtained in the opposite direction in each case with the analysis based on Camfferman (1987) obtaining significant results in the predicted direction. A further two studies failed to replicate the ‘time-reversed interference’ effect suggesting that the effect is not as robust as originally reported in Klintman (1983, 1984). The fact that different methods of analysis produced results in opposite directions is also cause for concern in these experiments and suggests that the effects may be artifactual.

Bem (2011) reports another series of experiments examining precognitive effects using reaction time measures and other standard behavioural measures. Bem’s approach was to take well established psychological effects and to ‘time-reverse’ the order of stimulus and response. For example, in one study, participants were given a precognition task that was a ‘time-reversed’ version of an affective priming experiment. In each trial, a picture (for example, a beautiful landscape) was presented on the screen and participants were required to indicate as fast as possible whether they thought the picture was pleasant or unpleasant. After they made their response, a prime word was briefly flashed on the screen, the valence of which was either congruent or incongruent with the picture. The hypothesis was that the mean reaction time to pictures on congruent trials would be faster than on incongruent trials even though the prime was presented after the response had been made. The results of Bem’s first retro-active priming study showed that participants were approximately 15ms faster to name the pictures when the prime word was congruent even though the category of trial was randomly determined and the prime was displayed after the participant made their response. Because the retro-active hypothesis stated that the results would mirror the effect of a standard, non-retroactive procedure, Bem also ran the same participants on a normal, forward-priming version of the experiment using the same stimulus materials. This standard procedure produced the expected results confirming that responses were approximately 22ms faster when a congruent prime was flashed before presentation of the picture. In a second study, Bem found the same pattern of significant results; participants were approximately 20ms faster on congruent trials in the retro-active priming procedure and approximately 29ms faster on congruent trials in the forward priming procedure.

In another study, termed precognitive habituation, Bem drew on the idea that repeated exposure to positively arousing and negatively arousing stimuli renders the stimuli less arousing (more neutral) upon subsequent viewing. Previous research established that this effect is strongest when the repeated exposure is subliminal (Bornstein, 1989). Therefore, Bem reversed the procedural order of events to investigate a retro-active version of this effect. Participants were asked to choose which of a pair of strongly arousing, negative pictures they preferred. After a response had been made, one of the pictures was randomly selected as a target and repeatedly displayed to the participant at a subliminal level. Bem hypothesised that subliminal presentation of the target after the preference task would retro-actively render the target less negatively arousing, biasing participants to prefer the target picture over the decoy picture in the preference task. This hypothesis was confirmed with participants preferring the target picture 53.1% of the time when 50% was expected by chance ($p = 0.01$). Trials with neutral picture pairs were also administered which obtained non-significant results as predicted. In a further replication attempt, Bem introduced positively arousing picture pairs and hypothesised that retro-active habituation would render the target picture less positively arousing, biasing participants to prefer the decoy picture. The results showed that participants did indeed prefer the decoy pictures on trials with positively arousing picture pairs and the previous results with negative picture pairs were also replicated.

In a final set of two studies, Bem ‘time-reversed’ a facilitation of recall procedure. Participants were given a list of nouns and then asked to recall as many as they could. This was followed by a practice session on a randomly selected subset of words from this list. Normally, rehearsal before recall enhances recall of practiced words. However, in both studies, Bem found that a practice session after the recall session resulted in recall of more to-be-practiced words, suggesting that the practice session retro-causally influenced recall. A larger effect was found in the second study which involved a deeper level of encoding in the practice phase.

In total, Bem (2011) reported nine time-reversed experiments, finding a significant main effect in all but one study. Many secondary analyses were also performed on subsets of data examining personality variables, gender differences and various stimulus presentation parameters but such findings are beyond the scope of this review. The mean effect size across all experiments was d = 0.22 which according to Cohen (1988) is suggestive of a small effect. A number of criticisms have been levelled at the methodology described in Bem's paper, mostly focussing on the statistical treatment of data. For example, several authors have raised concerns over the use of one-tailed tests (Alcock, 2011; Wagenmakers, Wetzels, Borsboom & van der Maas, 2011) and others have argued that frequentist statistical methods used in null hypothesis testing are problematic (LeBel & Peters, 2011) with some suggesting that Bayesian methods be used to assess the evidence for the precognitive hypothesis (Wagenmakers et al. 2011; Wagenmakers, Wetzels, Borsboom, Kievit & van der Maas, 2011). LeBel and Peters (2011) grant that Bem's methods often exceed the accepted standard of modal research practice in social psychology (Cook & Groom, 2004) but claim the extraordinary nature of Bem's hypotheses and confirmatory results suggest that such standards need to be improved.

Several independent replications of Bem's work have been attempted. Batthyany (2010) reports one positive replication of the precognitive habituation effect. Using high-arousal negative pictures, the results showed that participants preferred the target-to-be on 53.3% of trials, an effect that was of similar magnitude to that obtained in Bem's original study. However, participants in Batthyany's replication were tested in groups of five, each of whom viewed the same trial projected on a canvas (participants indicated their preference on each trial by writing on separate scoring sheets). Consequently, responses in each group cannot be regarded as independent which may have introduced a statistical 'stacking' effect similar to that described in the forced choice ESP literature (Davis, 1978) with the result that a p-value generated from a statistical test that assumes independence may be misleading. Ritchie, Wiseman and French (2012) and Robinson (2011) report exact replication attempts of the retro-active recall effect using the same software as used in the original Bem (2011) study. The Ritchie et al. study involved three separate replication attempts at independent laboratories, all of which failed to find any support for Bem's original findings despite the fact that the experiments were attempting to replicate the study with largest reported effect size (d = 0.42). Similarly, the results from Robinson (2011) were null. However, it is possible that a failure to replicate the retro-active recall effect was due to a lack of participant naivety about the details of the experimental procedure. During the study phase of the retro-active recall experiment, participants are not required to explicitly encode the list of words. Rather, encoding is assumed to be incidental since participants are only required to visualise each word in the list and are not supposed to be aware of the impending recall test. However, if the recall test was no longer a surprise to participants in the replication attempts, perhaps because they were made aware of the procedure from prior media coverage of Bem's original studies, they may have been using explicit encoding strategies to memorise the words in preparation for the recall test.[5] Whether this would make a difference to the expression of retro-active recall effects remains to be seen but, at least, this may be a reason to suppose that these particular replication attempts were not 'exact' as one might imagine.

Studies of non-intentional precognition using physiological measures

Physiological measures have also been used in non-intentional precognition experiments. Don, McDonough and Warren (1998) presented self-reported gamblers with playing cards in a gambling task while measuring ERPs. Cards were presented one at a time to participants while the electrical activity of their brain was monitored. All four playing cards were then presented together and participants were required to guess the winning card. Participants did not wager any money and performed the task 'just for fun'. In this respect, the design was very similar to the intentional forced choice precognition experiments using EEG described previously (Warren et al., 1992a, 1992b). The only difference in this case was that participants were unselected for previous psychic ability and were unaware that the experiment was testing for precognition. The resulting ERPs associated with targets and decoys were analysed and the results showed that there was a significantly more negative slow wave associated with viewing targets, with the effect occurring over the left hemisphere. This partially replicated their previous findings with a selected subject (Warren et al., 1992a, 1992b). In another replication attempt McDonough, Don and Warren (2002) ran 20 participants through the same task. Again, ERPs were used to measure any

---

[5] I would like to give credit to Michael Franklin who put forward this argument in the comments section of the online Ritchie et al. (2011) article.

precognitive effect and participants were selected for self-reported gambling habits. The results were marginally significant and suggested that processing of target stimuli was associated with a more negative slow wave but over right hemisphere sites, a result that only partially replicated previous findings. Thus, all studies in this series of replications involved a negative going component in the resulting ERPs, although as noted by McDonough et al. the hemispheric localisation of the effect was inconsistent from study to study. McDonough et al. (2002) also note that negative components of ERPs tend to be observed in studies involving attentional processing. In particular, they point to previous research identifying negative components peaking at approximately 260ms during the processing of infrequent stimuli amongst a sequence of frequently occurring stimuli (Näätänen, 1982) and suggest that the relatively infrequent target stimulus amongst a group of frequently occurring decoys may have been processed in a similar manner. Similarly, McDonough et al. (2002) draw a comparison between their results and those of McCallum, Barrett and Pocock (1989) who observed a negative going slow wave in frontal regions that was associated with stimuli delivered to attended channels in a selective auditory attention task and those of De Jong, Kok and van Rooy (1988) who also observed a frontal negative slow wave in response to attended stimuli. However, the precognitive ERP effect has yet to be replicated by an independent laboratory and the fact that its localisation differed between studies raises the possibility that the effect was due to methodological artifacts, perhaps originating from choices in data processing (Simmons, Nelson & Simonsohn, 2011).

Another method of testing for non-intentional precognition has become known as the *presentiment* paradigm. In a typical experiment, some measure of a participant’s physiology is continuously recorded while they are presented with a random sequence of dichotomous stimuli, such as calm versus emotional pictures. Participants are not instructed to respond to the stimuli in any specific way but instead are asked to remain passive while their physiological response is measured. For example, Radin (1997b) measured participants’ electrodermal activity (EDA) while presenting them with a random sequence of emotional and calm photographs. Each trial was initiated by the participant by pressing a key and, after a few seconds of blank screen, a randomly determined calm or emotional picture was presented for three seconds. This was followed by a blank screen for a few more seconds after which the trial was terminated and the participant was free to initiate the next trial when ready. Normally, an individual’s EDA increases in response to presentation of an emotional picture, whereas it remains relatively inactive in response to a neutral picture. Radin (1997b) observed this normal response but also observed higher EDA a few seconds before presentation of emotional pictures compared to calm pictures despite the fact that their occurrence was randomly determined. Since participants did not report any conscious awareness about which type of stimulus was about to be presented, the data can be interpreted as an unconscious anticipation of emotional stimuli, reflected in the activity of the autonomic nervous system. Radin (2004) reports a series of replications again using EDA as the dependent measure. One study out of three replication attempts yielded a significant positive result, with the remaining two experiments yielding non-significant results but in the predicted direction. When data from all four experiments was combined, there was an overall significant effect and pre-stimulus EDA activity positively correlated with independent ratings of stimulus emotionality suggesting that participants were also anticipating the degree of emotionality of the upcoming stimulus. However, this correlation was very small and achieved statistical significance due to a large number of data points ($r = 0.04$, $p = 0.008$, $N = 4569$). Radin also considers a number of alternative explanations for his results including sensory and statistical cues, procedural artifacts, selective reporting, fraud and expectation strategies but ultimately rejects them. The latter explanation, discussed in detail in Chapter 4 of the current thesis, concerns a bias that can arise from intra-trial expectation effects (Dalkvist & Westerlund, 2006; Dalkvist, Westerlund & Bierman, 2002; Wackermann, 2002). Essentially, participants may form an expectation about the upcoming stimulus based on the outcome of previous trials. For example, participants may believe that after a long sequence of calm trials, the next trial is likely to be an emotional trial (in fact, this belief is fallacious since the outcome of each trial is randomly determined and independent). As a result, their EDA may increase as a function of the number of consecutive calm trials and reset to baseline levels after viewing an emotional trial. If such behaviour occurs, then EDA will tend to be at a peak level before presentation of emotional trials which may lead to the appearance of a presentiment effect when, in fact, the result is due to an artifact. Radin (2004) considers this explanation and rejects it as a plausible account for his results because his data did not show that participants' skin conductance levels were increasing in response to consecutive calm trials.

Two successful replications of the presentiment effect have been reported that also measured EDA but used randomly presented audio startle stimuli to elicit a physiological response (May, Paulinyi & Vassy, 2005; Spottiswoode & May, 2003). Here, instead of trials being initiated by the participant, a continuous 'trial' was initiated by the experimenter at the start of the session and stimuli consisting of either a burst of white noise or a control period of silence were presented at random intervals. In both studies, EDA was observed to be higher a few seconds before noise stimuli compared to periods of silence. However, Broughton (2004) ran a presentiment experiment using calm and emotional pictures while measuring EDA but found no effect despite having selected 220 of the most arousing and 220 of the least arousing pictures from the International Affective Picture System stimulus set (Lang, Bradley & Cuthbert, 1995).

Two studies report positive results using heart rate as the dependent variable to investigate anticipatory responses to neutral and alerting acoustic stimuli (Tressoldi, Martinelli, Zaccaria, & Massaccesi, 2009; Tressoldi, Martinelli, Scartezzini & Massaccesi, 2010). However, a number of methodological problems can be identified that can potentially explain the results of these studies. Tressoldi et al. (2009) followed a simple presentiment design whereby each participant was presented with 10 neutral and 10 alerting sounds in random order. A significant effect was found in that heart rate before neutral stimuli was lower than before alerting stimuli ($d = 0.13$, $p < 0.0001$). However, using a fixed number of stimuli in this manner may introduce statistical cues that participants may notice and capitalise on in order to predict the upcoming stimulus at a rate greater than chance. Indeed, the experimental instructions actually informed participants that they would receive 10 neutral and 10 alerting trials during their session. A participant would therefore have only needed to keep a simple count of the number of trials of each category to, at least, predict the last trial with certainty. Unfortunately, Tressoldi et al. (2009) do not examine their data for this possibility or discuss statistical cueing as an alternative explanation for their results. The same methodological issue is present in Tressoldi et al. (2010) where positive results were also obtained.

Tressoldi, Martinelli, Zaccaria and Massaccesi (2011) use pupil dilation responses to neutral and alerting sounds to test for presentiment effects. They employed a novel design in which a preliminary ‘calibration’ session gathered data on individual participant pupillary responses to 10 neutral and 10 alerting sounds. For each participant, an average pupil dilation measure for each category of stimulus was calculated. Following that, the second session involved presenting the neutral and alerting sounds in a random order. However, this time, the computer compared the participant’s pre-stimulus pupil dilation on a given trial with their individual baseline values calculated from session one. This calculation was carried out online during the two second period before stimulus presentation. Thus, if a participant’s pre-stimulus pupil dilation on a given trial in session two was closer to their neutral baseline value then the computer predicted that the trial was about to play a neutral sound. If, however, the pre-stimulus pupil dilation was closer to their alerting baseline value, an alerting trial was predicted. In this way, a hit or a miss was recorded for each trial depending on the prediction that was made from the pre-stimulus pupil dilation data. However, the final analysis presented in Tressoldi et al. (2011) consisted of calculating the number of hits separately for each stimulus category. They report that the hit rate for alerting trials was 60.3% (binomial, $p < 0.00001$) but incorrectly assume that chance expectation was 50%. The reason this method of analysis is flawed is as follows. Suppose that in the calibration session, alerting sounds caused a participant's pupils to dilate more than neutral sounds resulting in a greater mean baseline value for that category. Now suppose that in session two, the same participant's pupils were on average more dilated compared to session one. This could have been due to a number of normal reasons such as adaptation to the lighting conditions. If such dilation was extreme (for the sake of argument), an alerting sound would have been predicted on every trial in session two leading to a 100% hit rate when alerting trials were analysed separately. Of course, the corollary of this scenario would be a 0% hit rate for neutral trials and, if all data was considered, the expected 50% hit rate would result. Indeed, the data presented in Tressoldi et al. (2011) suggests this pattern of results; the hit rate for neutral sounds was 44.6% (binomial $p < 0.01$), although the authors do not report this analysis. A valid method of analysing the data from this experiment would simply be to look at the overall hit rate which was 52.4% and significant with a one-tailed binomial test ($p = 0.03$). Therefore, weak evidence for a presentiment effect was indeed present in the data. However an explanation in terms of statistical cueing cannot be ruled out here since, like in Tressoldi et al. (2009) and Tressoldi et al. (2010), a fixed number of neutral and alerting sounds were administered and participants were made aware of the relative frequencies of each type before the session began.

A number of studies have tested for presentiment effects using neurophysiological measurements. McRaty, Atkinson and Bradley (2004a, 2004b) report a replication attempt using concurrent measurements of skin conductance, heart rate and brain activity. The same basic design as Radin (1997b) was used except that the number of calm and emotion trials presented to each participant was fixed, although sequenced in a random order. McRaty et al. found a significant presentiment effect when they examined heart rate and EEG but no effect was found for skin conductance. The EEG data showed that the ERP associated with the pre-stimulus period was more negative for emotional trials compared to calm trials over left and right frontopolar regions, suggesting a possible involvement of attentional processes. In another study, Bierman and Scholte (2002) used fMRI to test for precognitive brain activity preceding emotional and neutral pictures. A presentiment effect was found for females who showed anticipatory brain activity preceding erotic and violent pictures, mainly in the visual cortex. However, males only showed anticipation to erotic stimuli and the results for both analyses were marginally significant. The authors therefore regard the study as exploratory since a number of analyses were performed on subsets of the data. Furthermore, lack of an fMRI presentiment effect in frontal regions is inconsistent with the results of McRaty et al. (2004b) who found a precognitive negative going potential preceding emotional stimuli in frontal regions using EEG, although the poor temporal resolution of fMRI may have contributed to this discrepancy.

Another EEG presentiment experiment was carried out by Radin and Lobach (2007) who tested for precognitive anticipation of simple visual stimuli. The design followed closely that of Radin (1997b) where participants initiated each trial with a button-press and, after a few seconds, were randomly presented either with a flash of light or no flash. A significant presentiment effect was only obtained for females with the effect characterised as a more positive voltage preceding the flash condition compared to the no-flash condition over the occipital region. Males, on the other hand, did not show this effect. However, it is unclear why the results from this study were presented separately for each gender (an analysis of all the data was not included). Therefore, replication of such gender effects would be required before drawing firm conclusions from this study.

Radin and Borges (2009) report a presentiment study testing for anticipatory eye movements and pupil dilation before presentation of emotional and neutral pictures. Again, the general design was based on that of Radin (1997b). However, in an attempt to increase the effect size, the planned analysis compared the 5% most emotional to the 5% most calm pictures from the stimulus set, based on their independent ratings of arousal (stimuli were taken from the International Affective Picture System (Lang, Bradley & Cuthbert, 1995)). They hypothesised that emotional pictures would be preceded by greater pupil dilation which was confirmed. Like previous studies using EDA as a measure of autonomic arousal, the increase in dilation reflected the post-stimulus response to viewing the emotional content of the stimulus. In addition, participants moved their eyes to the left significantly more before emotional pictures. Radin and Borges (2009) interpret this finding as suggestive of precognitive, right-hemisphere processing of emotional valence. However, the emotional pictures used in their study may have contained more visually interesting features on the left compared to neutral pictures and therefore the anticipatory eye movements may have reflected a precognitive orienting of visual attention. This interpretation is supported by their data (Radin & Borges, 2009, figure 7, p. 207) where mean gaze position is biased to the left side of emotional pictures when they were actually viewing these stimuli.

Summary of laboratory studies

The above review presents a non-exhaustive overview of the range of methodologies and findings from studies testing for conscious and unconscious forms of anomalous anticipatory effects. Behavioural measures included forced choice, judgement and preference decisions, reaction time and word recall, whereas physiological measures included electrodermal activity, heart rate, brain activity, pupil dilation and eye movements. The extent to which methodological flaws can be identified in each study is variable, with some studies better designed to test for precognition than others. The issue of methodological flaws has been systematically addressed with regards to experimental paradigms such as the dream ESP studies (Child, 1985) and the ganzfeld database (Bem, 1994; Hyman, 1994, 2010; Storm et al. 2010a, 2010b). In the case of ganzfeld studies, Storm et al. (2010a) report that the effect sizes of studies conducted between 1992 and 2008 did not correlate with methodological quality ratings from two independent judges. However, when taking into account the entire ganzfeld database (1974 – 2008), they document a slight linear decline in mean study effect size over time, although they advise caution in assuming this decline can be explained in terms of increasing methodological quality. Indeed, Storm et al. note that a quadratic curve fitted to the same data suggests a significant ‘rebound’ in effect sizes due to studies conducted after around 2000. Preliminary results from a meta-analysis of all presentiment studies conducted to date are reported in Tressoldi (2011). The overall mean effect size for the presentiment database was 0.26 ($Z = 8.7$) using a fixed effects model and 0.28 ($Z = 6.07$) using a random effects model, while an effect size of 0.12 (fixed effects) and 0.13 (random effects) was reported for the entire ganzfeld database as a comparison. The results of a Bayesian analysis on the presentiment data was also reported which provided strong evidence in favour of the alternative hypothesis. However, Tressoldi (2011) reports these findings as a summary of the results of a meta-analysis submitted for publication (Mossbridge, Tressoldi & Utts, submitted) and it remains to be seen whether the published analysis will provide details on whether effect sizes in the presentiment database are related to study quality. Indeed, as described previously, some presentiment studies (e.g., Tressoldi et al., 2009, 2010, 2011) suffer from series methodological weaknesses and one may wonder what influence these particular studies may have on the results of the above mentioned meta-analysis.

A role for attention in precognition?

Although the existing literature on precognition contains a body of positive studies that would benefit from further replication, there has been relatively little research aiming to establish whether there are particular cognitive systems that play a central role in the expression of these effects. Nevertheless, a number of previous studies investigating precognition using non-intentional and physiologically based measures suggest the potential involvement of selective attention (Don et al., 1998; McDonough et al., 2002; McRaty et al. 2004b; Warren et al., 1992a, 1992b). Although psychologists may differ in their precise definition of attention, there is general agreement that it is involved in a great deal of perceptual and cognitive processing. Furthermore, it is not a unitary phenomenon but has various forms tailored to different processing contexts. For example, we can voluntarily maintain our focus of attention on certain features of the environment or on internal representations such as mental imagery. We can divide our attention between multiple tasks and select information from a range of available sources for more detailed processing. Attention can also be captured involuntarily by salient stimuli in the environment while thoughts and feelings can unexpectedly intrude on our stream of consciousness.

Given the pervasive role of attention in cognition, could this processing system play a significant role in precognition? Such a question has not been directly addressed in parapsychology with experimental and theoretical studies predominantly focussing on memory as a potential substrate upon which the anomalous process presumably underlying extra-sensory effects occurs (Broughton, 2006; Irwin, 1979; Palmer, 2006; Roll, 1966). A shortage of studies looking at the role of attention is perhaps surprising since the information conveyed in reports of extra-sensory experiences often appears to interrupt and demand attention, reminiscent of the way a salient stimulus can achieve the same effect.

An experimental paradigm was therefore developed that would allow for a test of precognition within the context of a task designed to measure the allocation of attention to visual stimuli using eye tracking techniques. This novel paradigm will be described in the next chapter followed by five empirical tests of the precognitive hypothesis using variations on this basic methodology.

Chapter 3: Testing for precognition using eye tracking techniques

Development of the current paradigm

This chapter describes a series of experiments using a novel paradigm designed to test for precognitive effects in the visual attention system using eye tracking as a behavioural measure. In order to examine unconscious forms of precognition, previous studies used an experimental design that can be described as a test of whether an individual's neuropsychological system responds to the initial presentation of a stimulus in anticipation of its repeated presentation, at a time when no such predictive response could be made. The general methodology of this paradigm, here termed *sequential presentation*, is shown in figure 2. Each trial involves measuring continuous physiological activity while a number of stimuli (usually four) are briefly presented in sequence. Following this, participants are typically given a forced choice task where they are required to guess which one of those stimuli will be chosen as the target for that trial (however, the forced choice task is not an essential design feature). After a decision has been made, the target is randomly selected and presented to the participant again as feedback. When the experiment is complete, physiological activity associated with the initial presentation of the target and decoys is then compared. This type of design tests the hypothesis that an individual can selectively respond to the initial presentation of the target stimulus at a time when its identity cannot not be deduced or inferred, whether that selective response is measured via electrodermal activity, heart rate, nervous system activity or some other physiological metric.

As discussed in Chapter 2, a number of studies have used this paradigm in conjunction with physiological and neurophysiological measures such as heart rate (Sartori et al., 2004), fMRI (Moulton & Kosslyn, 2008) and EEG (Don et al., 1998; McDonough et al., 2002; Warren et al., 1992a, 1992b). Of the studies that obtained positive results, the data suggest that the selective response to targets was at an unconscious level that did not influence overt decision making since overall performance on the forced choice task was at chance. However, the overt behaviour probed by the forced choice task may have been insensitive to the level at which the selective response to targets occurred. Indeed, a meta-analysis of forced choice precognition studies obtained an overall mean effect that was very small (Honorton & Ferrari, 1989) suggesting that forced choice tests are a relatively ineffective behavioural means to test for precognition. In addition, a number of studies based on the sequential presentation paradigm obtained suggestive evidence for the involvement of attention in the expression of anomalous anticipatory effects (Don et al., 1998; McDonough et al., 2002; Warren et al., 1992a, 1992b).

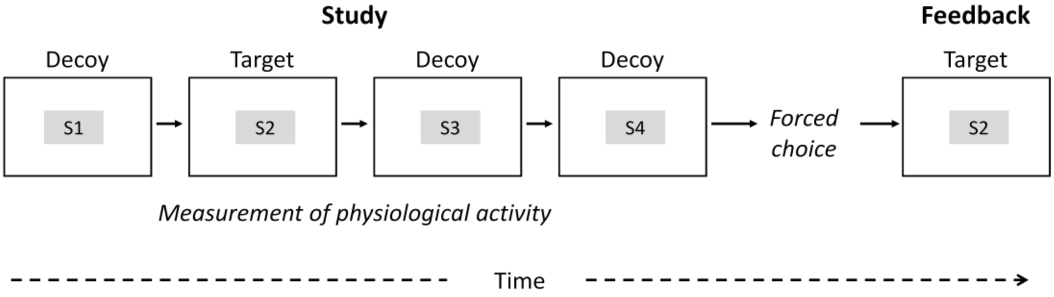


Figure 2. The Sequential Presentation Paradigm: In the study phase, stimuli S1-S4 are presented in a random sequence while physiological activity is recorded. The participant then makes a forced choice about the target identity. Finally, the actual target is randomly selected (in this case S2) and displayed as feedback.

If attentional processes underlie this type of anticipatory effect, as previous research suggests, use of a behavioural index of attention may be an alternative means to test for selective responses to target stimuli. Eye movements represent one such behavioural index of the visual attention system. Where and when we move our eyes reflects the kind of visual information we have decided to process in detail at a given point in time. Typically, we move our eyes approximately three to four times per second (Henderson, 2003; Rayner, 1998) and the majority of this behaviour consists of a succession of *fixations* interspersed with rapid *saccades* that serve to reposition the eyes on regions of interest.[6] Visual processing is effectively blocked during saccades so that we do not experience adverse motion blur and the vast majority of sensory input is derived from fixations which align fixated regions with the fovea, allowing them to be processed to a much higher degree of detail. Thus, eye movements are an expression of *overt visual selection* because they result in enhanced processing of regions and features of the visual field by a limited pool of

---

[6] The eyes are also able to track moving stimuli and while doing so adopt a smooth motion called *pursuit*. However, pursuit does not occur when stimuli are motionless as is the case in the current paradigm. Therefore, this thesis will limit discussion of eye movements to fixations and saccadic movements.

resources. While eye movements reflect overt selection, *covert selection* refers to our ability to shift attention between features or locations of the visual field while our gaze remains stationary. Although research has shown that spatial shifts in covert attention can confer a certain degree of processing benefit at selected locations in the peripheral field (Eriksen & St James, 1986; Posner, 1980), a clearly more effective method of spatial selection involves moving the eyes so that the fovea is aligned with the region of interest.

Several factors can influence whether a region or feature of the visual field is selected as the target for a fixation. Firstly, selection can be based on the physical properties of a stimulus. For example, attention might be involuntarily captured by a stimulus with a unique physical feature such as a small patch of red in a uniform grey background which may trigger an eye movement to that location. Visual search experiments show that salient distractor items can trigger an involuntary eye movement towards their location even when such behaviour is in conflict with current task goals (Theeuwes, Kramer, Hahn & Irwin, 1998). The physical properties of a fixated stimulus can also affect how long gaze remains held on that item (Blakely, Wright, Dehili, Boot & Brockmole, 2012), reflecting an influence on attentional disengagement. Itti and Koch (2000) propose that the visual cortex constructs a ‘saliency map’ representing the structural features of the visual environment based on information from sensory pathways such as luminance, contrast, orientation, colour, texture and spatio-temporal frequency. Regions of the map that represent highly salient features will tend to capture, attract or hold overt visual attention. Support for this idea comes from studies in which the location and duration of eye movements during scene viewing were correlated with the structure of computationally modelled saliency maps (Itti, 2006; Itti & Koch, 2000; Parkhurst, Law & Niebur, 2002).

Secondly, selection can be influenced by higher level factors originating from within an individual’s neuropsychological system such as memory, goals and expectations. For example, early experiments measuring eye movements showed that the location of fixations while viewing a painting varied according to the particular task instructions given to participants (Yarbus, 1967). Further studies showed that the number of fixations and overall time spent looking at objects in scenes varied according to whether participants were instructed to search or memorise the scene (Castelhano, Mack & Henderson, 2009). Studies have also failed to find evidence that models of visual saliency predict eye movement behaviour during scene viewing and that such behaviour is better modelled by taking top-down influences into account (Henderson, Brockmole, Castelhano & Mack, 2007; Torralba, Oliva, Castelhano & Henderson, 2006). Object memory can also influence attentional selection. Chanon and Hopfinger (2008) found that the time spent looking at objects during memorisation of a natural scene was greater when the same objects had been part of a previous encoding session. However, experimental context may be important for the effects of item memory on attentional allocation. Johnston, Hawley and Farnham (1993) showed that novel words capture attention when presented together with previously encoded words, whereas Christie and Klien (1995) found that familiar items had a greater influence on attention than unfamiliar items during a motion detection task.

Thus, previous studies consistently show that eye movements can act as an index of ongoing visual and cognitive processing priority and offer a more direct measure of attentional allocation than more traditional measures such as reaction time (Duc, Bays & Husain, 2008). Therefore, by measuring the location and duration of eye fixations in a precognition task, it may be possible to evaluate whether a target stimulus has been prioritised for more extensive processing at a given point in time. As far as the current author is aware, only one previous study has measured eye movements in a precognition task (Radin & Borges, 2009). However, that particular study used horizontal eye movements as an indicator of emotional processing. The current set of studies is therefore the first to use eye movements to test for precognitive effects in the visual attention system.

Building on the methodology from the sequential presentation paradigm described above, an experimental paradigm was developed that would allow for a test of precognitive effects in the visual attention system using eye movements as a behavioural measure. Like the sequential presentation paradigm, the current paradigm involved the initial presentation of target and decoy stimuli followed by the repeated presentation of the target as feedback. However, there were a number of differences. Firstly, target and decoy stimuli were presented simultaneously in an array, rather than in a sequence, in order to allow the visual attention system to express selective responses to each stimulus via eye movements. Secondly, rather than imposing forced choice task conditions (as was the case with the sequential presentation paradigm), the current experiments required participants to memorise the stimulus array in preparation for a recognition test without informing them that the experiment was an investigation of precognition. It was thought that presenting the task as a 'guessing game' might introduce response biases that may reduce the experiment's ability to detect potential precognitive effects. For example, analysis of forced choice ESP experiments using Zener cards has shown that participants display a tendency to guess each symbol an equal number of times over the course of their session and tend to avoid guessing a symbol twice in a row (Stanford, 1975). Since random sampling will most certainly generate unsymmetrical target distributions and target repetitions in forced choice experiments, this form of response bias would serve to reduce the signal to noise ratio. On the other hand, it was thought that a memorisation task might reduce these particular sources of response bias since guessing behaviour would be incidental rather than an explicit requirement.

The general design of the current paradigm is shown in figure 3. Each trial consists of a *study* and *test* phase. In the study phase, four stimuli are presented in an array while eye movements are recorded. During this time, participants are asked to attend to and memorise each stimulus in preparation for a yes/no recognition test. After a brief retention interval, the trial is randomly determined to proceed according to one of two conditions. In the *old* condition, the target stimulus is presented as a probe, whereas in the *new* condition the probe is a previously unobserved stimulus. The yes/no recognition test therefore functions as a means to manipulate the repeated presentation of the target stimulus.

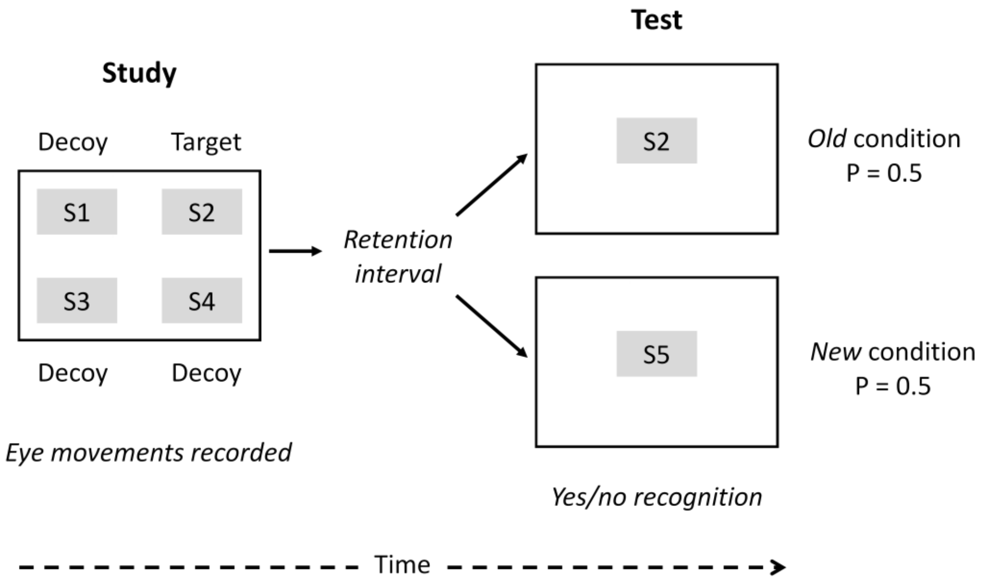


Figure 3. The current visual attention paradigm: In the study phase, participants are required to memorise an array of stimuli, S1-S4, while their eye movements are recorded. After a brief retention interval, a probe is presented for a yes/no recognition test. The target (in this case S2) is presented as a probe in old trials while a previously unobserved stimulus (S5) is presented in new trials. Old and new trials are presented in a random sequence.

Description of the general hypothesis

In the current thesis, precognition is defined as an anomalous correlation between present cognitive activity and a future event. Accordingly, the *precognitive hypothesis* proposes that the degree of attention allocated to target stimuli during the study phase will correlate with the repeated presentation of the target stimulus in the future. Since old trials present target stimuli in the future but new trials do not, there will be a difference in attentional allocation to targets when old and new trials are compared. In contrast, the null hypothesis predicts that there will be no difference. No assumptions are made about the type of information processing underlying a selective response to targets in the old condition. Therefore, the precognitive hypothesis is non-directional; attention may be directed towards or away from targets.

Dependent variables in the current paradigm

All of the experiments in the present thesis used eye movement variables to probe the allocation of attention during the study period. The duration and location of fixations are widely acknowledged to reflect processes involved in gaze control (Henderson, 2003). Two variables, *Total Dwell Time* (TDT) and *First Fixation Time* (FFT), were used and were assumed to reflect the degree to which the experimental stimuli were able to hold and attract attention respectively. A number of processes are likely to influence these measures such as those involved in sensory analysis, object recognition, encoding, maintenance, semantic analysis and potential expectations about target status. Each variable was defined as follows:

Total Dwell Time

TDT was defined as the total time spent fixating a given stimulus, measured over the entire study period. This was calculated by summing the duration of all individual fixations falling on the relevant stimulus. Saccades were not included in this measure, even if the saccadic movement traversed the relevant stimulus. TDT is likely to indicate the degree to which a given stimulus holds attention.

First Fixation Time

FFT was defined as the time taken to fixate the relevant stimulus for the first time, measured from trial onset to the beginning of the first fixation on the relevant stimulus. FFT is likely to indicate the degree to which a given stimulus attracts attention from the peripheral field.

In total, a series of five experiments were carried out using the general paradigm described above, each measuring TDT and FFT associated with each stimulus in the array. Experiment 1 was an initial test of the precognitive hypothesis using this novel experimental design.

Experiment 1: An initial test of the paradigm

Introduction

Experiment 1 was an initial test of the current paradigm that was designed to test for precognitive effects in the visual attention system using an experimental design adapted from past studies. Participants were presented with an array of real-world objects and were required to attend to and memorise each object in preparation for a recognition test. After memorisation, two conditions were randomly determined; in the old condition, the target object was presented again as a probe in the recognition test, whereas a novel object was presented in the new condition. Participants were required to indicate whether they had viewed the object before. It was hypothesised that the degree of attention allocated to the initial presentation of the target object, as indexed by eye movements, would be different between the two conditions. The paradigm was a non-intentional test of precognition since participants were unaware that the experiment was investigating anomalous anticipatory effects.

Methods

Participants

50 student participants (41 female, 9 male) were recruited by convenience from the psychology department subject pool resource at the University of Edinburgh. The average age of participants was 18.9 years (range, 17-31). All participants had normal or corrected to normal vision and were either native or non-native English speakers. Course credit was awarded for participation. Participants were required to sign a consent form as acknowledgement of the experimental aims, risks and data anonymity. The consent form included the right to terminate the session at any time without reason. The experiment met the British Psychological Society's ethical guidelines and was approved by the psychology department's ethical approval panel.

Stimuli

General stimulus characteristics

It was decided that images of real world objects and animals would serve as experimental stimuli. Use of such stimuli has several advantages. Firstly, the precognitive hypothesis does not make any assumptions about the kind of processing involved in selective responses to target stimuli. Studying images of real world objects and animals is expected to involve a number of processes such as sensory analysis, object recognition, semantic analysis, categorisation and lexical activation (Ungerleider & Bell, 2011). Therefore, use of these stimuli does not restrict a test of the precognitive hypothesis to selective responses based on one particular level of analysis. In contrast, use of simple geometric shapes restricts selective responses to information based on lower levels of analysis which may be the wrong level at which to assess the effect of interest. Secondly, use of images depicting real-world objects is comparable to a number of contemporary precognition studies of unconscious precognition where stimuli comprise of relatively complex, meaningful images (Bem, 2011; Broughton, 2004; Mcdonough et al., 2002; Moulton & Kosslyn, 2008; Radin, 2004; Radin & Borges, 2009; Storm et al., 2010a; Sartori et al., 2004) and may offer a more useful comparison of results.

Static images were used in the study phase while video clips were used as probe stimuli in the test phase. The use of video clips was based on evidence that videos resulted in a significantly higher hit rate compared to static images when used as target stimuli in a series of ganzfeld experiments (Bem & Honorton, 1994). However, static images were used in the study phase because it was felt that an array of simultaneously presented video clips would create an excessive source of distraction for participants during the memorisation task.

Array stimuli

Array stimuli comprised of static images and were constructed as follows. 132 digitised colour images of real-world objects and animals were obtained from the internet. Irrelevant background was removed from the outer contours of each item using image editing software and all digitised images were manually adjusted to be approximately equal in size (mean of 245 pixels in width and 215 pixels in height). Items were arranged into 33 groups of 4. Members of each group were chosen in an attempt to maximise within-group differences in visual and semantic properties.

Figure 4. Example of an array used in Experiment 1

3 groups of stimuli were used in practice trials while the remaining 30 groups were used in the experimental session. One item in each group was then arbitrarily selected by the author to be the target for that group. Target items in each group were fixed throughout the experiment. The remaining three items in each group served as decoys. 3 practice and 30 experimental arrays were then constructed. Arrays consisted of the four items in each group positioned on a uniform grey background (RGB = 192, 192, 192). In each array, the display

was divided into four equally sized quadrants and items were randomly allocated to one quadrant and positioned in the centre. After this procedure, the position of each item within the array did not vary throughout the experiment. Arrays were displayed at a resolution of 1024 x 768 pixels. An example of an array can be seen in figure 4.

Probe stimuli

Video clips were created to serve as probe stimuli in the test phase of each trial. Clips were taken from the internet and were also created by the author using a digital camera. Each clip featured a single item as the central focus of attention. All clips were 10 seconds in length and were presented at 30 frames per second at a screen resolution of 1024 x 768 pixels. For the experimental session, two clips were grouped with each array corresponding to the *old* and *new* condition; old clips featured the designated target item for the given array while new clips featured a new item. Target items featured in old clips were selected to match the structural features of the corresponding array image as closely as possible. The items featured in new clips were selected in an attempt to maximise visual and semantic differences between the new item and all other items in the array. In total, 30 old and 30 new clips were created for the experimental session and an additional 3 clips were created for use in the practice trials. All participants viewed the same practice trial clips in the same order. Figure 5 shows a still frame taken from an old clip featuring the target item seen in figure 4.

Figure 5. A still frame taken from a video clip in the old condition featuring a target item.

Apparatus

Stimuli were presented on a 19-inch monitor at a refresh rate of 60Hz. Eye movements were recorded using an SR Research EyeLink 1000 table mounted eye tracker, sampling at 1000Hz and accurate to within 0.01 degrees of visual angle. Participants viewed stimuli at a viewing distance of approximately 90cm with binocular vision. However, only the right eye was tracked as is standard procedure in the eye tracking literature. The experiment was implemented using Experiment Builder (SR Research) and raw data was captured using Data Viewer (SR Research). Raw data was then automatically transferred from Data Viewer to Microsoft Excel 2007 for further processing.

Procedure

Trial procedure

Participants answered advertisements stating that the experiment was investigating visual memory. Upon entering the experimental room, participants were sat comfortably at the eye tracker and viewed the following instructions on the monitor:

> You are about to participate in an experiment investigating visual memory. In each trial, you will be shown a screen containing four everyday objects. This will be followed by a short video clip featuring one object. Your task is to decide whether the object featured in the video clip was present in the previous screen. During the video clip, press the green button for 'yes' and red button for 'no'. Try to respond as quickly and as accurately as possible.

A calibration procedure was then performed requiring participants to fixate nine points arranged in a grid on the display which was then repeated for validation. Three practice trials were then presented. After the practice trials, participants were reminded of the instructions and eye position was re-calibrated.

Experimental trials then proceeded according to the following steps which are summarised in figure 6. A drift correction screen presented a fixation circle at the centre of the display. When the participant fixated the centre of this marker, the experimenter initiated the trial. An array was then presented for 4 seconds while eye movements were recorded. Each array was presented only once and in random order. The computer then displayed a blank screen for 1 second, after which a pseudo-RNG (using the Mersenne Twister algorithm as a core generator) randomly determined whether the trial was in the old or new condition with equal probability. If the trial was in the old condition, a video clip was presented featuring the target object for that trial. If the trial was in the new condition, a video clip was presented featuring a new object not present in the previous array. In the new condition, a given array was always followed by the same video clip assigned to that array. Each video clip played for 10 seconds and participants could respond ‘yes’ or ‘no’ with the use of a hand held controller at any time during the clip, although they were encouraged to respond as quickly and accurately as possible. Once the participant made their response, the video clip continued to play for the entire 10 seconds ensuring that all participants viewed the clip for the same amount of time. During each trial, the experimenter faced a monitor with his back to the eye tracker display. The experimenter’s monitor displayed information about eye position and calibration but did not display any information about which particular array was being viewed by the participant during the experiment. All participants received 3 practice and 30 experimental trials. Sessions lasted approximately 30 minutes including calibration procedures.

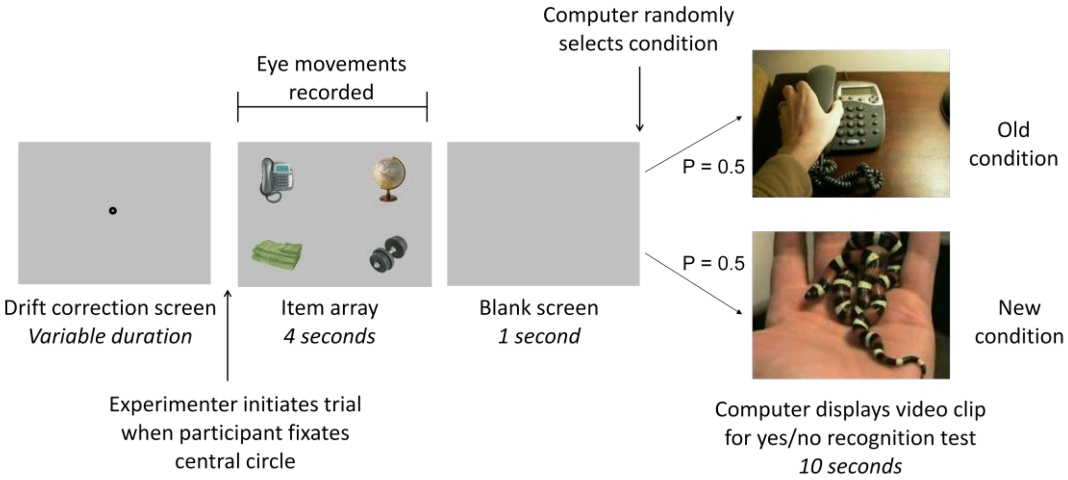


Figure 6. Schematic showing the basic experiment procedure of a single trial in Experiment 1. A drift correction screen was first presented. After the experimenter initiated the trial, an item array was then presented for 4 seconds while eye movements were recorded. This was followed by a blank screen for 1 second. The computer then randomly determined whether the trial proceeded in the old or new condition. In the old condition, a video clip was presented that featured the designated target item for that trial (in this case, a telephone). In the new condition, a video clip was presented featuring an item not encountered in the previous array. Old and new trials were determined with equal probability.

Detection of fixations and saccades

Experiment Builder allows for an *interest area* (IA) to be drawn around stimuli so that eye movements detected within the boundaries of an IA are automatically assigned to the appropriate stimulus. An IA was assigned to each item in a given array and was constructed by manually drawing a boundary around the item. On average, IA boundaries were approximately 5cm from the outer contours of each item which ensured that fixations that did not land directly on the item, but were nevertheless directed towards it, were captured by the interest area. Therefore, IAs varied in size and shape between items but remained fixed for a given item. All fixations and saccades that landed within an IA were automatically assigned to the appropriate item for later analysis. Although it is possible to define IAs after data has been collected, all IAs were constructed before data collection and remained fixed in size and position throughout the experiment. IAs were invisible to the participant during the experiment.

Figure 7. An example of an interest area drawn around an array item to capture eye movement data.

Randomisation

Emphasis has typically been placed on the importance of randomisation in experiments testing for precognition. This is because a test of the null hypothesis requires that the design controls for all means of inferring or deducing future events. Randomisation with replacement means that a particular outcome of a randomisation event has a constant probability throughout the experiment which ensures that outcomes cannot be predicted beyond what would be expected by chance. For that reason, randomisation with replacement is often used in precognition studies. One possible disadvantage of this method is that one is left with an unequal distribution of outcomes since their overall frequency is not constrained. This may be cause for concern if the randomisation procedure happens to produce a disproportionate frequency of one particular outcome by chance, to the extent that it may affect response characteristics and introduce noise into the data.

On the other hand, an experimenter may choose to constrain the overall frequency of outcomes with a balanced design and thus randomise without replacement. However, randomisation without replacement means that the probability of observing a particular outcome will fluctuate throughout an experimental session and will depend on the outcome of previous randomisation events. For example, let us imagine a 4-trial experiment with two possible conditions, A and B. Let us also suppose that the design required the same number of trials in the A and B condition at the end of the experiment (i.e., 2 of each). The probability of observing condition A or B on trial 1 would be 1/2. However, if trial 1 resulted in condition A, the probability of observing condition A on trial 2 is now 1/3. Theoretically, this may allow for a correlation between participants' behaviour and the fluctuations in outcome probability that result from this method, compromising a test of the null.

It was therefore decided that Experiment 1 use randomisation with replacement to determine the trial condition. Thus, the probability of observing an old or new trial was held constant throughout the experiment. Although the randomisation was implemented by a pseudo-RNG which will contain a small degree of bias, the Mersenne Twister core generator passes many stringent test of randomness and its output can therefore safely be regarded as unpredictable in the current context (Matsumoto & Nishimura, 1998).

Hypotheses

In Experiment 1, the identity of target stimuli was fixed throughout the experiment. Therefore, the precognitive hypotheses were as follows:

1) The mean TDT for target items will be different when old and new trials are compared, reflecting a difference in the degree to which target items hold attention in each condition.

2) The mean FFT for target items will be different when old and new trials are compared, reflecting a difference in the degree to which target items attract attention in each condition.

According to the null hypothesis, there will be no such differences given that the random selection of old and new conditions is not expected to correlate with behaviour during study. Since the type of information underlying the hypothesised selective response to targets in old trials is not specified, the hypotheses are bidirectional and two-tailed tests are used.

Planned Analyses

Two analyses are reported for each test of the precognitive hypothesis; one with participants as the unit of analysis (n = 47) and one with items as the unit of analysis (n = 30). Tests were two-tailed unless specified otherwise. In addition, a Bonferroni correction for two comparisons (TDT and FFT) was applied to each test of the precognitive hypothesis. Effect sizes are reported as Cohen's *d*. Data trimming procedures were applied post-hoc to assess the influence of outliers.

Results

Data Exclusions

Data from 3 participants was excluded before statistical analysis due to a poor eye tracking signal, leaving data from 47 participants (38 female, 9 male) with an average age of 18.9 years (range 17 – 31). In addition, data from trials in which participants responded either incorrectly or gave no response were excluded from further analysis. This was to ensure that only those trials in which participants paid attention to the task were included.

Reaction Times and Response Accuracy

The mean reaction time in the recognition task was 1298ms in the old condition and 2466ms in the new condition, which was found to be significantly different, $t(46) = -8.36$, $p < 0.0001$. This is likely to reflect differences in the time taken to attend to items in old and new video clips due to differences in their dynamic properties. Overall accuracy was high with 97.0% and 96.6% correct responses to probe stimuli in old and new trials respectively. This supports the validity of the experimental manipulation since items in old video clips were almost always identified as being old and vice versa for items in new video clips.

General viewing behaviour

All stimuli

Table 1 shows the relationship between array position and the sequence in which items were fixated for the first time. In general, items occupying the top left position of the array (position 1) were fixated first, followed by the top right (position 2), bottom right (position 4) and then bottom left (position 3). A similar relationship can be seen when FFT is examined for each position in the array, collapsed across conditions (see figure 8). Items in position 1 are fixated the earliest, followed by position 2, while items in position 3 and 4 are fixated last. A one way ANOVA on FFT with array position as a factor revealed a highly significant effect of position, $F(3,46) = 133.6$, $p < 0.0001$. This suggests that participants adopted an approximately clockwise viewing behaviour with a strong tendency to fixate the top left item first. This may reflect behaviour that occurs during reading where, in the English language, the top left word in a passage of text tends to be read first. However, items were sometimes re-examined multiple times after other items in the array had been fixated. Figure 8 also shows that mean TDT decreased as a function of array position with items in position 1 holding the greatest degree of attention while items in position 3 and 4 held the least degree of attention. This decrease was significant as revealed by a one way ANOVA on TDT with array position as a factor, $F(3,46) = 38.99$, $p < 0.0001$, and was perhaps due the approximately clockwise pattern of viewing behaviour adopted by the majority of participants during study. In other words, extensive viewing of late-fixated items occupying position 3 and 4 may have been interrupted by the termination of the study array.

Targets vs. decoys

Overall, the TDT for decoys (M = 797.9, SD = 38.1) was greater than for targets (M = 727.4, SD = 65.5), $t(46) = 6.1$, $p < 0.0001$, $d = 1.32$, and the FFT to targets (M = 1406.3, SD = 185.5) was greater than for decoys (M = 1357.2, SD = 143.4), $t(46) = 3.07$, $p < 0.01$, $d = 0.30$ suggesting that decoys were fixated earlier and for longer.[7] Since targets and decoys were arbitrarily selected before the experiment and fixed in position, this result is likely to reflect intrinsic differences in their visual and semantic properties and their average position in the array.

		Sequence			
		1st	**2nd**	**3rd**	**4th**
Array Position	**1**	81.1	12.9	2.6	3.4
	2	11.4	67.3	8.5	12.7
	3	4.8	15.2	21.4	58.6
	4	2.7	4.5	67.5	25.2

1	2
3	4

Array position

Table 1. The percentage of items that were fixated 1st, 2nd, 3rd and 4th as a function of position in the array. Array positions were 1 (top left), 2 (top right), 3(bottom left) and 4 (bottom right).

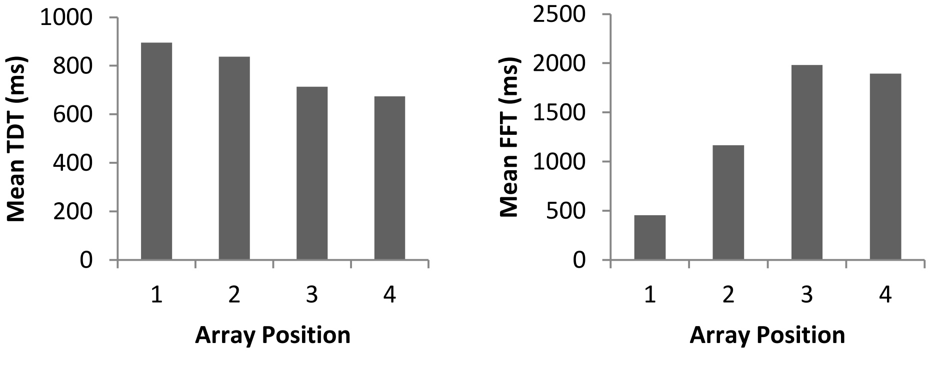


Figure 8. Mean TDT (ms) and mean FFT (ms) to all items as a function of position in the array in Experiment 1.

---

[7] The smaller standard deviations for decoys are a consequence of averaging across three stimulus items per cell. This applies to all experiments reported in the current thesis.

Randomisation check

The pseudo-RNG generated 802 trials in the old condition and 608 trials in the new condition. Since the ratio of old to new trials should have been approximately 1:1, this difference alone represented somewhat of an anomaly. Therefore, the current author presented this data to SR Research (designers of the eye tracking software used in the current thesis) and it was discovered through further correspondence that the discrepancy in output could be traced to an unexpected bug in the Experiment Builder eye tracking software. The difference in frequency between old and new trials was tested with a chi-squared goodness of fit test and was found to be highly significant, X^2 (1, 1410) = 26.4, $p < 0.0001$. It was therefore concluded that the sequence of trial conditions was unlikely to have been adequately random and may have contained a significant degree of underlying structure. Even though visual inspection of the trial sequence for each participant did not reveal any obvious repeating pattern, the potential presence of bias in the sequence of trial conditions limits conclusions that can be drawn from tests of the precognitive hypothesis.

Test of Hypothesis 1

Table 1 shows the mean TDT for target and decoy items in the old and new conditions. The results show that target items were fixated approximately 44 ms longer in the old condition compared to the new condition. Kolmogorov-Smirnov goodness of fit tests suggested that TDT data for old and new targets was normally distributed (participants, old, $D(47) = 0.128$, $p > 0.2$; new, $D(47) = 0.116$, $p > 0.2$, items, old, $D(30) = 0.124$, $p > 0.2$; new, $D(30) = 0.127$, $p > 0.2$), therefore parametric tests were used in the planned analysis. A paired t-test confirmed that this difference was statistically significant; participants, $t(46) = 2.41$, $p = 0.02$ (with Bonferroni correction), $d = 0.51$; items, $t(29) = 2.10$, $p = 0.04$ (with Bonferroni correction), $d = 0.32$. In addition, the TDT for decoys was greater in new trials compared to old trials suggesting that the increase in target TDT for old trials was not the result of a general increase in dwell time over all items in the old condition, but rather, a selective response to targets. In order to assess whether this result was due to the influence of outliers, a 10% trim was applied to the raw data. 5% of data points with the highest TDT values and 5% of data points with the lowest TDT values were removed. This removed all TDT values above 1403ms and less than 299ms. After this trimming procedure, the data showed that target items were fixated approximately 30ms longer in the old condition compared to the new condition. However, only the analysis by participants was significant with the trimmed data; participants, *t*(46) = 2.47, *p* = 0.02 (with Bonferroni correction), *d* = 0.42; items, *t*(29) = 1.47, *p* = 0.15 (with Bonferroni correction), *d* = 0.29. Therefore, trimming the data appeared to somewhat reduce the effect which may suggest that outliers were influencing the results from the untrimmed data. The TDT results can only be regarded as suggestive evidence in favour of hypothesis 1 for this reason. However, it should also be noted that the sequence of trial conditions was biased which brings this result into further doubt.

	Old	New
Target	748.7 (80.2)	703.7 (95.8)
Decoy	792.0 (44.1)	804.8 (45.4)

Table 1. Mean TDT (ms) for targets and decoys in old and new trials. Standard deviations are in parentheses. Values shown are for participant means from untrimmed data.

Test of Hypothesis 2

Table 2 shows the mean FFT for target and decoy items in the old and new conditions. Kolmogorov-Smirnov goodness of fit tests suggested that FFT data for old and new targets was normally distributed (participants, old, *D*(47) = 0.106, *p* > 0.2; new, *D*(47) = 0.125, *p* > 0.2, items, old, *D*(30) = 0.118, *p* > 0.2; new, *D*(30) = 0.112, *p* > 0.2), therefore parametric tests were used in the planned analysis. The mean FFT for targets did not significantly differ between the old and new condition; participants, *t*(46) = 0.16, *p* = 0.87, *d* = 0.03; items, *t*(29) = 0.88, *p* = 0.39, *d* = 0.01. Applying a 10% trim to the raw data did not affect this result; participants, *t*(46) = 0.05, *p* = 0.96, *d* = 0.01; items, *t*(29) = 0.72, *p* = 0.48, *d* = 0.09. Therefore hypothesis 2 was not confirmed.

	Old	New
Target	1414.5 (246.0)	1406.2 (266.1)
Decoy	1343.3 (156.6)	1377.4 (170.9)

Table 2. Mean FFT (ms) to targets and decoys in old and new trials. Standard deviations are in parentheses. Values shown are for participant means from untrimmed data

Discussion

Experiment 1 was an initial test of the novel experimental paradigm developed in the current thesis that was designed to examine the hypothesis that visual attention can selectively respond to target stimuli in anticipation of their repeated presentation in the future. As far as the current author is aware, this is the first time that eye movements have been used to directly probe visual attention for the presence of precognitive effects. One previous experiment (Radin & Borges, 2009) used eye tracking to test for horizontal eye movements preceding presentation of emotional and calm pictures. However, that study was not designed to directly probe the activity of visual attention as was done here. Therefore, the current paradigm contributes to the existing literature in this regard.

The results of Experiment 1 showed that although there was no difference in first fixation time when old and new trials were compared, participants spent more time fixating target items in the old condition compared to the new condition. This appeared to support hypothesis 1 and suggests that target items were more effective at holding attention in old trials. In contrast, the results for FFT were non-significant suggesting that even though target items held attention for a longer period of time, they did not attract attention from the periphery any more in the old condition. This null result may have been due to the strong tendency for participants to fixate items in a specific order. For example, approximately 80% of items in the top left of the array were the first to be fixated which may have significantly lessened the paradigm’s ability to detect a precognitive effect in the FFT data.

There are several potential explanations for the significant results observed in the TDT comparison. Firstly, one may argue that participants obtained information about targets by normal sensory channels, termed sensory leakage. Parapsychology experiments that aim to test for contemporaneous psi effects such as telepathy are particularly prone to criticism based on sensory leakage because the target stimulus and response co-occur in time. However, in precognitive designs, issues of sensory leakage are nullified since the critical manipulation occurs after responses have been recorded. Here, probe stimuli were presented after eye movements were recorded; therefore, there is no possibility that participants could have acquired sensory information about the trial condition. In addition, since the probe video clip was randomly selected by the computer at least one second after the recording period had ended, there was no possibility of participants picking up on subtle auditory cues as a result of the processes involved in accessing the video clip from the computer’s hard drive. However, in Experiment 1 the experimenter knew the identity of the fixed target items and was present in the experimental room for purposes of monitoring the adequacy of eye movement calibration during each session. Thus, it is conceivable that the experimenter could have alerted the participant to the identity of the target item by some unconscious manner. However, this could not have provided the participant with any information predictive of the trial condition because the experimenter was blind to this aspect of the experiment.

A second and more likely explanation is based on the fact that the output of the pseudo-RNG was biased and produced many more old trials than would be expected by chance. It is therefore possible that the sequence of old and new trials presented to each participant contained an underlying structure, a possibility that inevitably brings confirmation of hypothesis 1 into doubt. If it is assumed that a systematic bias in the sequence of old and new trials was present, then participants may have learned and responded to this underlying structure as each experimental session progressed. Indeed, Brugger and Taylor (2003) propose that positive results from forced choice ESP tests, in which participants are often given long runs of trials with feedback, may be explained by implicit sequence learning (Destrebecqz & Cleeremans, 2001). Furthermore, they suggest that implicit sequence learning may, in principle, account for the results of *any* forced choice ESP experiment that involves feedback (and, presumably, a sufficient number of trials) because all methods of randomisation, including those of ‘true’ RNGs, will contain a small degree of bias. This argument may have merit when applied to early forced choice ESP studies in which randomisation was based on relatively crude means such as manual shuffling of cards (Rhine, 1934). However, modern pseudo-RNGs such as the Mersenne Twister (used here) pass stringent tests of randomness (L’Ecuyer & Simard, 2007) and the number of randomisation events before this core generator repeats its output is well in excess of the number of trials administered to individual participants, which would seem to make learning of any underlying structure implausible. Nevertheless, since the actual data from Experiment 1 suggests that the sequence of trial conditions was biased, the implicit sequence learning hypothesis can at least be tested here, albeit in a post-hoc manner. If participants were indeed learning to respond to a bias in the sequence of old and new trials, we would expect the TDT effect to increase over the course of individual sessions. As can be seen from figure 9, a plot of this relationship shows that, contrary to the implicit learning hypothesis, there is a decrease in the TDT effect over time rather than an increase. However, the trend is not significant (r = -0.07, n = 30). It should be noted that since the plot is not cumulative, the decreasing trend is not a consequence of regression to the mean.

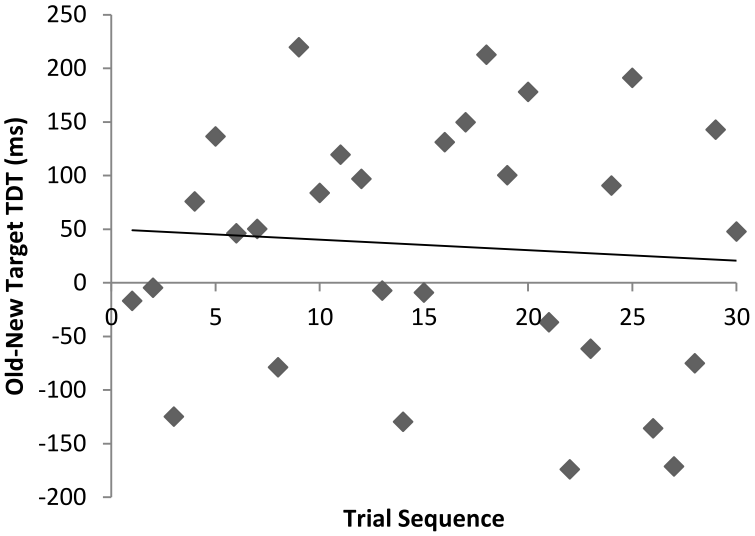


Figure 9. The difference in target TDT between old and new trials as a function of trial sequence.

Although this analysis does not support the implicit sequence learning hypothesis, it cannot be ruled out as an explanation since a more complex pattern of learning may exist that is not detectable by a simple linear correlation. However, confirmation of hypothesis 1 was not the result of a general increase in TDT over all items in the array. Rather, TDT increased for targets and decreased for decoys when old and new trials were compared. Therefore, for participants to have produced this result by implicit sequence learning, they must have been able to reliably predict both the trial condition *and* the position or identity of the target over the course of the session. Target items were arbitrarily selected by the current author before the experiment began rather than being selected at random during the experiment, therefore it is possible (but perhaps not likely) that participants acquired implicit knowledge of target identity through shared preference with the current author and were able, in conjunction with sequence learning, to bias their eye movements accordingly.

Because of the difficulty in interpreting the positive results of Experiment 1 due to the reasons described above, Experiment 2 was carried out as an exact replication of Experiment 1 but with adequate randomisation. If the results of Experiment 1 were to replicate with proper randomisation, explanations in terms of implicit sequence learning would become less likely.

Experiment 2: A replication attempt

Introduction

Experiment 1 obtained a significant difference in target TDT between the old and new condition. In this respect, the results were promising. However, the biased output of the process used to randomise the sequence of trial conditions meant that interpretation of this finding was problematic. The data was examined for the presence of implicit learning effects that might have arisen from responses to subtle patterns in the sequence of trials in each session. Although the data did not directly support the sequence learning hypothesis, caution should nonetheless be exercised when ruling out an explanation of the results of Experiment 1 in terms of statistical learning. In light of this methodological issue, an exact replication of Experiment 1 was carried out in Experiment 2 with proper randomisation in order to see whether the effect would replicate under these conditions. In other respects, Experiment 2 was identical to that of Experiment 1 except for the recruitment of a new set of participants who had not taken part in the previous experiment.

Methods

Stimuli, apparatus, hypotheses and planned analyses were identical to Experiment 1.

Participants

60 student participants (43 female, 17 male) were recruited by convenience from the psychology department subject pool resource at the University of Edinburgh. The average age of participants was 18.9 years (range, 18-24). All participants had normal or corrected to

normal vision and were either native or non-native English speakers. Course credit was awarded for participation. Participants were required to sign a consent form as acknowledgement of the experimental aims, risks and data anonymity. The consent form included the right to terminate the session at any time without reason. The experiment met the British Psychological Society's ethical guidelines and was approved by the psychology department's ethical approval panel.

Procedure

The procedure was the same as used in Experiment 1 except for a correction to the randomisation process. The software bug present in the SR Research Experiment Builder software, identified from Experiment 1, was caused by use of the Python function *random.randint(a,b)* within a particular Experiment Builder graphical interface function. A correspondence from SR Research recommended that an alternative method be used by which the same Python randomisation function could be used without causing a biased output. This new procedure was implemented in Experiment 2 which did not affect any other design feature. Preliminary checks on the randomisation output confirmed that it was operating correctly.

Results

Data Exclusions

Eye tracking signals were adequate for all participants, therefore no individual participant data was excluded. Data from trials in which participants responded either incorrectly or gave no response were excluded from further analysis. This was to ensure that only those trials in which participants paid attention to the task were included.

Reaction Times and Response Accuracy

The mean reaction time in the recognition task was 1227ms in the old condition and 2384ms in the new condition. Overall accuracy was high with 96.3% and 98.6% correct responses to probe stimuli in old and new trials respectively. This supports the validity of the experimental manipulation since probe items in old video clips were almost always identified as being old and vice versa for items in new video clips.

General viewing behaviour

All stimuli

Table 3 shows the relationship between array position and the sequence in which items were fixated for the first time. The data shows that, on average, items in position 1 were fixated first, followed by position 2, position 4 and finally position 3. This trend is reflected in the FFT data collapsed across all items and conditions (see figure 10) which shows that items in position 1 were fixated the earliest, followed by position 2, while items in position 3 and 4 were fixated last. A one way ANOVA on FFT with array position as a factor revealed a highly significant effect of position, $F(3,59) = 158.9$, $p < 0.0001$. This pattern of results closely mirrors that of Experiment 1 and shows that participants again tended to adopt a clockwise pattern of viewing behaviour. In addition, the TDT decreases as a function of array position with items in position 1 having the longest dwell time in a similar manner to Experiment 1.

Targets vs. decoys

Overall, the TDT for decoys (M = 774.1, SD = 52.0) was greater than for targets (M = 717.6, SD = 70.4), $t(59) = 6.42$, $p < 0.0001$, $d = 0.91$, and the FFT to targets (M = 1374.3, SD = 153.2) was greater than for decoys (M = 1325.0, SD = 162.2), $t(59) = 3.14$, $p < 0.01$, $d = 0.31$ suggesting that decoys were fixated earlier and for longer. Again, this data is comparable to that of Experiment 1.

		Sequence			
		1st	**2nd**	**3rd**	**4th**
Array Position	**1**	82.7	9.6	3.1	4.5
	2	10.5	68.8	7.6	13.0
	3	4.8	15.2	24.5	55.4
	4	1.9	6.3	64.8	27.0

1	2
3	4

Array position

Table 3. The percentage of items that were fixated 1st, 2nd, 3rd and 4th as a function of position in the array. Array positions were 1 (top left), 2 (top right), 3(bottom left), 4 (bottom right).

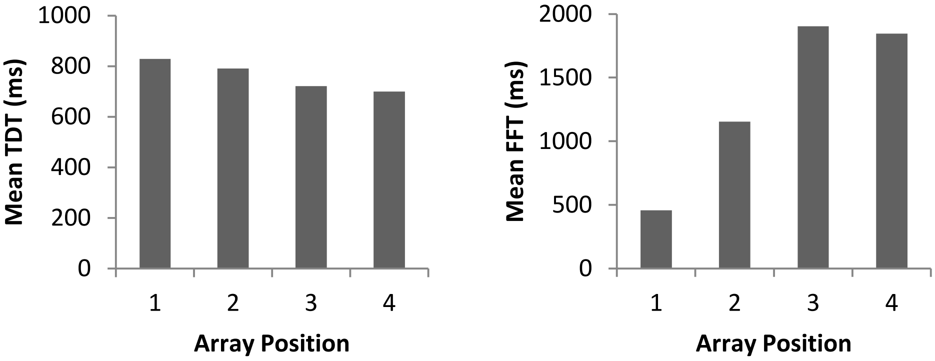


Figure 10. Mean TDT (ms) and mean FFT (ms) to all items as a function of position in the array in Experiment 2.

Randomisation check

The pseudo-RNG generated 864 trials in the old condition and 936 trials in the new condition. A chi-squared goodness of fit test confirmed that the distribution of old and new trials was not significantly different, X^2 (1, 1800) = 2.8, p = 0.09.

Test of Hypothesis 1

Table 4 shows the mean TDT for target and decoy items in the old and new conditions. Kolmogorov-Smirnov goodness of fit tests suggested that TDT data for old and new targets was normally distributed (participants, old, $D(60) = 0.081$, $p > 0.2$; new, $D(60) = 0.121$, $p > 0.2$, items, old, $D(30) = 0.095$, $p > 0.2$; new, $D(30) = 0.070$, $p > 0.2$), therefore parametric tests were used in the planned analysis. The results show that the TDT for targets was less in old compared to new trials but this difference was not statistically significant; participants, $t(59) = -0.77$, $p = 0.44$, $d = -0.12$; items, $t(29) = -0.26$, $p = 0.80$, $d = -0.03$. A 10% trim on the raw data did not affect this result; participants, $t(59) = -1.14$, $p = 0.26$, $d = -0.26$; items, $t(29) = -1.77$, $p = 0.09$, $d = -0.21$. Therefore, hypothesis 1 was not confirmed.

	Old	New
Target	711.2 (93.1)	722.1 (87.7)
Decoy	775.5 (60.7)	773.3 (55.1)

Table 4. Mean TDT (ms) for targets and decoys in old and new trials. Standard deviations are in parentheses. Values shown are for participant means from untrimmed data.

Test of Hypothesis 2

Table 5 shows the mean FFT for target and decoy items in the old and new conditions. Kolmogorov-Smirnov goodness of fit tests suggested that FFT data for old and new targets was normally distributed (participants, old, $D(60) = 0.100$, $p > 0.2$; new, $D(60) = 0.091$, $p > 0.2$, items, old, $D(30) = 0.097$, $p > 0.2$; new, $D(30) = 0.103$, $p > 0.2$), therefore parametric tests were used in the planned analysis. The mean FFT for targets did not significantly differ between old and new trials; participants, $t(59) = -0.61$, $p = 0.55$, $d = -0.12$; items, $t(29) = -0.72$, $p = 0.48$, $d = -0.04$. A 10% trim on the raw data did not affect this result; participants, $t(59) = 0.02$, $p = 0.98$, $d = 0.004$; items, $t(29) = -0.09$, $p = 0.93$, $d = -0.004$. Therefore, hypothesis 2 was not confirmed.

	Old	New
Target	1360.9 (287.7)	1386.9 (239.1)
Decoy	1298.6 (160.3)	1353.3 (192.0)

Table 5. Mean FFT (ms) to targets and decoys in old and new trials. Standard deviations are in parentheses. Values shown are for participant means from untrimmed data.

Discussion

Experiment 2 was an attempt to replicate the preliminary findings from Experiment 1 with adequate randomisation. The results provided no support for the precognitive hypothesis; there was no difference in target TDT and FFT when old and new trials were compared. There are a number of potential explanations for this failure to replicate. Firstly, it may simply suggest that the null is true. In Experiment 1, randomisation was affected by a software error and, consequently, many more old trials were produced than expected by chance. The results from Experiment 1 may therefore have been due to the presence of bias in the trial sequence. In contrast, when there was no obvious bias in the random selection of trial conditions in Experiment 2, there was no effect in the data. However, post-hoc analysis of the data from Experiment 1 revealed no clear pattern that would suggest implicit learning of a bias in the sequence of old and new trials and, in fact, the data seemed to indicate that the effect diminished over each session. Alternatively, the results from Experiment 1 may have been due to chance, with the bias in RNG output playing an insignificant role. Alternatively, the null may actually be false and the failure to replicate the results of Experiment 1 indicates that the precognitive effect is unreliable under the current experimental conditions. However, the absence of a theoretical framework for precognition means that there are many design factors that could, in principle, be held responsible for a failure to replicate the initial effect.

One factor that could have potentially contributed to poor replicability is the type of encoding occurring during study. Two main forms of encoding are available to participants during memorisation of real-world objects in the current paradigm; verbal encoding whereby the name of the object is encoded in a phonological store (Baddeley, 1986) and visual encoding whereby the visual features of the object are encoded in a visuo-spatial ‘scratchpad’ (Logie, 1986). Once encoded, each temporary store of information is maintained by a rehearsal mechanism. In Experiments 1 and 2, participants were free to memorise each item using a visual or verbal encoding strategy, or a combination of strategies. It is therefore possible that the apparent precognitive effect observed in Experiment 1 was largely dependent on one or the other of these encoding strategies.

Two further experiments were therefore carried out. In Experiment 3A, video clips were presented as probes in the same way as Experiments 1 and 2. This served as a further attempt to replicate the effect under the same conditions as the previous studies. However, Experiment 3B presented object names as probes in the recognition test. Accordingly, participants would be encouraged to adopt a verbal encoding strategy in order to carry out the recognition task effectively. Several methodological changes were also made to Experiments 3A and 3B. Firstly, target items were randomly selected from all available items in the array which ensured that targets and decoys were sampled from the same population of stimuli. Secondly, the position of items in the array was randomised. These changes to the randomisation scheme allowed for a comparison between target items in the old condition and *all* items in the new condition since the latter were categorically identical according to the precognitive hypothesis.

Experiment 3A and 3B: Improving randomisation and exploring the role of encoding strategies

Introduction

The results of Experiment 1 did not replicate in Experiment 2. This may have been due to a number of factors. Firstly, the null hypothesis may be true and the results of Experiment 1 can be interpreted in terms of inadequate randomisation or chance. Alternatively, the null may be false and the experimental conditions may not have been optimal for the effect to replicate across Experiments 1 and 2. Experiment 3A was designed as a further attempt to replicate the apparent effect observed in Experiment 1 while Experiment 3B was designed to explore whether encouraging participants to use a verbal encoding strategy would result in a more reliable effect. Accordingly, participants in Experiment 3A were presented with video clips in the test phase, as in Experiments 1 and 2. However, participants in Experiment 3B were presented with object names in the test phase, thus encouraging them to verbally encode the study items in order to perform the recognition task effectively. Three additional methodological changes were also implemented; target items were randomly selected from all available items in the array and the position of each item in the array was randomised. Thirdly, the stimulus set used in Experiments 1 and 2 was amended to include a number of new objects in order to make construction of corresponding video clips easier to accomplish.

Methods

Participants

50 student participants were recruited for both Experiments 3A and 3B (3A, 37 female, 13 male; 3B, 34 female, 16 male). The average age of participants was 20.7 years for Experiment 3A (range, 18-31) and 19.9 years for Experiment 3B (range, 18-28). All participants had normal or corrected to normal vision and were either native or non-native English speakers. Course credit was awarded for participation. Participants were required to sign a consent form as acknowledgement of the experimental aims, risks and data anonymity. The consent form included the right to terminate the session at any time without reason. The experiment met the British Psychological Society's ethical guidelines and was approved by the psychology department's ethical approval panel.

Stimuli

Array stimuli

55 of the array items used in Experiment 1 and 2 were used in Experiment 3. In addition, 65 new digitised colour photographs of real-world objects and animals were added to make a total of 120 items used in experimental trials. The new items were introduced to make it easier to construct corresponding target video clips. Images of real-world objects were obtained from internet sources and their background was removed using image manipulation software. Experimental items were then arranged into 30 groups of 4 while attempting to maximise within-group differences in their visual and semantic properties. Arrays were displayed at a resolution of 1024 x 768 pixels. 3 groups of items were used in practice trials while the remaining 30 groups were used in the experimental session. During runtime, the computer randomly positioned each item in one of the four quadrants of the array.

Probe stimuli

Experiment 3A: Video clips featuring each array item and an additional new item were created to make a total of 150 experimental video clips. Five video clips were grouped with each array; four clips featuring array items and one clip featuring a new item not present in the array. The new item videos were created with an effort to maximise visual and semantic differences between the new item and all other items in the array. All video clips were 10 seconds in length and were presented at 30 frames per second at a screen resolution of 1024 x 768 pixels.

Experiment 3B: Arial text at font size 40 was used for all item names in the test phase. Text was presented in the centre of the display for 10 seconds.

Procedure

Trial procedure

Participants answered advertisements stating that the experiment was investigating visual memory. Before each session began, participants were randomly allocated to either Experiment 3A or 3B. Upon entering the experimental room, participants were sat comfortably at the eye tracker and viewed instructions displayed on the monitor. A calibration procedure was then performed requiring participants to fixate nine points arranged in a grid on the display which was then repeated for validation. Three practice trials were then presented. After the practice trials, participants were reminded of the instructions and eye position was re-calibrated.

Experimental trials then proceeded according to the following steps which are summarised in figure 11. All randomisation events were implemented using the Mersenne Twister as a core generator (Matsumoto & Nishimura, 1998). A drift correction screen presented a fixation circle at the centre of the display. When the participant fixated the centre of this marker, the experimenter initiated the trial. The computer then randomly selected the position of each item within the array. The array was then presented for 4 seconds while eye movements were recorded. Each group of items was presented only once and in random order. A blank screen was then displayed for 1 second, after which the computer randomly determined the trial condition and which item from the array would serve as the target for that trial. If the trial was in the old condition, the target was presented as a probe item. If the trial was in the new condition, a new item was presented. Video clips were presented as probes in Experiment 3A and object names were presented as probes in Experiment 3B. In the new condition, a given array was always followed by the same probe item assigned to that array. Each probe item was presented for 10 seconds and participants could respond ‘yes’ or ‘no’ with the use of a hand held controller at any time during the clip, although they were encouraged to respond as quickly and accurately as possible. Once the participant made their response, the probe item remained on the display for the entire 10 seconds ensuring that all participants viewed probe items for the same amount of time. During each trial, the experimenter faced a monitor with his back to the eye tracker display. The experimenter’s monitor displayed information about eye position and calibration but did not display any information about which particular array was being viewed by the participant during the experiment. All participants in Experiment 3A and 3B received 3 practice and 30 experimental trials.

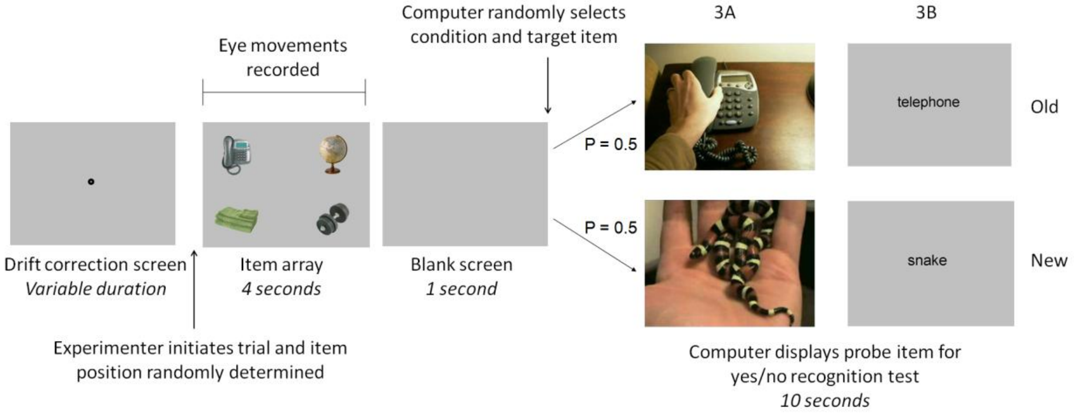


Figure 11. The basic experimental procedure of a single trial in Experiments 3A and 3B. A drift correction screen was first presented. Once the central circle was fixated, the experimenter initiated the trial and the computer randomly determined the position of items within the array. The array was then presented for 4 seconds while eye movements were recorded. After a blank screen, the computer randomly determined the target item and the trial condition. Video clips were presented as probe items in Experiment 3A whereas object names were presented as probe items in Experiment 3B.

Detection of fixations and saccades

Interest areas were drawn in each quadrant of the array and were 400 pixels in width and 300 pixels in height. These dimensions were sufficient to surround all items and ensured that fixations that did not land directly on the item, but were nevertheless directed towards the item, were captured by the interest area. All fixations and saccades that landed within an IA were automatically assigned to the appropriate item for later analysis. All IA’s were constructed before data collection, remained fixed in size and position and were invisible to participants during the experiment.

Hypotheses

In Experiments 3A and 3B, target and decoy items were randomly selected from the same population of items rather than being fixed as in Experiments 1 and 2. Therefore, a test of the precognitive hypothesis could compare targets in the old condition with all items in the new condition since all items are categorically equivalent in the new condition. The precognitive hypotheses for Experiment 3A and 3B were therefore as follows:

1) The mean TDT for target items in the old condition will be different than mean TDT for all items in the new condition, reflecting a difference in the degree to which target items hold attention in the old condition.
2) The mean FFT for target items in the old condition will be different than mean FFT for all items in the new condition, reflecting a difference in the degree to which target items attract attention in the old condition.

Again, since the type of information underlying the hypothesised selective response to targets in the old condition is not specified, the hypotheses are non-directional and two-tailed test are used.

Planned Analyses

All statistical tests were carried out on participants as the unit of analysis. Item analyses were not carried out because the randomisation with replacement procedure resulted in a low number of data points for many items. Tests were two-tailed unless specified otherwise. In addition, a Bonferroni correction for two comparisons (TDT and FFT) was applied to each test of the precognitive hypothesis. Effect sizes are reported as Cohen’s *d*. Data trimming procedures were applied post-hoc to assess the influence of outliers.

Results

Data Exclusions

Data from 1 participant in Experiment 3A was excluded because no response was given in the recognition test on any trial. In addition, data from trials in which participants responded either incorrectly or gave no response were excluded from further analysis. This was to ensure that only those trials in which participants paid attention to the task were included.

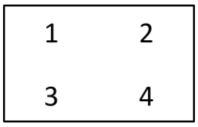
Reaction Times and Response Accuracy

The mean RT in Experiment 3A was 1080ms in the old condition and 1867ms in the new condition, $t(48) = -6.39$, $p < 0.0001$. Mean RT in Experiment 3B was 1090ms in the old condition and 1060ms in the new condition which was not statistically significant, $t(49) = 1.11$, $p = 0.27$. The discrepancy in RT results between Experiment 3A and 3B was most likely due to differences in the time taken for items in the old and new video clips to capture attention in the recognition test. Overall accuracy was high; Experiment 3A resulted in 93.2% and 96.8% correct responses in old and new trials respectively; Experiment 3B resulted in 97.3% and 98.9% correct responses to old and new trials respectively.

General viewing behaviour

The general viewing behaviour in both Experiment 3A and 3B closely mirrored that in previous experiments. Participants tended to fixate the item in the top left corner first and continued in an approximately clockwise fashion. Table 6 shows the relationship between array position and fixation sequence and figure 12 shows mean TDT and mean FFT for all items in each array position for Experiments 3A and 3B.

			Sequence			
			1st	**2nd**	**3rd**	**4th**
Array Position	Expt. 3A	**1**	85.3	10.6	1.7	2.4
		2	5.5	70.7	10.3	13.5
		3	7.4	15.3	17.3	59.9
		4	1.8	3.4	70.7	24.1
	Expt. 3B	**1**	86.9	8.9	2.2	5.3
		2	9.1	68.6	10.0	15.6
		3	5.4	19.6	24.7	53.5
		4	1.9	6.2	66.4	28.8



Array position

Table 6. The percentage of items that were fixated 1st, 2nd, 3rd and 4th as a function of position in the array for Experiment 3A and 3B. Array positions were 1 (top left), 2 (top right), 3(bottom left), 4 (bottom right).

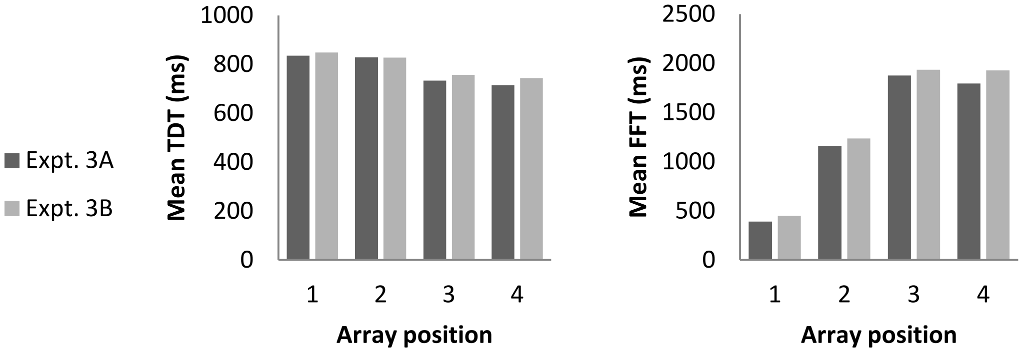


Figure 12. Mean TDT (ms) and mean FFT (ms) to all items as a function of position in the array for Experiments 3A and 3B.

Randomisation check

In Experiment 3A, 745 old trials and 755 new trials were generated. In Experiment 3B, 744 old trials and 756 new trials were generated. Neither frequency deviated significantly from chance expectation, 3A: X^2 (1, 1500) = 0.06, $p = 0.81$; 3B: X^2 (1, 1500) = 0.08, $p = 0.78$.

Test of Hypothesis 1

Table 7 shows the mean TDT for targets, decoys and all items in the old and new condition for Experiments 3A and 3B. Kolmogorov-Smirnov goodness of fit tests suggested that all TDT data was normally distributed (all p > 0.2) therefore parametric tests were used. In Experiment 3A, targets in the old condition were fixated approximately 1.5ms less than items in the new condition and this difference was not significant, $t(48) = -0.12$, $p = 0.90$, $d = -0.02$. In Experiment 3B, targets in the old condition were fixated approximately 11.4ms more than items in the new condition and this difference was also not significant, $t(49) = 0.74$, $p = 0.46$, $d = 0.14$. A 10% trim of the raw TDT data from each experiment did not affect this result; 3A, $t(48) = 0.15$, $p = 0.88$, $d = 0.03$; 3B, $t(49) = -0.06$, $p = 0.95$, $d = -0.01$. Therefore, hypothesis 1 was not confirmed for either experiment.

		Old	New
Expt. 3A	Target	776.7 (93.5)	792.1 (107.4)
	Decoy	778.1 (58.7)	773.6 (52.1)
	All items	777.8 (50.7)	778.2 (46.9)
Expt. 3B	Target	803.2 (110.3)	796.4 (94.0)
	Decoy	792.7 (47.6)	790.3 (42.1)
	All items	795.2 (29.5)	791.8 (34.3)

Table 7. Mean TDT (ms) for targets, decoys and all items in the old and new condition for Experiment 3A and 3B. Standard deviations are in parentheses. Values are from untrimmed data.

Test of Hypothesis 2

Table 8 shows the mean FFT for targets, decoys and all items in the old and new condition for Experiments 3A and 3B. Kolmogorov-Smirnov goodness of fit tests suggested that all FFT data was normally distributed (all p > 0.2) therefore parametric tests were used. In Experiment 3A, targets in the old condition took approximately 17.5ms longer to fixate for the first time than items in the new condition but this difference was not significant, $t(48) = 0.53$, $p = 0.60$, $d = 0.07$. A 10% trim on the raw data did not affect this result, $t(48) = -0.17$, p

= 0.86, *d* = -0.03. Hypothesis 2 was therefore not confirmed for Experiment 3A. In Experiment 3B, targets in the old condition took approximately 86.8ms longer to fixate for the first time than items in the new condition and this difference was significant, *t*(49) = 2.16, *p* = 0.04, *d* = 0.36 (Bonferroni corrected for two comparisons). This result suggests that target items in the old condition did not attract attention as effectively as items in the new condition. In order to assess the influence of potential outliers on the result from Experiment 3B, a 10% trim was applied to the raw data; 5% of data points with the highest FFT values and 5% of data points with the lowest FFT values were removed. This removed all FFT values above 3087ms and less than 227ms. After this trimming procedure, the effect was substantially reduced; the data showed target items in the old condition were only fixated approximately 6.7ms later than items in the new condition. This difference was non-significant, *t*(49) = 0.23, *p* = 0.82, *d* = 0.03 and suggests that the result of the main analysis was likely to have been influenced by outliers. Confirmation of hypothesis 2 for Experiment 3B should therefore be treated with caution.

		Old	New
Expt. 3A	Target	1311.5 (309.8)	1278.2 (256.0)
	Decoy	1300.8 (160.5)	1299.4 (185.6)
	All items	1303.4 (165.2)	1294.0 (166.9)
Expt. 3B	Target	1468.0 (301.1)	1409.2 (334.7)
	Decoy	1364.2 (168.2)	1372.1 (158.7)
	All items	1389.4 (166.2)	1381.3 (162.2)

Table 8. Mean FFT (ms) for targets, decoys and all items in the old and new condition for Experiment 3A and 3B. Standard deviations are in parentheses. Values are from untrimmed data.

Discussion

Experiments 3A and 3B were both designed as further tests of the precognitive hypothesis using the experimental paradigm developed in the current thesis. Experiment 3A was an attempted replication of the apparent effect observed in Experiment 1, again using dynamic probe stimuli, but with more extensive randomisation procedures. Experiment 3B presented participants with object names as probe stimuli in an attempt to explore whether encouraging participants to adopt a verbal encoding strategy would lead to a more reliable effect. However, no compelling evidence for precognition was obtained in either Experiment 3A or 3B. There was a significant result found in Experiment 3B when FFT was examined but trimming the data to assess the influence of outliers drastically reduced the effect and rendered the analysis non-significant. It would therefore be prudent to regard the significant effect found with the untrimmed data as due to a spurious effect of outliers.

The presence of null results can be interpreted in a number of ways. As discussed previously, this could simply indicate that the null is, in fact, true. If we compare the data from experiments 3A and 3B, not only did the results fail to reach significance but the differences between conditions were also in opposite directions, which adds another element of inconsistency. Methodologically, Experiments 3A and 3B were perhaps superior to previous experiments; since targets were randomly sampled and their position was randomly determined, a bias in the placement of target position *and* a bias in the sequence of old and new trials would have to be present to generate a spurious result.

However, a consequence of randomising variables without balancing their overall frequency is that their final distribution is likely to be unsymmetrical. For example, the proportion of targets occupying each position in the array is likely to be uneven in the old condition. The data from each experiment so far has shown that the position an item occupies in the array affects the TDT on that item; for example, items occupying position 1 are fixated approximately 100ms longer than items in positions 3 and 4. In addition, randomly sampling targets with replacement means that the old condition was likely to contain different numbers of each array item serving as the target, with items differing in their inherent visual attractiveness. Therefore, although non-significant deviations in these distributions are not likely to lead to spurious false-positive effects, they may add noise to the data making reliable observation of an underlying precognitive effect more difficult, especially if the extraneous variables have a relatively large effect on eye movement measures compared to a presumably small precognitive effect.

Experiment 4 therefore used a completely balanced design in an attempt to minimise sources of variance due to unsymmetrical distributions of targets and their position in the array. Each stimulus was to be presented the same number of times as a target and decoy, the same number of times in each array position, and the same number of times in each condition. The total number of old and new trials was also balanced across the experiment. Using a completely balanced design does, however, introduce an element of predictability in the sequence of trial conditions and stimulus parameters. For example, if participants receive the same number of trials in the old and new condition, the probability that the last trial will be old or new is either 0 or 1, depending on the particular trial sequence history. Nevertheless, participants must be in a position to know, in advance, the total number of trials and the relative frequency of each condition in order to fully capitalise on this element of certainty. Furthermore, it is possible to examine the data for evidence of sequence learning that may account for any positive results that are found.

Two further methodological changes were introduced in Experiment 4. Firstly, study items were taken from a set of standardised stimuli consisting of colourised drawings of real-world objects and animals (Rossion & Pourtois, 2004). Several norms are available for these stimuli such as ratings of familiarity, visual complexity and naming time. By using these standardised stimuli, the relationship between normative data and a potential precognitive effect could be explored. Secondly, the Rossion and Pourtois (2004) colourised drawings were also used as probes in the test phase in order to explore whether presenting exactly the same visual stimulus at study and test would result in a more reliable effect.

Experiment 4: Identical stimuli at study and test

Introduction

The results of Experiment 3A and 3B failed to support the precognitive hypothesis and thus failed to replicate the apparent effect observed in Experiment 1. Experiment 4 was designed to further explore the potential role of the probe stimulus in the current paradigm. Several methodological changes were introduced. Firstly, standardised stimuli developed by Rossion and Pourtois (2004) were used as study items. The Rossion and Pourtois set is a colourised version of a stimulus set originally developed by Snodgrass and Vanderwart (1980) and consists of pictorial representations of concrete nouns. The original Snodgrass and Vanderwart (1980) images have been used in a number of previous experiments, for example in studies of attention (Dux & Harris, 2007; Pashler & Harris, 2001; Scholl, 2000), memory (Karlsen & Snodgrass, 2004; Kohler, Moscovitch & Melo, 2001), object identification

(Dell'Acqua & Job, 1998; Snodgrass & Corwin, 1988), the Stroop effect (Dishon-Berkovits & Algom, 2000) and priming (Damian, 2000; Mitchell, 2006). In addition, a number of studies have collected normative data for these stimuli (Barry, Morrison & Ellis, 1997; Rossion & Pourtois, 2004; Snodgrass & Vanderwart, 1980). This provided an opportunity to explore whether a precognitive effect, if present, would be systematically related to any particular normative variables which may help to elucidate any potential underlying cognitive or perceptual processes responsible for the observed effect.

Secondly, test items were also taken from the same stimulus set. In other words, the target item and the probe item in the old condition were exactly the same visual stimulus. In Experiments 1, 2 and 3A, probe items were video clips that featured target objects in dynamic contexts. There may have been several disadvantages of using video clips as probes in these experiments. Firstly, the degree of visual correspondence between a given target array item and the object featured in the video clip was not an exact correspondence. This was naturally due to the dynamic nature of the video clip, which presented objects in a variety of changing perspectives. However, items featured in video clips also differed slightly from target array items in other structural ways such as form, surface details and exact dimensions. If the hypothesised precognitive effect is sensitive to the degree of visual similarity between stimuli presented at study and test, then it may be advantageous to present exactly the same visual stimuli in the study and test phase. By way of comparison, some well established psychological effects are sensitive to such considerations. For example, in perceptual priming experiments, the strength of the priming effect can be reduced if the prime does not exactly match the physical characteristics of the stimulus on its repeated presentation. For instance, changing an object's orientation (Srinivas, 1995, 1996) or exemplar (Biederman & Cooper, 1991; Cave, Bost & Cobb, 1996) between study and test can reduce the priming effect while studies of perceptual priming using word stimuli show that the priming effect is reduced by changing typography (Jacoby & Hayman, 1987). This is not to imply that perceptual priming and precognition share any underlying mechanism. Rather, this suggests that such considerations be empirically tested since priming experiments and the current paradigm are, at least, methodologically comparable in that both involve the repeated presentation of stimuli, albeit in reverse order in the present case.

Thirdly, in order to remove potential sources of noise, the trial condition, item position and target identity were completely balanced across the study. In Experiments 3A and 3B, both the position of items in the array and the identity (and therefore the position) of the target item was randomised with replacement during the experiment runtime. As a consequence of this unbalanced randomisation procedure, the old condition was composed of an asymmetrical distribution of target items and their position in the array. Although this is not likely to inflate the rate of false positives, it may have made detection of an underlying precognitive effect more difficult especially if the size of the hypothesised effect is small compared to the variance component introduced via this randomisation procedure. Indeed, the data from all previous experiments showed that the TDT and FFT varied considerably as a function of an item's position in the array. Array position had the strongest effect on FFT, with the difference in mean FFT varying by as much as 1500ms. Array position affected TDT to a lesser extent with a largest difference of approximately 100ms between positions. Another source of noise in previous experiments may have been the unequal distribution of items appearing as targets, again due to randomisation with replacement. Items are expected to vary in their inherent ability to attract and hold attention due to differences in their physical and semantic properties. For example, if the randomisation procedure happened to select a greater number of visually salient items as targets, this may have also introduced another element of noise into the data.

In order to remedy these sources of noise, Experiment 4 introduced a balanced design. Each item was presented for the same number of times as a target in each array position and in each condition across the entire study. Items were also presented the same number of times as decoys in each array position and in each condition. Finally, each participant received the same number of trials in the old and new condition (however, they only viewed a particular item once during their session).

Methods

Participants

96 student participants (74 female, 22 male) were recruited by convenience from the psychology department subject pool resource at the University of Edinburgh and through university based internet boards. The average age of participants was 21.5 years (range, 18-37). All participants had normal or corrected to normal vision and were either native or non-native English speakers. Course credit was awarded for participation. Consent forms were

provided to participants who were required to sign them as acknowledgement of the experimental aims, risks and data anonymity. The consent form included the participant’s right to terminate the session at any time without reason. The experiment met the British Psychological Society’s ethical guidelines and was approved by the University of Edinburgh psychology department’s ethical approval panel.

Stimuli

Experiment 4 used stimuli created by Rossion and Pourtois (2004) (retrieved from http://www.nefy.ucl.ac.be/facecatlab/stimuli.htm), which were colourised versions of stimuli originally developed by Snodgrass and Vanderwart (1980) and consisted of coloured pictorial representations of concrete nouns from various categories such as clothing, furniture, tools, kitchen utensils, musical instruments, fruit, vegetables and animals. Stimuli from the Rossion and Pourtois (2004) set were used as study items and test items. Items were displayed on a white background, each in one of the four quadrants of the array, and were arranged into 52 groups of 5. In assigning items to groups, the main concern was to achieve some degree of within-group variability in the conceptual and visual features of each item. Therefore, items were arranged into groups on the basis of their familiarity ratings provided in Rossion and Pourtois (2004). In the Rossion and Pourtois (2004) study, participants rated the familiarity of each item on a 5 point scale with a rating of 1 assigned to items that were judged to be very unfamiliar and a rating of 5 for very familiar items. Familiarity was defined as “the degree to which you come in contact with or think about the concept". Accordingly, all 260 items were first sorted by their familiarity rating and the item with the lowest familiarity rating was then grouped with the 53rd item in the list along with the 105th, 157th and 209th items. This process was repeated, starting with the 2nd lowest rated item which was grouped with the 54th, 106th, 158th and 210th, etc. This ensured that the items within each group had a broad range of familiarity ratings. Rossion and Pourtois (2004) also provide subjective ratings of item complexity, measured by asking participants to rate each image on “the amount of detail or intricacy of line in the picture". Since the basic correlation between ratings of familiarity and complexity was medium sized and negative ($r = -0.53$), this method of grouping items by ratings of familiarity was expected to also result in a modest degree of within-group variability in visual complexity. Once items were assigned to groups, 1 item in each group was randomly selected to act as the new item while the 4 remaining items were used in the study phase. Arrays were displayed at a resolution of 1024 x 768 pixels. 4 groups of stimuli were used in practice trials while the remaining 48 groups were used in the experimental session. All participants received the same stimuli in practice and experimental trials.

Procedure

Trial procedure

Participants answered advertisements stating that the experiment was investigating visual memory. Upon entering the experimental room, participants were sat comfortably at the eye tracker and viewed instructions displayed on the monitor. A calibration procedure was then performed requiring participants to fixate nine points arranged in a grid on the display which was then repeated for validation. Three practice trials were then presented. After the practice trials, participants were reminded of the instructions and eye position was re-calibrated.

Experimental trials then proceeded according to the following steps which are summarised in figure 13. A drift correction screen presented a fixation circle at the centre of the display. When the participant fixated the centre of this marker, the experimenter initiated the trial. An array was then presented for 4 seconds while eye movements were recorded. A blank screen was then displayed for 1 second followed by presentation of the probe item for yes/no recognition. In the old condition, the target for the given trial was presented, whereas in the new condition a novel item was presented. Each probe item was presented for 10 seconds and participants could respond ‘yes’ or ‘no’ with the use of a hand held controller at any time during the clip, although they were encouraged to respond as quickly and accurately as possible. Once the participant made their response, the probe item remained on the display for the entire 10 seconds ensuring that all participants viewed probe items for the same amount of time. During each trial, the experimenter faced a monitor with his back to the eye tracker display. The experimenter’s monitor displayed information about eye position and calibration but did not display any information about which particular array was being viewed by the participant during the experiment.

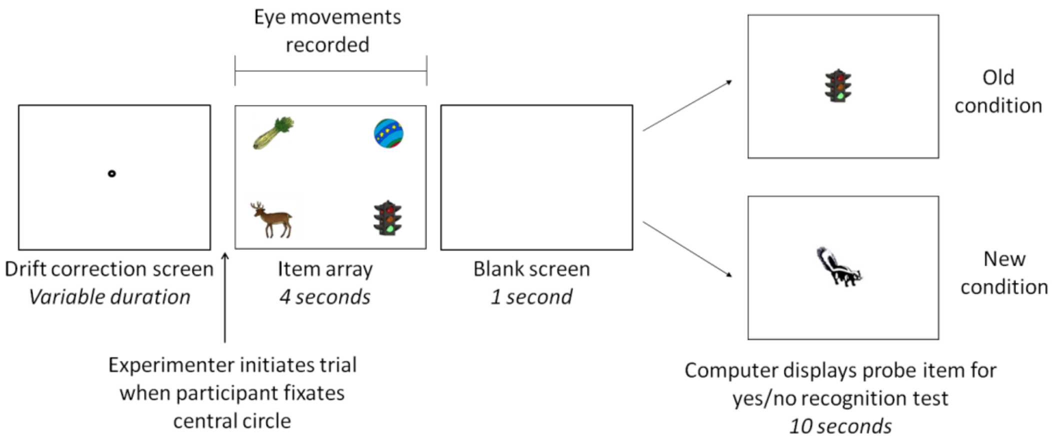


Figure 13. Schematic showing the basic experimental procedure of a single trial. A drift correction screen was first presented with a fixation circle located at the centre. Once this circle was fixated, the experimenter initiated the trial. An item array was then presented for 4 seconds while eye movements were recorded followed by a blank screen for 1 second. The computer then displayed a probe in the test phase. In the old condition, the target item for the given trial was presented whereas in the new condition a novel item was presented.

Balancing of variables

Each participant received 24 trials in the old condition and 24 trials in the new condition presented in a random order with no constraints on the number of trials in a particular condition appearing in succession. The Mersenne Twister core generator (Matsumoto & Nishimura, 1998) was used to randomise the order of trials. Each array of stimuli was presented once to each participant. A fully balanced design was employed such that each item appeared in each array position for the same number of times in each condition across the study. In addition, each item appeared as the target for the same number of times in each array position and in each condition. 96 participants were required for three repetitions of the fully balanced stimulus list. This meant that after all participants had been recruited; each item appeared 12 times as a target (3 times in each array position) and 36 times as a decoy (9 times in each array position) for each of the old and new condition.

Detection of fixations and saccades

Interest areas were drawn in each quadrant of the array and were 400 pixels in width and 300 pixels in height. These dimensions were sufficient to surround all items and ensured that fixations that did not land directly on the item, but were nevertheless directed towards the item, were captured by the interest area. All fixations and saccades that landed within an IA were automatically assigned to the appropriate item for later analysis. All IAs were constructed before data collection, remained fixed in size and position and were invisible to participants during the experiment.

Hypotheses

The hypotheses were the same as in Experiment 3A and 3B in that the allocation of visual attention to items would differ between old and new trials. Since all items in the new condition were categorically equivalent, targets in old trials were compared with all items in new trials. The hypotheses in Experiment 4 were therefore as follows:

1) The mean TDT for target items in the old condition will be different than the mean TDT for all items in the new condition.
2) The mean FFT for target items in the old condition will be different than the mean FFT for all items in the new condition.

Again, since the type of information underlying the hypothesised selective response to targets in old trials is not specified, the hypotheses are bidirectional and two-tailed tests are used.

Planned Analyses

Two statistical tests are reported per comparison; one with participants as the unit of analysis (n = 96) and one with items as the unit of analysis (n = 192). All tests were two-tailed and were performed on untransformed data. In addition, a Bonferroni correction for two comparisons was applied to tests of Hypothesis 1 and 2. Effect sizes are reported as Cohen's *d*. Data trimming procedures were applied post-hoc to assess the influence of outliers.

Results

Data Exclusions

If data from a particular participant was rejected before analysis (for example, due to a poor eye tracking signal), another participant was recruited to obtain a total of 96 participants required for full balancing. To ensure the experiment was fully balanced, data from all trials was included regardless of whether the participant gave an incorrect response in the recognition test.

Reaction Times and Response Accuracy

The mean RT for trials with correct responses was 841ms in the old condition and 818ms in the new condition, $t(94) = 2.06$, $p = 0.04$ (two-tailed). Overall accuracy was high; 96.8% and 97.6% correct responses in old and new trials respectively.

General viewing behaviour

The general viewing behaviour in Experiment 4 continued to replicate that of previous experiments. Participants tended to fixate the item in the top left corner first and continued in an approximately clockwise fashion. Table 9 shows the relationship between array position and fixation sequence and figure 14 shows the mean FFT for all items in each array position for Experiment 4.

		Sequence			
		1st	**2nd**	**3rd**	**4th**
Array Position	**1**	82.7	11.5	1.8	4.0
	2	11.6	72.1	9.0	7.3
	3	4.2	10.3	33.6	51.8
	4	1.5	6.1	55.6	36.8

1	2
3	4

Array position

Table 9. The percentage of items that were fixated 1^{st}, 2^{nd}, 3^{rd} and 4^{th} as a function of position in the array in Experiment 4. Array positions are shown schematically.

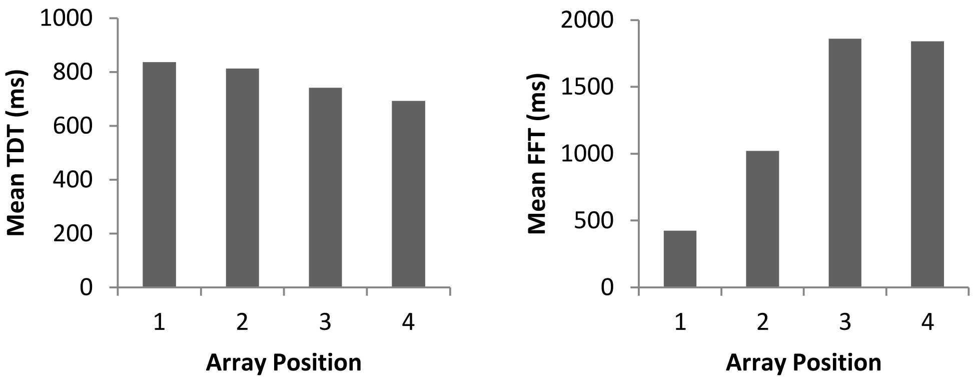


Figure 14. The mean FFT (left) and mean TDT (right) for all items as a function of position in the array in Experiment 4.

Participants tended to fixate items in position 1 for the longest amount of time (M = 836.8, SD = 112.3) followed by position 2 (M = 812.3, SD = 108.1), position 3 (M = 740.9, SD = 105.2) and finally position 4 (M = 692.7, SD = 101.3). A one-way ANOVA with array position as a factor revealed a highly significant effect of array position, $F(3,191) = 201.8$, $p < 0.001$. Further paired t-tests (with Bonferroni corrections) revealed that TDT diminished as a function of position such that pos1 > pos2, $t(191) = 3.36$, $p < 0.001$, pos2 > pos3, $t(191) = 9.92$, $p < 0.001$, pos3 > pos4, $t(191) = 7.24$, $p < 0.001$.

Test of Hypothesis 1

Kolmogorov-Smirnov goodness of fit tests suggested that TDT data for old targets and all new items was normally distributed (participants, old-targets, $D(96) = 0.095$, $p > 0.2$; new, $D(96) = 0.110$, $p > 0.2$; items, old-targets, $D(192) = 0.049$, $p > 0.2$; new, $D(192) = 0.043$, $p > 0.2$), therefore parametric tests were used in the planned analysis. Table 10 shows the mean TDT for items in the old and new conditions. Target items in the old condition were fixated approximately 14.6ms less than items in the new condition. However after Bonferroni correction for two comparisons, this difference was marginally non-significant for both participant and item analyses; participants, $t(95) = -1.82$, $p = 0.07$, $d = -0.23$; items, $t(191) = -1.84$, $p = 0.07$, $d = -0.13$. Nevertheless, the pattern of data was consistent with the precognitive hypothesis in that the degree of attention allocated to targets and decoys appeared to be different in the old condition but not in the new condition. However, the difference between targets and decoys in the old condition could not be assessed statistically since data from each group was not obtained independently and would thus violate the independence assumption of the statistical test.

In order to assess the influence of potential outliers, a 10% trim was applied to the raw data; 5% of data points with the highest TDT values and 5% of data points with the lowest TDT values were removed. This removed all TDT values above 1341ms and less than 334ms. After this trimming procedure, the data showed that target items in the old condition were fixated for approximately 14.5ms less than items in the new condition (see table 10). This difference was now found to be significant for both participant and item analyses; participant, $t(95) = -2.12$, $p = 0.04$, $d = -0.28$; items, $t(191) = -2.35$, $p = 0.02$, $d = -0.19$ (Bonferroni corrected for two comparisons). Although the effect size expressed in milliseconds was slightly less, effect sizes expressed as Cohen's *d* were larger than the main analysis due to a reduction in the standard deviation of item and participant means obtained from the trimmed raw data. The result of this trimmed analysis suggests that the marginally non-significant result from the main analysis may have been affected by outliers. When data was trimmed, a stronger central tendency was revealed. Hypothesis 1 was therefore confirmed; although this confirmation should be treated with caution since significance was obtained only after the data was trimmed.

	All data		10% trimmed	
	Old	New	Old	New
Targets	756.7 (121.1)	770.8 (116.1)	744.3 (88.1)	760.1 (80.8)
Decoys	774.5 (101.5)	771.5 (96.5)	759.7 (71.1)	758.4 (70.6)
All items	770.0 (98.2)	771.3 (93.8)	755.8 (66.2)	758.8 (64.7)

Table 10. Mean TDT (ms) for targets, decoys and all items in the old and new condition. Analyses for all data and 10% trimmed data are shown. Standard deviations are in parentheses. Values shown are from items as the unit of analysis.

Test of Hypothesis 2

FFT was found to be normally distributed in all datasets (Kolmogorov-Smirnov: participants, old-targets, $D(96) = 0.088$, $p > 0.2$; new, $D(96) = 0.081$, $p > 0.2$; items, old-targets, $D(192) = 0.056$, $p > 0.2$; new, $D(192) = 0.051$, $p > 0.2$) therefore paired t-tests were used in all comparisons. Table 11 shows the mean FFT for targets, decoys and all items in the old and new condition. The mean FFT for targets did not significantly differ from items in the new condition; participants, $t(95) = 0.39$, $p = 0.69$, $d = 0.02$; items, $t(191) = 0.68$, $p = 0.50$, $d = 0.09$. A 10% trim also resulted in a non-significant difference (participants, $t(95) = 1.33$, $p = 0.19$, $d = 0.12$; items, $t(191) = 0.97$, $p = 0.33$, $d = 0.15$). Therefore, hypothesis 2 was not confirmed suggesting that the degree to which items attracted attention from the periphery did not differ between conditions.

	Old	New
Target	1308.3 (167.0)	1282.8 (191.1)
Decoy	1300.7 (127.6)	1298.9 (120.6)
All items	1302.6 (112.1)	1294.8 (110.6)

Table 11. Mean FFT (ms) to targets, decoys and all items in the old and new conditions. Standard deviations are in parentheses.

Exploratory Analyses

Multiple regression: predictor variables

Since Hypothesis 1 was confirmed, an exploratory analysis was carried out on the TDT data in order to examine whether the apparent precognitive effect was systematically related to normative data on the stimuli used in this study. In developing their original stimulus set, Snodgrass and Vanderwart (1980) selected concrete nouns based on several inclusion criteria as follows:

1. The concepts must be unambiguously picturable.
2. They must include exemplars from the widely used category norms of Battig and Montague (1969).
3. They must represent concepts at the basic level of categorization (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

Once the picture set had been created, normative data was collected on several stimulus dimensions including ratings of *visual complexity*, *image agreement* and *concept familiarity*. Rossion and Pourtois (2004) also provide similar normative data for their colourised stimulus set using the same procedures reported in Snodgrass and Vanderwart (1980). Other studies have collected data on *naming time* in British English (Barry et al., 1997). It was decided to explore the extent to which the precognitive effect was related to the four stimulus dimensions of complexity, image agreement, familiarity and naming time as described below. These variables have been proposed to reflect activity at various processing stages of visual object recognition and picture naming (Alario et al., 2004; Barry et al., 1997) and are thought to represent distinct aspects of the picture (Snodgrass & Vanderwart, 1980). Since participants were expected to be engaged in both object recognition and subvocal picture naming in the current paradigm, a potential correlation between these normative variables and the TDT effect may shed some light on the contribution of these processing stages in the expression of the effect. Each normative variable will now be described.

Visual Complexity

Visual complexity refers to subjective ratings of how complex the form of the picture appears to be. Normative data on this variable was taken from Rossion and Pourtois (2004) who asked participants to rate each picture in their colourised set on “the amount of detail or intricacy of line in the picture” (p. 223). Participants were asked to rate the complexity of the drawing on a 5 point scale and rate the picture itself rather than the real-world object it represented. A rating of 1 was given to pictures with very low complexity and a rating of 5 to pictures with very high complexity. This variable is thought to reflect processing at the level of object recognition including structural analysis and pattern recognition (Alario et al., 2004). There is some evidence that object recognition processes take longer with pictures of more complex objects compared to simple objects (Attneave, 1957; Ellis & Morrison, 1998) but other studies have reported so such differences (Biederman, 1987; Paivio, Clark, Digdon & Bons, 1989). If more complex objects take longer to process visually, this may impact on the time spent fixating these objects in the current paradigm. In addition, the structural intricacy of relatively more complex objects may make them more visually salient and therefore more likely to attract and hold attention (Itti & Koch, 2000). If so, we might expect more complex items to be fixated earlier and for longer than less complex items as participants view the array.

Concept familiarity

Familiarity refers to extent to which the concept, represented by the picture, has been encountered in everyday life. In the Snodgrass and Vanderwart (1980) study, participants were asked to rate each picture "according to how usual or unusual the object is in your realm of experience” (p. 183) and familiarity was defined as “the degree to which you come in contact with or think about the concept” (p. 183). Participants were asked to rate the concept itself rather than way it was drawn. A rating of 1 indicated very low familiarity and a rating of 5 indicated very high familiarity. Familiarity is thought to be a pictorial version of word frequency and is likely to reflect how well the picture activates its semantic representation (Hirsh & Funnell, 1995). Participants contributing to the Rossion and Pourtois (2004) norms were French speakers, therefore familiarity norms were taken from Barry et al. (1997) who used British participants. This was to avoid potential differences in familiarity judgements that may arise from participants of a different nationality, which may occur if the occurrence of particular concepts varies between nationalities and cultures. Familiarity norms from Barry et al. (1997) were obtained using the original line drawings from Snodgrass and Vanderwart (1980) rather than the colourised versions used in Experiment 4. However, since judgements of familiarity were based on the concept itself rather than on the visual characteristics of the picture, it is likely that these norms are applicable to the colourised stimulus set.

Image agreement

Image agreement is the extent to which the image generated from an object's name matches the picture of the same name. Ratings of image agreement were taken from Rossion and Pourtois (2004) who presented participants with the name of an object followed by three seconds of blank screen and participants were required to form a mental image of the named object. The actual picture was then presented and participants rated how closely their image matched the visual characteristics of the picture. A rating of 1 indicated a low match while a rating of 5 indicated a high match. It has been suggested that this variable acts at the level of object recognition and reflects the degree to which the activation of stored structural descriptions of a particular object match the presented picture (Alario et al., 2004; Barry et al., 1997).

Naming time

Naming time is the time taken to name the object upon presentation of the picture. Norms were obtained from Barry et al. (1997) who used British English speaking participants. Naming latencies were recorded in milliseconds and were measured from picture onset to onset of a verbal response. Participants were required to name each object "as distinctly and quickly as possible" (p. 568). However, the Barry et al. study used the original set of line drawings created by Snodgrass and Vanderwart (1980) to obtain their normative data whereas the dependent variables in Experiment 4 were obtained using the colourised version of that stimulus set. As shown by Rossion and Pourtois (2004), the inclusion of colour significant speeds naming time, therefore use of the Barry et al. norms may not accurately reflect the time taken to name coloured pictures. However, Rossion and Pourtois (2004) also obtained naming times for line drawings and the correlation between their colour and line drawing naming time data is large (r = 0.78). Therefore, it will be assumed that use of naming times obtained from line drawings will not have a notable effect on this exploratory analysis. The time taken to name a pictorial object will reflect process involved in object recognition, semantic analysis, lexical retrieval and articulation (Levelt, 2001). In previous research, participants were presented with arrays of objects taken from the Snodgrass and Vanderwart picture set and were asked to memorise each item in preparation for a memory test (Zelinsky & Murphy, 2000). The study showed that the total time spent looking at objects was related to the number of syllables in the object's name with longer names resulting in greater dwell time. Therefore one might expect the variable of naming time to correlate with item TDT in the current paradigm.

Multiple regression: outcome variables

Using items as the unit of analysis, the four variables of complexity, familiarity, image agreement and naming time were included as predictors in two simultaneous multiple regressions using the following as the outcome variable:

1) *TDT-All*, defined as the TDT per item collapsed across both old and new conditions
2) *ΔTDT(Target-New)*, defined as the difference in TDT between targets in the old condition and items in the new condition

Analysis 1 aimed to explore the extent to which attention was allocated to items on the basis of the stimulus dimensions of complexity, familiarity, image agreement and naming time whereas analysis 2 aimed to explore the extent to which the apparent precognitive effect was related to these same stimulus dimensions. Analysis 2 was performed on the 10% trimmed data since that dataset appeared to provide a stronger effect in the main analysis.

Results of the multiple regression analyses

TDT-All analysis

The overall regression equation was significant, accounting for 36% of the variance, F (4,187) = 26.3, $p < 0.0001$, $R^2 = 0.36$. With the exception of familiarity, all variables entered into the multiple regression emerged as highly significant predictors of TDT-All. Naming time was the strongest predictor ($p < 0.001$) and was positively related to TDT-All suggesting that items taking longer to name held attention for longer. This is consistent with previous research showing that naming time is a major determinant of the time spent looking at objects during memorisation, with longer dwell times on items that have more syllables in their name (Zelinsky & Murphy, 2000). Thus, the high correlation between TDT-All and naming time suggests that participants were subvocally naming items while viewing the array. Image agreement had the next largest effect ($p < 0.001$) and was negatively related to TDT. This may be because items that more closely match stored structural descriptions of the object are more easily identified and encoded into memory and less time is therefore spent fixating these items. Complexity also had a significant independent effect ($p < 0.01$) and was positively related to TDT-All. This may have arisen because complex items are more visually salient (Itti & Koch, 2000) and are therefore more likely to attract and hold attention or perhaps that object recognition processes take longer to complete with more complex stimuli (Attneave, 1957; Ellis & Morrison, 1998). Finally, familiarity had a marginally significant effect on predicting TDT-All ($p = 0.057$) with a negative relationship suggesting that less attention was allocated to more familiar items. This may be because the semantic network of familiar items is more easily activated which may facilitate the encoding of these items into short term memory thus reducing fixation time.

ΔTDT(Target-New) analysis

After entering all variables simultaneously, the overall regression equation was non-significant, F (4,187) = 1.52, $p = 0.20$, $R^2 = 0.031$. Both image agreement and naming time had non-significant effects. However, both complexity and familiarity had independent effects that were marginally non-significant (both $p < 0.06$) and were both negatively related to the dependent variable. If this suggestive relationship does not represent a false positive finding, it may indicate that the precognitive effect observed in Experiment 4 was greater for more familiar and more visually complex items.

	Outcome variables		Predictor variables			
	TDT-All	ΔTDT (Target-New)	Comp.	Fam.	Imag.	Nam.
Comp.	0.293**	-0.083	1			
Fam.	-0.314**	-0.099	-0.418**	1		
Imag.	-0.263**	0.068	0.037	-0.211*	1	
Nam.	0.454**	0.016	0.053	-0.373**	-0.027	1

Table 12. Basic correlations between the predictor variables and TDT-All, and ΔTDT(Target-New). ** denotes $p < 0.001$, * denotes $p < 0.01$

	Predictor	*B* Coeff.	Standard Error	*β*	*t* value	*p*
TDT-All	Comp.	22.33	6.43	0.226	3.47	0.001
	Fam.	-12.83	6.71	-0.138	-1.91	0.057
	Imag.	-35.24	7.33	-0.290	-4.81	0.000
	Nam.	0.234	0.039	0.383	5.98	0.000
ΔTDT (Target-New)	Comp.	-12.825	6.712	-0.153	-1.91	0.058
	Fam.	-13.394	6.996	-0.170	-1.91	0.057
	Imag.	3.821	7.650	0.037	0.50	0.309
	Nam.	-0.020	0.041	-0.038	-0.49	0.313

Table 13. Multiple regression analyses on each outcome variable, TDT-All and ΔTDT(Target-New).

Discussion

Experiment 4 was a further test of the precognitive hypothesis and implemented several changes in design. Namely, a new stimulus set was introduced, study and test items were drawn from the same stimulus pool and a balanced design was introduced to counter potential confounding effects of target position and identity. After applying a 10% trim to the raw data, the results of the main analysis showed that participants spent less time fixating target items in the old condition compared to items in the new condition, suggesting that less attention was allocated to targets. This result is consistent with the precognitive hypothesis, since old trials consisted of the repeated presentation of the target whereas new trials did not. Moreover, the pattern of data suggested that the amount of time spent fixating targets was less than the amount of time spent fixating decoys in the old condition, while no such difference was apparent in the new condition. This is also consistent with the precognitive hypothesis since targets and decoys are categorically equivalent in the new condition and were predicted to show no difference in attentional allocation (within-condition differences cannot be tested statistically since target and decoy data was not obtained independently). However, the direction of the effect was opposite to that found in Experiment 1 where target items were fixated for longer. This represents somewhat of an inconsistency in findings, but as this pertains to the results of the thesis as a whole, issues of result directionality will be raised in the general discussion of Chapter 5.

In addition, there was suggestive evidence from an exploratory analysis that the TDT effect was negatively correlated with item complexity and familiarity. This suggests that less attention was allocated to targets if they were rated as more visually detailed or more familiar. However, the precognitive effect sizes from the main analysis and the exploratory analysis were small, raising the possibility that the results were due to methodological problems. Therefore, the following section will address a number of potential explanations for the results of Experiment 4 based on known factors.

Explanation 1: Inappropriate use of statistical tests

Two-tailed tests were used in the main analysis and were corrected for two multiple analyses. Therefore, it cannot be argued that the result were due to inflation of type I error from inappropriate use of one-tailed testing or lack of correction for multiple analyses, a criticism that has been levelled at previous research (Wagenmakers, et al., 2011).

Explanation 2: Selective data reporting

The main analysis in Experiment 4 was carried out on all available data. No trials were excluded on the basis of incorrect responses in the recognition task, nor were any outliers removed. Therefore, the result of the main analysis cannot be attributed to data exclusion. However, while the experiment was being carried out, data from a small number of participants was removed from the study prior to any statistical analysis. This was based on a qualitative assessment of the eye tracking data for each participant after they had been run through the experiment. For example, if an excessive amount of missing data was detected (due to loss of the eye tracking signal) or if an excessive number of fixation durations less than 100ms were observed (due to an intermittent eye tracking signal), then data from these participants was removed and another participant was recruited to take their place. However, since the data had not been statistically analysed at that time and the number of excluded participants was low ($n = 6$), it is unlikely that removal of these participants introduced a selection bias into the dataset.

Explanation 3: Sensory cues

In previous experiments, the target and/or trial condition was randomly selected by the computer after the dependent variable had been recorded. In those cases, issues of sensory cueing are ruled out entirely. However, in Experiment 4, the sequence of target positions and trial conditions was randomised at the start of the session and was therefore predetermined at the start of each trial. One might argue that audible cues originating from retrieval of images from the computer’s hard disk at the start of each trial provided participants with information about the trial condition or target item. However, this type of explanation seems implausible since the computer controlling the display of stimuli was located inside a closed cabinet some distance away from the participants and, upon inspection, the experimenter could hear no audible cue from the hard disk when seated at the eye tracker. Explanations in terms of sensory cues can therefore be safely ruled out.

Explanation 4: Statistical cues

The most plausible explanation for the results of the TDT analysis is one based on statistical cues, present as a consequence of the bias contained in a random sequence of trials that have a balanced frequency of conditions or stimulus parameters. In Experiment 4, participants received 24 trials in each condition and the 24 target items presented in each condition were evenly distributed over the four array positions (6 times in each position). Therefore, although trials were presented in random order, the balanced protocol raises the possibility that participants may have used statistical information from previous trials to predict (consciously or unconsciously) the upcoming trial condition and target position at a level above that expected by chance alone. This explanation is of particular importance not only to this study, but to experimental psychology as a whole since balanced designs are routinely used in studies of cognition and perception. If the results of Experiment 4 are indeed the result of conscious or unconscious use of statistical cues, then many other studies that use balanced designs may be susceptible to similar effects which may impact on the interpretation of their results.

It should firstly be noted that in order for statistical cues to result in the observed effect in Experiment 4, participants would need to make a prediction about both the current trial condition *and* the position of the target item in the array based on past information. A *cumulative count model* (CCM) may be able to predict the current trial condition above a probability of 0.5 by counting the number of previous trials in each condition and making a prediction according to the condition with the lowest count. On average, not only would a CCM increase the probability of correctly predicting the trial condition above 0.5, but the model need not include information about the total number of trials in the session. Similarly, to predict the current target position above a 0.25 probability, a count could be made of all array positions that the target occupied in previous trials (but only old trials since information about target position in new trials was not available to participants) and a prediction made on the basis of the position with the lowest count. If participants were implementing some form of CCM, either consciously or unconsciously, then once a prediction of the current trial condition and target position had been made, eye movements could accordingly be biased towards or away from the predicted target position which would, on average, result in a dependent variable differential when targets and decoys were compared. A CCM would also result in an increase in prediction accuracy as the session progressed since statistical information about the upcoming trial condition and target position would become more reliable over the course of a session.

Is it plausible that participants would engage in statistical processing of this sort during the experiment? Given that participants were required to memorise each array item rather than make any prediction about the trial condition or target position, this seems intuitively unlikely. Inter-trial memory effects have been demonstrated in other paradigms but are relatively short lived. For instance, an effect labelled ‘priming of pop out’ occurs during visual search experiments when the search target on a particular trial occupies the same location as the target on the previous trial (Maljkovic & Nakayama, 1996). Here, the deployment of attention to the repeated target location is facilitated, an influence that extends to about five to eight subsequent trials (Maljkovic & Nakayama, 1996). Clearly, a CCM prediction based only on information from a handful of preceding trials would be at a considerable disadvantage compared to a prediction based on information from all preceding trials. However, the current paradigm differs in many ways to the one used to elicit priming of pop out effects and it cannot be ruled out that participants were making predictions about current trials based on all available information since the start of their session.

Nevertheless, the data from Experiment 4 can be examined to see if it is consistent with the predictions of a CCM. Firstly, a cumulative count simulation can be run on the actual sequence of trials observed in Experiment 4 to examine whether the predictions of such a model correlate with eye movement behaviour. Accordingly, a CCM was applied to the actual sequence of trials generated for each participant in Experiment 4 and a prediction about the current trial condition and target position was made on the following basis:

1) To predict the current trial condition, the model counted the number of previous trials in each condition for a given participant. The condition with the lowest count was predicted as the current trial condition. If the count for each condition was equal, no prediction was made since no statistical cue was present on that occasion.

2) If the current trial was predicted as “old” from step 1, the model counted the number of times the target appeared in each position for all previous trials that were in the old condition. The position with the lowest count was predicted as the target position on the current trial. If all position counts were equal, no prediction was made.

The above model should produce the following pattern of data if it is to explain the results of Experiment 4: On occasions when the current trial is predicted to be in the old condition, predicted targets should have a lower TDT than predicted decoys. In fact, the opposite was the case. When the model predicted the current trial as old, predicted targets had a mean TDT of 778.3ms (n = 1068) whereas predicted decoys had a mean TDT of 765.4ms (n = 3204).

Secondly, a CCM would predict an increase in the size of the effect over the course of the session since the probability of a correct prediction increases and participants would also have had more opportunity to recognise any underlying pattern to the placement of target positions. In order to test this prediction, the data from Experiment 4 was first split by the first and last half of each participant's session with the last half predicted to exhibit a greater effect. However, the data shows that the effect is largely confined to the first half of each session (see table 14). The difference in TDT between targets and new items in the first 24 trials was approximately 25.8ms with items as the unit of analysis, $t(191) = -2.64$, $p < 0.01$, $d = -0.21$ (Bonferroni corrected for 2 comparisons), and approximately 27.8ms with participants as the unit of analysis, $t(95) = -3.10$, $p < 0.01$, $d = -0.40$ (Bonferroni corrected for 2 comparisons). In contrast, the difference in TDT between targets and new items was trivial in the last 24 trials (by participants, 4.4ms, p = n.s.; by items, 3.3ms, p = n.s.).

	First 24 trials		Last 24 trials	
	Old	New	Old	New
Target	750.1 (135.7)	770.3 (158.7)	763.5 (155.1)	766.9 (147.3)
Decoy	781.7 (122.4)	776.5 (110.4)	765.6 (106.3)	766.3 (109.9)
All items	773.6 (113.7)	775.9 (107.2)	766.3 (102.1)	766.9 (99.5)

Table 14. Mean TTD (ms) to targets, decoys and all items in the old condition split between the first and last 24 trials in each session. Standard deviations are in parentheses. Values obtained from items as the unit of analysis.

As a further test, the difference in TDT between targets and items in the new condition was plotted against the trial sequence averaged over all participant sessions. The CCM would predict an increase in the size of the effect over the course of the session but, as shown in figure 15, the data actually suggests that the effect declines as a function of trial sequence, although the correlation is not significant (*r* = 0.11, n.s.). It should be noted that each data point is not cumulative so the decline in effect cannot be attributable to regression to the mean. Together, these analyses suggest that participants were not capitalising on the statistical information contained in the sequence of trial conditions and target positions and argue against an interpretation of the results of Experiment 4 in terms of statistical cueing.

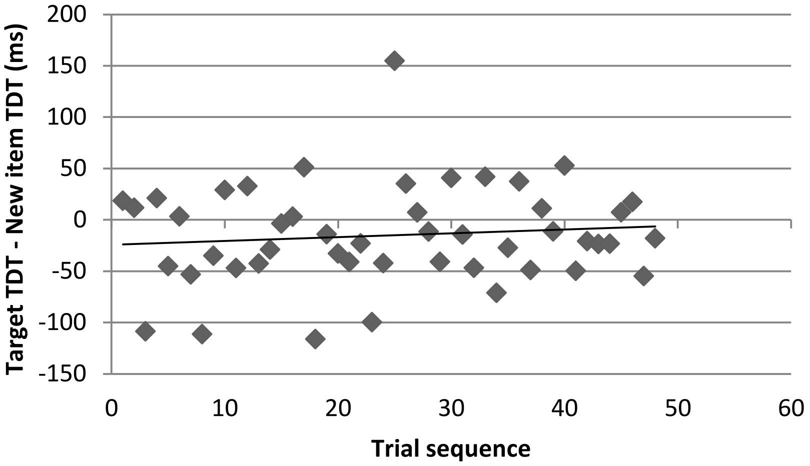


Figure 15. The difference between Target TDT and TDT for all items in the new condition as a function of trial sequence, averaged over all participant sessions.

However, several cautionary points are warranted here. Firstly, the analyses may be confounded by the effects of target position and target item identity since each individual analysis does not contain a balanced frequency of these parameters. For example, a greater percentage of targets may have been presented in array positions associated with lower TDT in the first 24 trials. Indeed, the data suggests that this is the case, although the bias is small; for the first 24 trials in the old condition, 48.6% of targets were collectively presented in positions 1 and 2 (the array positions associated with greater TDT) whereas 51.4% of targets were collectively placed in positions 3 and 4. However, since a reciprocal bias occurs in the last 24 trials due to the balanced design, we might expect a similar TDT difference in the opposite direction yet no such difference is observed. However, the analyses assume that participants' responses to statistical cues are constant throughout the session. If participants are initially very responsive to such cues, but later on in the session become much less responsive, then this may have offset any increase in prediction accuracy resulting from more reliable statistical cues as the session progressed. Therefore, although the results cannot be readily explained by statistical cueing, such an account cannot be ruled out entirely.

In summary, confirmation of hypothesis 1 is not readily explainable in terms of inappropriate statistical tests, selective data reporting, sensory cues or statistical cues. However, there remains one potential explanation that will now be discussed in detail in the following chapter. This explanation, termed the *expectation bias*, has received some attention in previous literature as a potential explanation for the positive results of previous studies of precognition, in particular the presentiment studies. The following chapter will firstly describe the mechanics of the bias and how it has been proposed as a potential explanation for previous studies of precognition. Secondly, a simple model of the bias will be presented as it applies to the current paradigm and an estimate made of its potential contribution to positive results. Finally, data from Experiments 1 to 4 will be used to assess the extent to which the bias may have been present in each case.

Chapter 4 - Exploring the effects of an expectation bias in the current paradigm: Experiments 5A and 5B

Introduction

Many of the experiments described in Chapter 2 were testing for precognition using physiological measures. Arguably, the most popular type of study to employ such methodology has become known as the *presentiment* experiment. In a typical experiment, participants sit at a computer screen and are simply required to passively view a series of emotional and calm pictures presented to them in a random sequence. During this time, some measure of their autonomic arousal is continuously recorded both before and during presentation of the pictures. It is well established that the autonomic nervous system responds to the presentation of emotionally arousing stimuli. For example, arousal will be higher when an emotional picture is shown compared to when a calm picture is shown. However, the presentiment hypothesis proposes that participants' level of autonomic arousal will also be higher moments before viewing emotional pictures compared to calm pictures, despite the fact that the two types of stimuli are presented in a random sequence. A number of presentiment experiments have reported anomalous anticipation of emotional stimuli using measures of skin conductance (Bierman & Radin, 1997; Radin, 2004), heart rate (McRaty et al., 2004), pupil dilation (Radin & Borges, 2009) and fMRI (Bierman & Scholte, 2002) and other studies have also used the same basic paradigm to test for anticipatory brain activity preceding emotionally neutral stimuli such as flashes of light (Radin & Lobach, 2007).

A test of the presentiment hypothesis involves calculating mean values of the dependent variable in each condition and comparing means to establish whether there is a statistically significant difference between conditions. If the dichotomous stimuli are presented in a random sequence, it is assumed that the overall mean values in each condition will be the same according to the null hypothesis. However, contrary to what one might expect, calculating means in this way can under certain circumstances lead to an artifactual difference between the mean values even when the sequence of stimuli is perfectly random. In order for this bias to arise, the pre-stimulus value of the dependent variable on a given trial must, to a certain extent, be systematically related to the outcome of previous trials. This bias has been described in detail in a number of publications (Dalkvist & Westerlund, 2006; Dalkvist, Westerlund & Bierman, 2002; Radin, 2006; Wackermann, 2002) and has been discussed as a potential explanation for the results of presentiment studies (Broughton, 2004; Radin, 2004; Spottiswoode & May, 2003).

An example of the how the bias can arise may be described in the following manner. Consider a presentiment experiment involving 4 participants, each receiving only two trials where the outcome of each trial is randomly determined to be either an emotional picture (E) or a calm picture (C). As shown in table 15, four participants are required to account for all possible permutations of trial outcomes from a two-trial experiment. Now suppose that a participant's level of pre-stimulus arousal is zero on the first trial and that each participant engages in the so-called 'gambler's fallacy' (Tversky & Kahneman, 1974), here characterised by an erroneous belief that the probability of a particular outcome on the second trial is dependent on the outcome of the first trial. Specifically, let us say that each participant believes that the second trial is likely to be E merely because the previous trial was C. Now also suppose that if the first trial is C then the participant's pre-stimulus arousal increases by one unit on the second trial due to the false expectation that the second trial will be E.

As can be seen from table 15, each participant's pre-stimulus arousal level starts at 0, representing no expectation about the outcome of the first trial. However, the pre-stimulus arousal level on the second trial changes as a function of the outcome from the first trial; if the first trial is C then the participant's pre-stimulus arousal level on the second trial is 1; if the first trial is E then the pre-stimulus arousal level on the second trial is 0. The *mean arousal* columns in table 15 specify the participant mean arousal values for C and E trials calculated across the two trials in their particular session. The *overall mean* values are the overall mean arousal levels for C and E trials calculated across participant means. We can see that, on average, the overall mean arousal for E trials (0.333) is greater than the overall mean arousal for C trials (0.167), representing a bias. Moreover, if participant 1 and 4 are removed to allow a within-subjects analysis (since they did not receive at least one C or E trial) then the difference between C and E trials becomes greater (0.5). Because this bias has been largely described by modelling participant expectations about upcoming trials in presentiment experiments (Dalkvist et al. 2002; Dalkvist & Westerlund 2006; Radin, 2004) it shall be henceforth be termed the *expectation bias*.[8]

Participant	Trial	Outcome	Arousal Level	Mean Arousal C trials	Mean Arousal E trials
1	1 2	C C	0 1	0.5	n/a
2	1 2	C E	0 1	0	1
3	1 2	E C	0 0	0	0
4	1 2	E E	0 0	n/a	0
			Overall Means:	0.167	0.333
				Bias = 0.167	

Table 15. A simple simulation of the expectation bias involving a two-trial experiment with four participants.

There is, however, a way in which the expectation bias can be avoided altogether in this two-trial example. If overall mean values for pre-stimulus arousal are simply calculated across all trials, without firstly calculating individual participant means, then the overall mean values in each condition are 0.25 and no bias occurs. However, as pointed out by Dalkvist and Westerlund (2006), every possible permutation of trial outcomes must be included in the calculation and each permutation must be present an equal number of times in order for this method to result in zero bias. With a two-trial experiment, this does not pose a practical problem since there are only four possible permutations. However, for an experiment with 30 trials per participant, Dalkvist and Westerlund (2006) calculate that 1.074 x 10^9 participants would be required to ensure that all possible sequences are included with equal frequency,

---

[8] Two similar terms are used in the current chapter; the *expectation effect* and the *expectation bias*. The term *expectation effect* refers to a change in the dependent variable over successive trials as a result of expectation based on the outcome of previous trials. The term *expectation bias* refers to the resulting difference between the mean values of the dependent variable when two conditions are compared.

thus reducing the bias to zero. Therefore, for practical reasons, presentiment experiments cannot include all possible permutations of trial outcomes and, under these circumstances, the mean arousal level will on average be larger for E trials compared to C trials whether means are calculated across all trials or across individual participant means.

The ‘expectation bias’ is, of course, not reliant on expectation effects *per se* and is not restricted to experiments measuring pre-stimulus arousal. All that is required is that the experiment employs two or more stimulus categories or conditions and the dependent variable varies in some systematic way with the outcome of previous trials. For example, sequential dependency effects have been observed during choice reaction time tasks (Cho et al., 2002; Soetens, Boer, & Hueting, 1985; Sommer, Leuthold & Soetens, 1999; Tanaka & Shimojo, 1996), visual search (Maljkovic & Nakayama, 2000) and saccadic go/no-go tasks (Emeric et al., 2007). Since, in all these cases, the dependent variable systematically varies in accordance with the outcome of previous trials, we would expect a bias to arise for the same reasons as described above. Moreover, if the experiment employs a balanced design, as many psychology studies do, it may also be susceptible to biases based on statistical cueing.

Dalkvist et al. (2002) report the results of a series of simulated presentiment ‘experiments’ that modelled participant expectations about upcoming stimuli based on the outcome of previous trials. In a manner similar to that illustrated in table 15, participants’ expectations were modelled as an increase in arousal after preceding ‘calm’ trials. Dalkvist et al. (2002) describe a simple *binary* model whereby pre-stimulus arousal on a given trial is set at 1 if the preceding trial was C and remains at that level until an E trial is encountered which resets the pre-stimulus arousal level to zero on the following trial. In the *linear* model, the pre-stimulus arousal on a given trial increases by 1 unit for every consecutive C trial unit it is reset to zero by an E trial. They also consider expectation functions with exponential and sigmoidal characteristics. In the *exponential* model, pre-stimulus arousal increases exponentially as a function of the number of previous C trials, whereas in the *sigmoid* model, pre-stimulus arousal is initially unresponsive to the first few consecutive C trials but accelerates as the number of consecutive C trials increases while finally levelling off at some maximum arousal level. Like the binary and linear models, pre-stimulus arousal in the exponential and sigmoid models is reset to zero by the occurrence of an E trial.

They estimate the bias effect size resulting from each iteration of the simulation according to the following formula, henceforth termed $Bias_a$:

$$\text{Bias}_a = \frac{\text{Mean}_E - \text{Mean}_C}{\text{Mean}_{All}} \text{ X } 100$$

Here, $Mean_E$ and $Mean_C$ are the overall mean pre-stimulus arousal values for 'emotional' and 'calm' trials respectively, whereas $Mean_{All}$ is the overall mean value collapsed across both conditions. As such, this is a relative percentage measure representing the size of the bias with respect to the average pre-stimulus arousal level. So, returning to the two-trial example illustrated in table 15, bias_a for this hypothetical experiment would be calculated as follows:

$$\text{Bias}_a = \frac{0.333 - 0.167}{0.25} \text{ X } 100$$

$$= 66.7\%$$

Thus, a two-trial 'experiment' of this type produces different mean arousal levels for calm and emotional trials and this bias is approximately two thirds the size of the overall mean arousal level. For each of the binary, linear, exponential and sigmoid models of expectation behaviour, Dalkvist et al. (2002) report the results of "about one million" (p. 68) simulated experiments where each experiment involved 16 participants, each receiving 32 trials. When the ratio of emotional to calm pictures was 1:1 and means were calculated over individual participant means, the size of bias_a for the binary, linear, exponential and sigmoid models was reported as 6.39%, 12.48%, 11.34% and 6.39% respectively. However, when means were calculated across individual trials, the bias_a value associated with each model was drastically reduced with no model producing a bias larger than 0.1%. Thus, Dalkvist et al. (2002) suggest that studies avoid calculating overall means across individual participant means because this method is particularly prone to generating the bias. Instead, they recommend that analyses calculate means across all trials in the experiment in order to reduce the impact of the bias, although they do caution that more complex patterns of expectation may lead to larger biases. After examining the size of the bias in relation to sample size, they conclude that it decreases in size as the number of trials increases and approaches zero as the number of trials approaches infinity.

As the above formula shows, the measure of bias$_a$ adopted by Dalkvist et al. (2002) is a relative measure with respect to the average pre-stimulus arousal level. Another potentially informative measure of bias is the size of the difference between the mean pre-stimulus arousal levels in each condition relative to the increase in arousal resulting from the modelled expectation behaviour. As far as the author is aware, this measure has not been previously described in the literature on the expectation bias and represents a novel means to assess the size of a simulated expectation bias relative to the expectation effect itself. Henceforth termed *bias$_b$*, the measure can be expressed according to the following formula:

$$\text{Bias}_b = \frac{\text{Mean}_E - \text{Mean}_C}{\text{Mean}_{PreC} - \text{Mean}_{PreE}} \text{ X } 100$$

Here, $Mean_E$ and $Mean_C$ are defined in the same way as in bias$_a$. However, $Mean_{PreC}$ is the mean pre-stimulus arousal level for trials in which the previous trial was C. Similarly, $Mean_{PreE}$ is the mean pre-stimulus arousal level for trials in which the previous trial was E. Again, if we return to the binary modelled two-trial example from table 15, bias$_b$ would be calculated as follows:

$$\text{Bias}_b = \frac{0.333 - 0.167}{1 - 0} \text{ X } 100$$

$$= 16.7\%$$

Thus, it can be seen that the size of bias_b is smaller than that of bias_a and provides an estimate of the size of the expectation bias relative to the increase in arousal resulting from expectation behaviour. It should also be noted that although the term $Mean_{Pre\ C} - Mean_{Pre\ E}$ will always a value of 1 in the binary model, the value of this term in linear, exponential and sigmoid models will vary as a function of how many consecutive C trials were generated by each randomised sequence. Values of bias_b may be useful when examining data from real presentiment experiments. For example, it could be argued that if values of bias_b tend to be relatively small across a number of reasonable models of expectation effect, we would expect to find relatively large expectation effects within datasets that generate significant presentiment effects, if such effects are indeed due to a bias of this kind. If, on the other hand, examination of experimental data does not provide evidence for the presence of relatively large expectation effects, despite the presence of concurrent presentiment effects, then an explanation for such positive results in terms of an expectation bias becomes less plausible.

A series of simulations were therefore carried out with the following aims. Firstly, to replicate and extend the simulation results of Dalkvist et al. (2002) by providing estimates of bias_a and bias_b modelled according to simple rules of participant expectation in a presentiment experiment with dichotomous stimuli. Secondly, to assess the potential contribution of the expectation bias to the current paradigm by modifying the model parameters accordingly. In addition, the data from each experiment in the current thesis was examined to determine whether expectation effects were present and, if so, to evaluate whether the size of such effects could provide a plausible basis for the positive results from Experiment 1 and 4.

Experiment 5A: A simulation of the expectation bias in a simple presentiment experiment

The aim of Experiment 5A was to run a series of simulated experiments modelling the ‘expectation bias’ and using the same model parameters described by Dalkvist et al (2002) in order to replicate their results. In addition, the methodology was extended to include a new relative measure of bias not previously described in the literature, here termed $bias_b$, that compared the difference between pre-stimulus arousal levels in each condition to the change in pre-stimulus arousal resulting from the modelled expectation behaviour.

Methods

Participants and sample size

The same number of participants and trials per iteration were used in Experiment 5A as were used in Dalkvist et al. (2002) so that the results of each study could be reasonably compared. Therefore, in line with the Dalkvist et al. study, 16 ‘participants’ were run in each iteration of the simulation. Each participant received 32 trials. 5000 iterations were run for each series.

Trial outcomes

Two trial conditions were specified in the model, ‘calm’ and ‘emotional’. The outcome of each trial was determined randomly, with replacement, and independently both within and between participants. The RANDBETWEEN function in Microsoft Excel 2007 was used for this purpose to determine the condition of each trial with equal probability.

Models

Two models of expectation behaviour were specified. In the *binary* model, arousal on the first trial for a given participant was always set at 0. If the outcome of trial n was ‘calm’ then the arousal level on trial $n+1$ was set at 1. If the outcome of trial n was ‘emotional’ then the arousal level on trial $n+1$ was reset to 0. In the *linear* model, arousal on the first trial for a given participant was also set at 0. However, if the outcome of trial n was ‘calm’ then the arousal level on trial $n+1$ was the arousal level of trial n plus 1 arousal unit. Like the binary model, if the outcome of trial n was ‘emotional’ then the arousal level on trial $n+1$ was reset to 0. An example of the expectation function output for each model can be seen in table 16.

Trial	Outcome	Binary arousal	Linear arousal
1	“Calm”	0	0
2	“Calm”	1	1
3	“Calm”	1	2
4	“Emotional”	1	3
5	“Calm”	0	0
6	“Emotional”	1	1
7	“Emotional”	0	0
8	“Emotional”	0	0

Table 16. Illustration of the binary and linear models of expectation behaviour.

Bias calculations

Bias$_a$ and bias$_b$ were calculated according to the formulae presented above. For each iteration, values of bias$_a$ and bias$_b$ were obtained using mean arousal values calculated across all trials and across individual participant means. The overall bias estimates were then determined by averaging across all iterations (n = 5000). The *by trial* method of calculating mean arousal values simply summed the arousal levels across all trials in each condition and divided by the number of trials in each condition. The *by participant* method firstly calculated the mean arousal levels for each participant in each condition and then calculated the overall mean arousal levels across participant means.

Results and Discussion

The results of each series of simulations are presented in table 17. It can be seen that the bias$_a$ estimates are, on the whole, slightly larger than those reported in Dalkvist et al. (2002) and the standard deviations are considerably larger. This may be due to the smaller number of iterations performed in the current series of simulations (5000 per series) compared to the “about a million” that were reported to have been carried out in Dalkvist et al. (2002) or it may reflect differences in the random number generators used in each case. However, the results are qualitatively similar in that the size of bias_a produced by the *by participant* method was much larger than that produced by the *by trial* method of calculating mean arousal values. Similarly, the linear model produced a larger bias_a than the binary model, in line with the results from Dalkvist et al. (2002).

		M_C	M_E	$M_E - M_C$	Bias_a %	Bias_{Dalk} %	$M_{Pre\ C}$	$M_{Pre\ E}$	Bias_b %
By Part.	**Bin.**	0.469	0.501	0.032	6.67	6.39	1	0	3.22
		(0.032)	(0.032)	(0.044)	(9.18)	(1.26)	(0)	(0)	(4.43)
	Lin.	0.882	1.002	0.120	13.23	12.48	1.88	0	6.36
		(0.127)	(0.094)	(0.101)	(11.16)	(1.61)	(0.13)	(0)	(5.60)
By Trials	**Bin.**	0.483	0.486	0.003	0.59	-0.05	1	0	0.28
		(0.031)	(0.031)	(0.044)	(9.19)	(1.25)	(0)	(0)	(4.44)
	Lin.	0.933	0.943	0.010	1.78	0.008	1.93	0	0.86
		(0.139)	(0.083)	(0.116)	(12.14)	(1.77)	(0.14)	(0)	(5.88)

Table 17. Results of a series of simulated experiments with 16 participants each receiving 32 trials, with 5000 iterations per series. Binary and linear models are shown, with means calculated by participants and by trials. Bias_{Dalk} refers to the values published in Dalkvist et al. (2002) using the same model parameters except for the number of iterations. Standard deviations are in parentheses.

The results also show that when the *by participant* method of calculating means was used, the measure of bias_b was roughly half the size of bias_a. For example, for the linear model, the difference in arousal between ‘calm’ and ‘emotional’ trials is approximately 6.4% of the increase in arousal resulting from the modelled expectation behaviour. Thus, if one were to assume that the linear model is a good approximation of participants’ expectation behaviour in a real experiment, one might expect to find an expectation effect approximately 15 times larger than the observed presentiment effect within the same dataset if the observed presentiment effect was really due to an expectation bias of this kind. Moreover, if means are calculated by trials, the situation is more extreme since the values of bias_b are no greater than

1% of the modelled increase in arousal due to expectation. Dalkvist and Westerlund (2006) note that most published presentiment studies use the by-trial method of calculating mean pre-stimulus arousal values in each condition. Therefore, one might expect to find much larger expectation effects when examining the data from such studies if the results are due to an expectation bias. Perhaps the simplest way to detect expectation effects in previous studies would be to split the data by trials that were immediately preceded by calm versus emotional trials; the expectation bias hypothesis would predict greater pre-stimulus arousal on current trials that were preceded by calm trials compared to emotional trials, an effect that would hold regardless of whether the current trial was calm or emotional.

The results of Experiment 5A were therefore in general agreement with the data from Dalkvist et al. (2002) thereby replicating their results. In addition, bias_b estimates varied in a similar manner according to the particular model of expectation and the method of calculating mean arousal values. In Experiment 5B, a further series of simulations were therefore run in order to model the bias as it might apply to the current paradigm.

Experiment 5B: Simulating the expectation bias in the current paradigm

In a typical presentiment experiment, there is only one outcome variable per trial; a calm or emotion picture. However, there are two potential outcome variables for every trial in the current paradigm; the trial condition and the target position. Therefore, a model of expectation bias as applied to the current paradigm must take both of these outcome variables into account. Old trials provide participants with information about the array position occupied by the target item whereas new trials do not. One could therefore imagine a model of expectation whereby participants modify their viewing behaviour on the current trial based on knowledge of the position occupied by the target on the previous trial. For example, suppose that participants spend longer fixating the array position occupied by the preceding target if the preceding trial was in the old condition, and that viewing behaviour is unaffected if the preceding trial was in the new condition. This kind of model might be thought of as an ‘expectation effect’ but, clearly, a number of other perceptual and cognitive effects may form the basis of this model such as priming or procedural learning (Fecteau & Munoz, 2003). Whatever process underlies such sequential dependencies, a bias will emerge for the same reasons as described previously. However, the size of the bias when modelled in the context of the present paradigm remains to be established, in addition to its potential contribution to the observed results. Thus, several series of simulations were run based on binary and linear models of expectation involving experiments with 96 participants, each receiving 48 trials. These participant and trial parameters were the same as used in Experiment 4 and would therefore provide a reasonable comparison with the results of this experiment, which seemed to have obtained suggestive evidence for a precognitive effect. In addition, a random and balanced method of determining the trial condition and target position were used for each series of simulations as described below.

Methods

Participants and sample size

96 participants were run on each iteration of the simulation. Each participant received 48 trials. 5000 simulations were run per series.

Trial outcomes

Two outcomes were specified for each trial; the condition (‘old’ or ‘new’) and target position (1-4). In the *random* design, the condition of each trial and the target position were determined randomly with replacement (using the RANDBETWEEN function in Microsoft Excel 2007), and independently both within and between participants. This reproduced the randomisation conditions present in Experiments 3A and 3B. The *balanced* design used the same frequency and distribution of trial conditions and target positions as used in Experiment 4; each participant received the same number of trials in each condition and the target was specified in each array position an equal number of times both within participants and across an individual ‘experiment’. However, the sequence of trial conditions and target placements presented to participants was randomised for each successive iteration of the simulation (using the RAND function in Microsoft Excel 2007).

Models

Versions of the binary and linear models were used in the current series of simulations. The *binary* model was based on the following rules. The dependent variable (DV) was always set to 0 on trial 1 at each array position for every participant. If trial *n* was in the old condition, the DV on trial *n+1* was set at 1 unit on the array position occupied by the target on trial *n*. However, if trial *n* was in the new condition, the DV was reset to 0 at all array positions on trial n+1. These rules reflect the fact that old trials provide information about the previous target position but new trials do not. In the *linear* model, the increase in the DV was additive within a particular array position rather than being set at 1, in the same way as described previously. An example of the output from each expectation model is shown in table 18.

			Binary				Linear			
Trial	Condition	Target position	Pos. 1	Pos. 2	Pos. 3	Pos. 4	Pos. 1	Pos. 2	Pos. 3	Pos. 4
1	Old	1	0	0	0	0	0	0	0	0
2	Old	3	1	0	0	0	1	0	0	0
3	Old	3	0	0	1	0	0	0	1	0
4	New	n/a	0	0	1	0	0	0	2	0
5	Old	2	0	0	0	0	0	0	0	0
6	New	n/a	0	1	0	0	0	1	0	0
7	New	n/a	0	0	0	0	0	0	0	0
8	New	n/a	0	0	0	0	0	0	0	0

Table 18. An 8-trial example of the binary and linear models of expectation behaviour as applied to the current paradigm. Values of the dependent variable are shown for each array position (1-4).

Bias calculations

$Bias_a$ was a relative percentage measure defined as:

$$\text{Bias}_a = \frac{\text{Mean}_{\text{OldT}} - \text{Mean}_{\text{New}}}{\text{Mean}_{\text{All}}} \text{ X } 100$$

Here, $Mean_{OldT}$ is the overall mean value of the DV associated with target positions (i.e., referring to subscript 'T') in the old condition (i.e., referring to subscript 'Old') whereas $Mean_{New}$ is the overall mean value of the DV associated with all array positions in the new condition. $Mean_{All}$ is the overall mean value of the DV collapsed across both categories. Thus, for the binary modelled 8-trial example illustrated in table 18, bias_a would be calculated as follows:

$$\text{Bias}_a = \frac{(0 + 0 + 1 + 0) / 4 \quad - \quad (0.25 + 0.25 + 0 + 0) / 4}{(0 + 0.25 + 0.25 + 0.25 + 0 + 0.25 + 0 + 0) / 8} \text{ X } 100$$

$$= \quad 100\%$$

$Bias_b$ was also a relative percentage measure defined as:

$$\text{Bias}_b = \frac{\text{Mean}_{\text{OldT}} - \text{Mean}_{\text{New}}}{\text{Mean}_{\text{PreOldT}} - \text{Mean}_{\text{PreNew}}} \text{ X } 100$$

Here, $\text{Mean}_{\text{OldT}}$ and Mean_{New} are defined in the same way as in bias_a. However, $Mean_{PreOldT}$ is the overall mean value of the DV associated with array positions corresponding to the target position (i.e., referring to subscript 'T') on immediately preceding old trials (i.e., referring to subscript ‘PreOld’), whereas $Mean_{PreNew}$ is the overall mean value of the DV associated with all array positions on trials that were immediately preceded by a new trial. The term $Mean_{PreOldT} - Mean_{PreNew}$ thus refers to the mean increase in the DV as a result of the modelled expectation behaviour. The value for this term will be 1 in the binary model but will vary in the linear model as a function of the particular sequence of target positions generated in each simulated experiment. Thus, for the binary modelled 8-trial example illustrated in table 18, bias_b would be calculated as follows:

$$\text{Bias}_b = \frac{(0 + 0 + 1 + 0) / 4 \quad - \quad (0.25 + 0.25 + 0 + 0) / 4}{(1 + 1 + 1 + 1) / 4 \quad - \quad (0 + 0 + 0) / 3} \text{ X } 100$$

$$= 12.5\%$$

For each iteration of the simulation, values of bias_a and bias_b were obtained using mean DV values calculated across all trials and across individual participant means. The overall bias estimates were then determined by averaging across all iterations (n = 5000). The *by trial* method summed DV values across all trials and divided by the number of trials. The *by participant* method firstly calculated the mean DV values for each participant and then calculated the overall mean DV values across participant means.

Results and Discussion

The results of each simulation are presented in table 19. It should firstly be noted that the bias estimates are mostly negative. This is due to the fact that the model specified an increase in the DV on the relevant array position. If the model specified the reverse relationship, then a positive bias would have been generally produced.

When the outcome of each trial is randomly determined with replacement, the results are qualitatively similar to the results from Experiment 5A; calculating overall mean values across individual participant means produces a larger bias than calculating means over all trials. The bias estimates are also smaller than those reported in Experiment 5A. This is likely to be due to two factors. Firstly, more participants were run in each iteration (96 as opposed to 16) and each participant received more trials (48 as opposed to 32). The results are therefore consistent with the conclusions of Dalkvist et al. (2002) who note that the size of the bias decreases as the number of trials increases. Secondly, the number of possible outcomes on a given trial is greater in the current paradigm compared to a typical presentiment experiment; if there are more possible outcomes, the effects of expectation on the resulting bias are diminished. In particular, the estimates of bias_b are approximately 6 to 10 times less than those produced from the previous series of simulations.

However, when the trial outcomes are pre-determined and balanced across each experiment (but randomly sequenced for each iteration), the estimates of bias_a and bias_b are much larger. This is likely due to two factors. Firstly, there are statistical cues present as a result of the balancing. The modelled expectation behaviour, whereby the DV increases by one unit on the array position occupied by the target on the previous trial, has essentially the same effect as a response based on statistical cueing because, on average, the current target is less likely to occupy the same position as the target on the previous trial compared to a different position. Of course, the statistical advantage gained from basing current responses on the outcome of only the immediately preceding trial is much less than the advantage gained from basing current responses on all available information from the first trial, such as is the case with a cumulative count model described in Chapter 3. In other words, when an experiment uses a balanced design, ‘expectation’ behaviour is no longer fallacious and, in fact, can lead to an effect that reflects the processing of statistical cues embedded in the sequence of trials. Secondly, a computational bias arising from expectation will also contribute to the estimates of bias_a and bias_b in addition to the statistical cueing effect. Therefore, these two factors in combination may have contributed to the larger estimates in the balanced design. It is also noteworthy that, due to the balanced distribution of conditions and target positions, the by-participant and by-trial method of calculating means produce equivalent values for each variable in table 19.

			M_{OldT}	M_{New}	M_{OldT} – M_{New}	$Bias_a$ %	$M_{PreOldT}$	M_{PreNew}	$Bias_b$ %
Rand.	**By Part.**	**Bin.**	0.120 (0.007)	0.125 (0.003)	-0.005 (0.007)	-4.29 (5.85)	1 (0)	0 (0)	-0.51 (0.70)
		Lin.	0.136 (0.009)	0.143 (0.003)	-0.007 (0.009)	-5.15 (6.35)	1.14 (0.009)	0 (0)	-0.61 (0.76)
	By Trials	**Bin.**	0.123 (0.007)	0.122 (0.003)	0.001 (0.007)	-0.095 (5.99)	1 (0)	0 (0)	0.005 (0.73)
		Lin.	0.139 (0.009)	0.140 (0.003)	-0.0001 (0.009)	-0.25 (6.38)	1.14 (0.01)	0 (0)	-0.01 (0.78)
Bal.	**By Part.**	**Bin.**	0.104 (0.006)	0.125 (0.002)	-0.021 (0.007)	-18.16 (6.25)	1 (0)	0 (0)	-2.07 (0.67)
		Lin.	0.114 (0.008)	0.140 (0.002)	-0.026 (0.007)	-20.75 (6.31)	1.11 (0.01)	0 (0)	-2.34 (0.66)
	By Trials	**Bin.**	0.104 (0.006)	0.125 (0.002)	-0.021 (0.007)	-18.16 (6.25)	1 (0)	0 (0)	-2.07 (0.67)
		Lin.	0.114 (0.008)	0.140 (0.002)	-0.026 (0.007)	-20.75 (6.31)	1.11 (0.01)	0 (0)	-2.34 (0.66)

Table 19. Results of 5000 simulated experiments with 96 participants each receiving 48 trials. The method of determining trial outcome is shown as random or balanced. Binary and linear models are also shown, with means calculated by participants and by trials. Standard deviations are in parentheses.

Expectation Effects in Experiments 1 to 4

The series of simulations run in Experiment 5B show that a robust expectation bias arises whenever overall means are calculated across participant means and that the bias is largest when a balanced design is used. Nevertheless, the estimates of bias_b are still only 2% of the modelled increase in the dependent variable for balanced designs. The data from each experiment in the current thesis was therefore examined in order to establish whether expectation effects were present in each case and whether the size of the expectation effects could plausibly account for the positive results in Experiment 1 and Experiment 4. For each experiment, the TDT data was re-categorised in order to derive values of $\text{Mean}_{\text{PreOldT}}$ and $\text{Mean}_{\text{PreNew}}$ as these terms were defined above. TDT data was used because this was the only variable to have shown an apparent precognitive effect. The values of bias_b suggest that, as a conservative estimate, the difference between $\text{Mean}_{\text{PreOldT}}$ and $\text{Mean}_{\text{PreNew}}$ should be at least 20 times the size of the apparent precognitive effect if the expectation bias is to offer a plausible explanation for the positive results of Experiment 1 and 4. It would also be of interest to establish whether expectation effects occurred in the experiments that generated null results.

Table 20 shows the values of $\text{Mean}_{\text{OldT}}$, Mean_{New} (or $\text{Mean}_{\text{NewT}}$ for Experiment 1 and 2 since the comparison was between fixed target items in the old and new condition), $\text{Mean}_{\text{PreOldT}}$ and $\text{Mean}_{\text{PreNew}}$ for each experiment. Means were calculated across individual participant means. The data shows that the difference between $\text{Mean}_{\text{PreOldT}}$ and $\text{Mean}_{\text{PreNew}}$ was relatively small compared to the difference between $\text{Mean}_{\text{OldT}}$ and Mean_{New} in all experiments. A significant expectation effect was found in both the full and trimmed dataset of Experiment 3B suggesting that participants spent less time fixating the array position occupied by the target on the previous trial: all data, $t(49) = -2.60$, $p < 0.05$, $d = -0.50$; 10% trimmed data, $t(49) = -2.32$, $p < 0.05$, $d = -0.46$ (both tests Bonferroni corrected for two comparisons). However, no significant precognitive effect was found in Experiment 3B despite the presence of the expectation effect suggesting that an expectation effect of that magnitude was not sufficient to contribute to a significant expectation bias. Furthermore, no evidence for expectation effects was found in Experiments 1 and 4, both of which had obtained suggestive evidence for precognitive effects. Indeed, the difference between $\text{Mean}_{\text{PreOldT}}$ and $\text{Mean}_{\text{PreNew}}$ in both Experiments 1 and 4 was actually *less* than the difference between $\text{Mean}_{\text{OldT}}$ and Mean_{New}. In summary, there is no compelling evidence to suggest that the results of Experiment 1 or 4 can be explained by the effects of an expectation bias.

		M_{OldT}	M_{New}	$M_{OldT} – M_{New}$	$M_{PreOldT}$	M_{PreNew}	$M_{PreOldT} - M_{PreNew}$
Expt 1	All data	748.7 (80.2)	703.7 (95.8)	45.0*	771.5 (103.8)	776.2 (34.5)	-4.7
	10% trim	738.6 (68.8)	708.4 (78.6)	30.2*	763.5 (71.2)	762.1 (31.5)	1.4
Expt 2	All data	711.2 (93.1)	722.1 (87.7)	-10.9	749.9 (114.3)	762.2 (47.3)	-12.3
	10% trim	708.8 (72.0)	728.0 (77.5)	-19.2	742.6 (89.2)	745.7 (39.3)	-3.1
Expt 3A	All data	776.7 (93.5)	778.2 (46.9)	-1.5	787.0 (119.4)	778.4 (48.5)	8.6
	10% trim	760.9 (67.5)	759.4 (41.3)	1.5	771.2 (84.9)	761.7 (44.6)	9.5
Expt 3B	All data	803.2 (110.3)	791.8 (34.3)	11.4	759.5 (87.6)	794.3 (36.2)	-34.8*
	10% trim	778.5 (80.6)	778.6 (36.1)	-0.1	756.4 (64.5)	782.2 (43.2)	-25.8*
Expt 4	All data	756.7 (308.7)	771.3 (310.8)	-14.6	783.7 (320.2)	769.9 (312.2)	13.8
	10% trim	744.3 (234.3)	758.8 (236.5)	-14.5*	762.5 (239.5)	756.8 (236.3)	5.7

Table 20. An analysis of expectation effects in each of the experiments presented in the current thesis. * refers to effects with p-values less than 0.05.

Other studies have also failed to find expectation effects in presentiment data. Broughton (2004) ran a standard presentiment experiment measuring pre-stimulus skin conductance response to calm and emotional stimuli and failed to observe either a presentiment effect or any kind of expectation effect. Radin (2004) combined the data from two of his presentiment studies in order to establish whether the expectation bias may account for his positive results. However, rather than looking at the pre-stimulus values of the dependent variable (skin conductance level) on trials that were preceded by calm versus emotional trials, Radin tested for a correlation between the pre-stimulus skin conductance level and the number of successively preceding calm trials for a subset of participants pre-selected for displaying strong presentiment effects. The resulting correlation was not significantly positive (as would be predicted from the expectation hypothesis) and Radin concludes that the expectation bias is not a viable explanation for his results.

However, drawing firm conclusions from simple models of expectation is problematic for a number of reasons. Firstly, as Dalkvist et al. (2002) point out, there are various expectation strategies that participants may adopt, all of which may lead to some degree of bias. There may also be more complex expectation functions that have not been considered here which lead to larger biases. In other words, the results of simulations and the conclusions that are drawn from them are dependent on the starting assumptions and there is no guarantee that the assumptions are appropriate for a particular dataset from a ‘real’ experiment. Different participants may also be adopting different expectation strategies which may complicate matters further.

One could argue that, in general, it is better to run an experiment that rules out a particular bias by design rather than examining the data post-hoc in an effort to establish whether the bias had a significant effect on the results. This argument may be particularly relevant to parapsychological studies since most research hypotheses do not have a firm theoretical foundation. Based on the results on the above simulations, perhaps the best way to confidently rule out the influence of expectation effects is to randomise all experimental variables with replacement and calculate mean values across all trials. Unfortunately, randomisation with replacement may generate an unequal distribution of variables which may ultimately introduce unwanted noise into an experiment if an insufficient sample size is used. One also loses a certain degree of statistical power by averaging across all trials when a within-subjects analysis could be used instead. On the other hand, balancing variables across the experiment introduces statistical cues and results in larger estimates of bias_a and bias_b whether means are calculated across all trials or across individual participant means. However, the data presented here suggests that, in general, participants’ responses on current trials were not significantly influenced by the outcome of the previous trial and that when such behaviour did occur this did not lead to a bias in the comparison between conditions. Furthermore, when a balanced design was used in Experiment 4, the presence of statistical cues did not facilitate expectation effects. This suggests that the expectation bias does not represent a significant problem for interpreting the experimental results of the current thesis.

Chapter 5: General Discussion and Conclusions

Summary of the current series of experiments

Seemingly prophetic experiences have been reported throughout history and continue to be interpreted in terms of a paranormal process by many people. For example, surveys show that about 25% of the population believes in the ability to predict the future by paranormal means (Moore, 2005). Such a widespread belief, coupled with case collections of ostensible precognition, provide an incentive to examine such experiences in a scientific manner. One approach, termed anomalistic psychology (Zusne & Jones, 1982) examines whether such experiences can be explained by perceptual and cognitive biases that cause an individual to believe that a paranormal event may have taken place when, in fact, it has not (Blagrove, French & Jones, 2006; French, 2003; French & Wilson, 2006; Gilovich, Vallone, & Tversky, 1985; Pronin, Lin, & Ross, 2002). Another approach asks whether such experiences may represent the result of an underlying process that has not yet been described by current scientific models of perception and cognition. The latter approach has been the one adopted in the current thesis.

Numerous studies attempting to address this question in a controlled laboratory setting have reported evidence for apparently anomalous anticipatory responses to upcoming stimuli using a variety of methodologies. The dependent variable has comprised of behavioural measures such as forced choice and reaction time and physiological measures such as electrodermal activity, heart rate, brain electrical activity, BOLD response, pupil dilation and eye movements. The results from a number of these studies suggest that attentional processes may have been involved in the expression of the anticipatory effect (Don et al., 1998; McDonough et al., 2002; McRaty et al., 2004b; Warren et al., 1992a, 1992b). The current thesis was an attempt to build upon this research by testing for the presence of precognitive effects in the activity of the visual attention system using eye tracking techniques. Five empirical studies were carried out using a novel experimental paradigm designed for this purpose and an additional two studies modelled the potential impact of expectation artifacts on the data obtained from this paradigm. Each trial required participants to memorise an array of real-world objects while their eye movements were recorded and, after a brief retention interval, a probe stimulus was presented for a yes/no recognition test. Two conditions were employed and were randomly determined. In the old condition, the probe was an item viewed during study, termed the target. In the new condition, the probe was a novel item. Experiments tested for the presence of precognition by examining whether there was a selective response to target items during the study phase of old trials. Two dependent variables were considered; the total dwell time (TDT) defined as the total time spent fixating an item over the recording period and the first fixation time (FFT) defined as the time taken to fixate an array item for the first time.

Experiment 1 was designed as an initial test of the paradigm and obtained suggestive evidence in support of the precognitive hypothesis. Specifically, TDT was greater for target items when presented in the old condition compared to the new condition suggesting that targets held more attention when they were to be repeatedly presented in the future. However, a randomisation error resulted in a biased sequence of trials which brought the results of Experiment 1 into question. Nonetheless, no clear evidence for sequence learning could be found in the data, indicated by no increase in the effect over the course of individual participant sessions. Experiment 2 was an attempt to replicate the results of Experiment 1 using identical methods, except for a correction in the random assignment of conditions. Experiment 2 obtained null results, failing to replicate the apparent precognitive effect found in the initial study. A number of methodological changes were made in Experiments 3A and 3B. A more extensive randomisation procedure was implemented which involved the random selection of target items and their position in the array in order to further reduce the possibility of spurious effects. Experiment 3A was a further attempt to replicate the suggestive effect found in Experiment 1 but failed to find any evidence in support of the precognitive hypothesis. Experiment 3B was designed to explore the potential effect of encouraging participants to adopt a verbal encoding strategy during the study phase by presenting object names as probes in the recognition test. A significant effect was found in the FFT data but a trimming procedure nullified this result suggesting that it was likely caused by outliers. Experiment 4 introduced a balanced design in order to control for the confounding effects of differences in the salience of target items and their position in the array. The TDT data showed that participants spent less time fixating targets in the old condition compared to items in the new condition with the effect marginally failing to reach significance. However, the effect was significant after trimming the data to remove outliers and suggested that less attentional resources were allocated to targets. Experiment 4 also explored whether the apparent effect was related to normative variables associated with the experimental stimuli. The results of this exploratory analysis showed a weak and suggestive correlation between the apparent precognitive effect and both the familiarity and complexity ratings for each item. In other words, participants tended to spend less time fixating target items that were relatively more familiar and more visually complex. Finally, Experiment 5A and 5B examined whether an expectation bias could account for the results of the current thesis. Using a simple model of expectation in a series of simulated experiments of comparable sample size to Experiment 4, it was found that the resulting bias was no greater than 2.5% of the size of the expectation effect built into the simulation. Furthermore, examining the actual data from Experiments 1 to 4 did not reveal any compelling evidence to suggest that expectation effects were present in the studies that obtained positive results.

Methodological issues

Do the results of the current thesis, as a whole, provide evidence for the presence of precognitive effects in the activity of the visual attention system or are they more plausibly explained by known factors? Before this question is directly addressed, it may be helpful to discuss a number of related conceptual and methodological issues. Firstly, the analyses presented in the current thesis rely on the use of null hypothesis significance testing (NHST). As Cohen (1994) points out, the null hypothesis is often formulated as a ‘nil’ hypothesis stating that the means of two groups are identical. This is the case in the current thesis where the null hypothesis states that mean measures of TDT and FFT are identical in the old and new conditions. The use of NHST has been criticised on a number of occasions. For example, Cohen (1990) states,

> A little thought reveals a fact widely understood among statisticians: The null hypothesis, taken literally (and that's the only way you can take it in formal hypothesis testing), is always false in the real world. It can only be true in the bowels of a computer processor running a Monte Carlo study (and even then a stray electron may make it false). If it is false, even to a tiny degree, it must be the case that a large enough sample will produce a significant result and lead to its rejection. (p. 1308)

Similarly, in discussing the precognition studies carried out by Bem (2011), LeBel and Peters (2011) criticise Bem’s use of the null hypothesis on the basis that “it is almost by definition false” (p. 4) and argue that observing a statistically significant difference between two groups is then dependent on a large enough sample size that will reveal this underlying difference above the level of statistical noise. Accordingly, LeBel and Peters (2011) suggest that the small effect found in Bem's (2011) Experiment 2, reported as a significant hit rate of 51.7% where 50% was expected by chance, is an example of this inevitable principle. However, the question then arises; what processes could lead to the null hypothesis being false in Bem's study? In experimental tests of precognition, the null (or nil) hypothesis is usually formulated on the basis that experimental conditions are randomly assigned. Therefore, in order to claim that the 'null is always false' in this context, one must postulate that the random assignment of experimental conditions deviates from true randomness. Brugger and Taylor (2003) point out that all forms of random event generation will contain some degree of bias. Although this is undoubtedly true, the degree of bias present in most modern software-based pseudo-RNGs is very small. For example, all of the experiments described in the current thesis used the Mersenne Twister algorithm to generate random events. The period (the number of random events generated by an RNG before its deterministic output is repeated) of the Mersenne Twister is 2^{19937} (Matsumoto & Nishimura, 1998). [9] In experimental contexts, this clearly rules out the null being false on the basis of RNG periodicity. In addition, the Mersenne Twister passes several stringent tests of randomness (L'Ecuyer & Simard, 2007) including the Diehard battery of tests (Marsaglia, 1985). Likewise, Bem (2011) reports that the range of RNGs used in his experiments all pass the Diehard tests of randomness. It is therefore implausible that the extremely small bias present in the output of these RNGs can be held responsible for the results of experiments where the total number of randomisation events (such as the random assignment of conditions or other stimulus parameters) is relatively low. Arguments that appeal to the ubiquitous falsity of the null (or nil) hypothesis in purely experimental contexts, although logically sound, are thus of little practical significance when randomisation is shown to be adequate. Indeed, such arguments can be safely ignored. In addition, a number of authors argue that the claim of the null hypothesis always being false is meant to apply to non-experimental studies. For example, Cohen (1994) points out that,

> Most of the criticism of NHST in the literature has been for this special case where its use may be valid only for true experiments involving randomisation (e.g., controlled clinical trials) or when any departure from pure chance is meaningful (as in laboratory experiments on clairvoyance). (p. 1000)

---

[9] Although the Mersenne Twister was used to randomise the sequence of trials in Experiment 4, statistical cueing remains an issue for that particular experiment because a balanced design was employed.

Similarly, Hagen (1997) suggests that frequently cited commentaries on this issue such as Cohen (1990) and Meehl (1978) are often taken out of context and that Meehl (1990) was clear on stating that the ubiquitous falsity of the null hypothesis did not apply to “pure experimental studies” (p. 204). Indeed, LeBel and Peters (2011) even acknowledge that “it might be argued that the use of a null hypothesis of no difference is theoretically appropriate in Bem’s tests for precognition” (p. 4), an admission that seems to undermine their critique of Bem’s use of NHST in this respect.

A second related issue is that experiments purporting to be empirical tests of psi, including precognition, define the effect of interest somewhat negatively. With experimental tests of precognition, including the studies reported in the current thesis, the effect under test is defined as an *anomalous* correlation between responses in the present and the presentation of stimuli in the future. This definition means that in order for an observed result to qualify as a precognitive effect, known means by which such a correlation could come about have to be excluded; if it is assumed that such means have indeed been controlled for then a claim for an anomalous result can be made. One critical response to this general methodological rationale has been to argue that successfully ruling out known factors is problematic (Alcock, 2003; Hyman, 1994). Such concerns are bolstered by the fact that in certain cases, known factors have been identified *post-hoc* that could potentially explain positive results that were assumed to have been obtained under conditions where all known factors were ruled out (Dalkvist & Westerlund, 2006; Hyman, 1994). In other words, the extent to which known factors can be said to have been ruled out is dependent on the ingenuity of the experimenter or critic. Thus, whether or not the results of the current series of experiments can be interpreted as evidence for the precognitive hypothesis depends critically on these issues.

Are the positive results due to known factors?

Since the test stimuli in the current series of experiments were presented to participants at least one second after critical responses had been made and recorded, there was no possibility that sensory information from the test items was available to participants during the recording period. Such issues of sensory leakage are relevant to other categories of psi experiment where remote stimuli are presented contemporaneously, for example in ganzfeld telepathy designs, but they are not applicable to precognition designs where the remote stimulus is presented after the response has been recorded. Therefore, discussion of known factors in the current paradigm will be limited to consideration of chance processes and the possibility that participants were able to obtain statistical information that biased their responses to target items.

The first possibility that can be considered is that the null is true and that the significant results obtained from individual studies were false positives generated by random variation. At least two factors suggest that this may be the case. Firstly, the replication rate of the current paradigm was poor. Five experimental tests of the precognitive hypothesis were carried out in total, involving approximately 300 participants. Two studies obtained positive results that were marginally significant while the remaining three studies obtained null results. Secondly, the positive results from Experiment 1 and Experiment 4 were inconsistent. In Experiment 1, the results suggested that participants spent more time fixating targets yet in Experiment 4 the opposite was the case. Statistical noise would tend to produce low replicability and inconsistent, bidirectional false positive results. It should also be noted that significant results were only obtained in Experiment 4 when the data was trimmed post-hoc, an improvement that may also have been due to chance.[10] Some authors have used null results to argue that psi does not exist. For example, Moulton and Kosslyn (2008) ran an fMRI experiment attempting to test for selective brain activation to target stimuli and observed null results in the BOLD response data and in performance on their behavioural task. On the basis that no evidence for underlying brain activity in response to target stimuli was found, they conclude that psi does not exist. However, their conclusion may have been premature considering that they carried out only one study with relatively few participants. Furthermore, affirming the null is inherently problematic since a lack of positive findings can arguably be attributed to suboptimal testing conditions, especially if the claimed phenomenon does not have a firm theoretical foundation as is the case with psi effects in general.

---

[10] However, a Bonferroni correction for two comparisons was applied to each test of the precognitive hypothesis which can reduce the statistical power of the test and increase the likelihood of committing a type II error (Perneger, 1998). Uncorrected statistical tests would have rendered the untrimmed data significant.

Another possible interpretation of the results from the current thesis is that the positive findings were due to various forms of bias. Indeed, the only significant results from Experiment 1 and Experiment 4 were observed when bias was likely to have been present, specifically in the randomisation of trial conditions and target positions. In Experiment 1, the RNG used to randomise the sequence of old and new trials produced a ratio of old to new trials of approximately 4:3 when 1:1 was expected by chance. In Experiment 4, the sequence of trial conditions and target positions was expected to contain structure due to the balanced design employed in that particular study. Each participant received the same number of trials in the old and new condition and was also presented with the same number of target items in each array position. Thus, the probability of a target item appearing in a particular array position on a given trial was not held constant, as would be the case using a method of randomisation with replacement. In a balanced design of this sort, the probability of a particular outcome on a given trial is dependent on the outcome of previous trials. For example, there may be occasions where the probability that a given trial will be in the old condition is greater than the probability that the trial will be in the new condition. A similar situation is expected to arise for the probabilities associated with target position within the array.

There are a number of potential ways in which a bias in the sequence of trial conditions and target positions could have led to spurious results. Firstly, it is possible that participants' eye movements were influenced by implicit learning of non-random patterns embedded in the trial sequence. Brugger and Taylor (2003) put forward implicit sequence learning as a mechanism that may account for the results of many positive findings in parapsychology where subjects are given feedback on their performance from trial to trial. For example, they refer to experiments carried out by Tart (1976) who hypothesised that participants could learn to use ESP in a forced choice guessing task by responding to internal cues that were reinforced by immediate feedback. Some of the results from Tart's studies were highly significant but upon closer examination, Gatlin (1979) claimed to have identified significant non-random structure in the trial sequences from Tart's data that may have been learned by participants. Nevertheless, in the current thesis, no evidence for implicit sequence learning was found when the data from Experiment 1 was examined; the TDT effect did not increase over the course of individual sessions but, on the contrary, tended to decrease although this negative relationship was non-significant. Similarly, no evidence for implicit learning was found in Experiment 4 where the effect also decreased as each session progressed, although again this was not a significant decline. This provides evidence against the implicit learning hypothesis since it would predict an increase in the size of the effect over time as participants learn to respond to the structure embedded in the sequence of trials presented to them. It should also be noted that a previous psi study using a fully counterbalanced design failed to find a significant effect (Moulton & Kosslyn, 2008) despite this study involving a greater number of trials across which individual participants could have learned to respond to sequential bias compared to the present series of studies.

In addition to implicit sequence learning, one could argue that the non-random structure contained in the sequence of trial conditions and target positions from Experiment 1 and Experiment 4 just happened to match participants' inherent response biases. It is almost certain that participants will display response biases in any precognition study and the results of the present thesis were no exception. For example, participants tended to fixate items in a stereotypical clockwise manner with a greater degree of attention allocated to items located in the top left region of the array. If the assignment of experimental conditions and parameters is truly random then such biases will not inflate the rate of false positives above that expected by chance. However, if the sequence of trials contains an element of non-random structure, then the likelihood of response biases matching this patterning is no longer solely based on random variation. Such an account has been put forward to explain the results of past psi studies. For example, Colwell, Schroder and Sladen (2000) report an experiment on 'the sense of being stared at' in which a receiver made a binary choice on a given trial as to whether they were being looked at by a remote observer. A feedback and a no-feedback condition were employed and the results showed that participants performed better than chance when they were given trial-by-trial feedback. Crucially, subsequent analysis showed that there were more alternations in the sequence of stare and no-stare trials than expected by chance but only in the feedback condition. Brugger (1997) points out that participants who are asked to produce random sequences by purely subjective means also tend to make more alternations than expected by chance, and on this basis Colwell et al. (2000) conclude that their results may have been due to a fortuitous match between the non-random trial sequence and participants' response biases.

A number of simple models of the expectation bias were also explored in detail in Chapter 4 but it was found that the artifactual effects from such models were trivial at best when considered in the context of the current series of experiments. For example, in order for an expectation bias to plausibly account for the results of Experiment 4, an expectation effect of at least 20 times the magnitude of the observed results would have to have been present in the data. However, no evidence for an expectation effect was found for this experiment. Indeed, the one study that did reveal an expectation effect (Experiment 3B) did not show evidence for a precognitive effect suggesting that the presence of expectation effects of such magnitude were not sufficient to generate an artifactual result.

In summary, no compelling evidence was obtained suggesting that expectation biases or implicit sequence learning were responsible for the positive results obtained in the current thesis. However, more complex models of expectation not considered here may be able to account for such results. Perhaps the most likely explanation for the results of Experiment 1 and Experiment 4 in terms of known factors is that participant response biases fortuitously matched the bias present in the sequence of trial conditions and target positions generated in each experiment. Therefore, the null hypothesis cannot be entirely ruled out as an explanation for the results of the current thesis as a whole. Further replication attempts would be needed to resolve these issues. For example, successful replications would serve to lower the probability that the positive results obtained here were due to chance or a match between response bias and sequential bias. On other hand, if further replication attempts failed to reproduce the results of Experiment 1 or Experiment 4 then it would be more likely that the results from these studies were due to the known factors described above.

Exploring the precognitive hypothesis

The two-stage framework

Even though the null hypothesis cannot be adequately ruled out in the current series of experiments, it may be useful to explore ways in which the positive results may be explained if they do indeed represent genuine precognitive effects. The most popular way of conceptualising extra-sensory effects has been to suggest that they occur according to a two-stage process (Broughton, 2006; Irwin, 1979; Roll, 1966; Stanford 1974; Tyrell, 1946). In *stage 1*, a novel physical principle is thought to generate information about a normally inaccessible target event at some level, or levels, within an individual's neuropsychological system. Subsequently, *stage 2* is thought to involve established physical principles, as opposed to paranormal ones, whereby the information derived from stage 1 is processed by perceptual and cognitive systems appropriate to the locus, or loci, at which the information is established. Two-stage models generally assume that information arising from stage 1 is generated at a level below conscious awareness and that processing of this information during stage 2 may result in a conscious experience or some form of behavioural response.

A description of the range of models proposed as a basis for stage 1 is beyond the scope of this thesis. Nevertheless, it should be noted that recent theoretical developments in the physics of retro-causation (e.g., Cramer, 2006; Sheehan, 2006) may offer a plausible framework upon which to model stage 1 processes involved in precognition and, more generally, retro-causal effects in the brain. Indeed, retro-causation has been suggested as a candidate stage 1 mechanism to account for the results of many precognitive effects described in Chapter 2 (Radin, 2006). Similarly, Bierman (2010) notes that the inherent time-symmetry of physical formalisms allows for so-called retarded and advanced solutions that “are identical but reflected in time” (p. 278) and suggests that this may provide a theoretical basis for precognition. Although it is normally assumed that ‘time-reversed’ solutions to the fundamental equations of physics are in some way prohibited in the observed universe, Bierman (2010) speculates that consciousness can function to partially restore this inherent time-symmetry which may result in the appearance of time-reversed effects in cognition. However, Bierman is clear to point out that his theory does not allow for the future to influence the past, but rather that “at any moment a signal has determinants that are from past as well as from future boundary conditions” (p. 283). Precisely how consciousness is able to restore time-symmetry is not discussed, although he does suggest that the degree to which this restoration occurs is proportional to the degree of coherence in the activity of the brain and proposes a number of testable predictions from his theory.

Considering that such theorising is at a speculative stage and in the absence of a firm theoretical foundation for stage 1, theorists and experimentalists attempting to study potential stage 2 processes are forced to treat stage 1 as a necessary but unknown means by which information is generated within an individual’s neuropsychological system. Nevertheless, it is sometimes assumed that stage 1 at least involves some kind of information transfer from the environment to the percipient. For example, Bem (2011) defines psi as “anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms” (p. 407). Others are neutral with regard to the particular type of physical principle that may underlie stage 1. For example, Broughton (2006) stresses that although the character of extra-sensory experiences makes it appear as if information has been transferred from the environment to the percipient, it should not be assumed from the outset that stage 1 necessarily involves information transfer. Similarly, Irwin (1979) uses the term ‘psi input’ to refer to the initial information generated from stage 1 but emphasises that,

> ...psi input is defined to be *within* the system, and should not be regarded as an environmental stimulus. It may prove to have its origins in such as stimulus, or it may just ‘happen’ in the system, but in no way should use of the term *psi input* be taken in itself as an assumption of the former view in preference to the latter. (p.64)

Irwin’s model of psi functioning does, however, propose a locus for the generation of information from stage 1. Specifically, he suggests that an anomalous process results in the activation of target-relevant representations held in memory. Other theorists have also suggested that memory may play a central role in the generation of extra-sensory effects (Broughton, 2006; Roll, 1966; Stanford, 1974). According to such models, psi is thought to directly activate target-relevant memory representations in the absence of any preceding signals from sensory or top-down pathways. Upon activation, these memory representations are then subject to further processing as if they had been triggered by sensory or higher-level signals.

Irwin (1979) proposes that the most likely processing locus for stage 1 is the activation of memory representations coding for structural features of target stimuli. In support of this hypothesis, he cites evidence from a series of informal telepathic picture drawing experiments reported by Sinclair (1962) and Warcollier (1938) where ‘receivers’ were generally more successful at describing the structural features of target objects rather than identifying them by name or by describing semantic associations. For example, there were some occasions where the receiver accurately described the structural features of the target object but misidentified the target as a semantically unrelated object with similar physical characteristics. According to Irwin’s proposal, this may have occurred because the stage 1 process directly activated stored structural representations for objects that had similar physical characteristics to the target but were semantically unrelated to it, leading to misidentifications. However, the results of these informal experiments are also compatible with the assumption that stage 1 information is initially generated within low level sensory analysis pathways where the pattern of activation corresponds to the structural features of the target. These presumably weak sensory signals may then have been processed in such a manner as to activate representations from long term memory which were admitted into conscious awareness in the form of visual imagery. According to this account, the apparent misidentification of targets reported in Sinclair (1962) and Warcollier (1938) may have been due to weak psi-mediated signals from sensory pathways resulting in a relatively diffuse activation of stored structural descriptions with features common to a range of objects from different semantic categories. Irwin ultimately rejects such an account on the basis that activation of sensory pathways from stage 1 would predict a pattern of performance in psi tasks that is sensitive to parameters that affect normal sensory discriminability such as size, form or orientation, and claims that reviews of early card guessing experiments (Pratt, Rhine, Smith, Stuart & Greenwood, 1940) do not support this prediction. However, Broughton (2006) argues that relatively little information may be generated by stage 1 and that this may account for an apparent lack of sensory processing characteristics.

Therefore, although most theorists seem to agree that memory is likely to be involved in the expression of extra-sensory effects, it is unclear whether it is more appropriate to regard stage 1 as directly activating long term memories or whether the activation of target-relevant memory representations is an indirect result of stage 1 activity occurring elsewhere in the processing hierarchy. Regardless of which interpretation may turn out to be correct, two-stage models assume that the information generated from stage 1 is made available for further processing and integration with ongoing cognitive and perceptual processes during stage 2. The current thesis examined the activity of one such ongoing process, that of visual attention. One of the main functions of attention is to select activated representations for more detailed processing, including those from long term memory. The activity of attention is therefore consistent with the proposed function of stage 2 in the two-stage framework and is particularly compatible with memory models of psi where stored representations activated by stage 1 would be available for selection and enhanced processing.

Models of attentional selection

One influential theory of attention, put forward by Broadbent (1958), proposed that selection occurs by means of a filter that is applied to all incoming sensory information at early levels of analysis and functions to prevent further processing of unattended input on the basis of the input's representation of physical features. Input that does not match the filter's parameters is prevented from being processed any further and is lost through decay. However, inputs matching the parameters of the filter are allowed to pass and can gain access to a limited capacity, serial processing system that Broadbent considered necessary for conscious awareness. A modification of Broadbent's filter theory was proposed by Treisman (1964) who suggested that the filter acts to attenuate inputs rather than block them out completely. Broadbent later adapted his original theory to take Treisman's attenuator model into account (Broadbent, 1982). Deutsch and Deutsch (1963) proposed an alternative framework to filter theory which was motivated by evidence suggesting that complex information could pass through the filter on unattended channels (Moray, 1959). The model first supposes that all inputs achieve a high level of perceptual processing and are each given a weighting that represents that input's importance to the individual. Selection then acts on these signals according to their relative weights by means of a 'shifting reference standard'; the signal with the highest importance is the one that is selected. Deutsch and Deutsch's theory essentially allows selection to occur on the basis of a signal's meaning as well as its physical characteristics. Norman (1968) proposed a model of attention that emphasised the relationship between selection and memory. Like the model of Deutsch and Deutsch (1963), Norman's model proposed that all sensory inputs automatically activate representations held in long term memory before any selection occurs. Once representations have been activated from long term memory, they are assigned 'pertinence' values that are based on the nature of the input and its context. Processing of these representations then proceeds and their pertinence values can change according to the level at which they are processed. In Norman's model, selection is based on the representation with the highest pertinence value at any given time. Since pertinence values are continually being updated, selection can act on representations at various stages of processing. Attention is also regarded as an important component in models of working memory (Baddeley, 2007; Cowan, 1988, 1999). Cowan (1988, 1999) proposes an embedded process model of working memory whereby exogenous and endogenous inputs temporarily activate representations held in long term memory. A limited capacity focus of attention then selects representations from this temporarily activated pool and prevents their decay, thereby forming the basis of working memory content. Like in earlier models of attention, unattended representations are not subject to further processing and quickly decay.

Recent models of attention have described selection as a competitive process (Desimone & Duncan 1995; Knudsen, 2007). Knudsen (2007) presents a model of visual attention where, at any given moment, a pool of activated representations are available for selection, including those originating from perceptual analysis, long term memory, motor control and higher level cognition (see figure 16). Activated representations compete for selection at various levels of processing (Desimone & Duncan 1995) and, at any given level, representations with the greatest signal strength are selected and undergo further processing such as access to working memory and consciousness. Overt orienting of visual attention, via eye movements, is ultimately achieved by gaze control circuitry that directs the eyes to a portion of the visual field that has been selected for further processing. Gaze control circuitry receives input from both working memory and directly from the pool of activated representations that have won the competitive selection process. The former pathway corresponds to conscious and voluntary control of attention while the latter pathway represents involuntary and unconscious control. However, competitive selection of activated representations occurs in both cases. The signal strength of activated representations can be influenced by bottom-up or top-down pathways. In bottom-up pathways, filters are applied to incoming sensory information and function to increase the signal strength of representations encoding salient physical features of stimuli such as high contrast or spatial infrequency. In the current paradigm, array items that are physically salient are expected to attract attention and perhaps hold attention for longer periods on this basis. Top-down pathways involve higher-level signals that can bias the sensitivity of neural circuits underlying activated representations. In this way, activated representations that are important or relevant to an individual’s goals can be selected for further processing. In the current paradigm, top-down signals are likely to be involved in the selection of the next array item that requires encoding into short term memory once encoding of the currently fixated item has reached a threshold level.

What kind of information processing may underlie the selective response to targets?

If it is assumed that the positive results from the current thesis represent genuine precognitive effects, the model proposed by Knudsen (2007) can be used to speculate that during the study phase of old trials, an anomalous process generated target-relevant information somewhere in the processing scheme illustrated in figure 16 and that this information was subsequently able to affect gaze control. Lack of a theoretical basis for stage 1 makes it difficult to predict a precise locus for the initial generation of target information. However, one possibility is that information generated by stage 1 directly or indirectly influenced either the strength of activated representations coding for target-relevant features or influenced the ease with which such representations were activated during the study phase. The model proposed by Knudsen (2007) would predict that an increase in the activation strength of target-relevant representations would provide these representations with an advantage in the competitive selection processes which would more likely lead to an effect on overt attentional control.

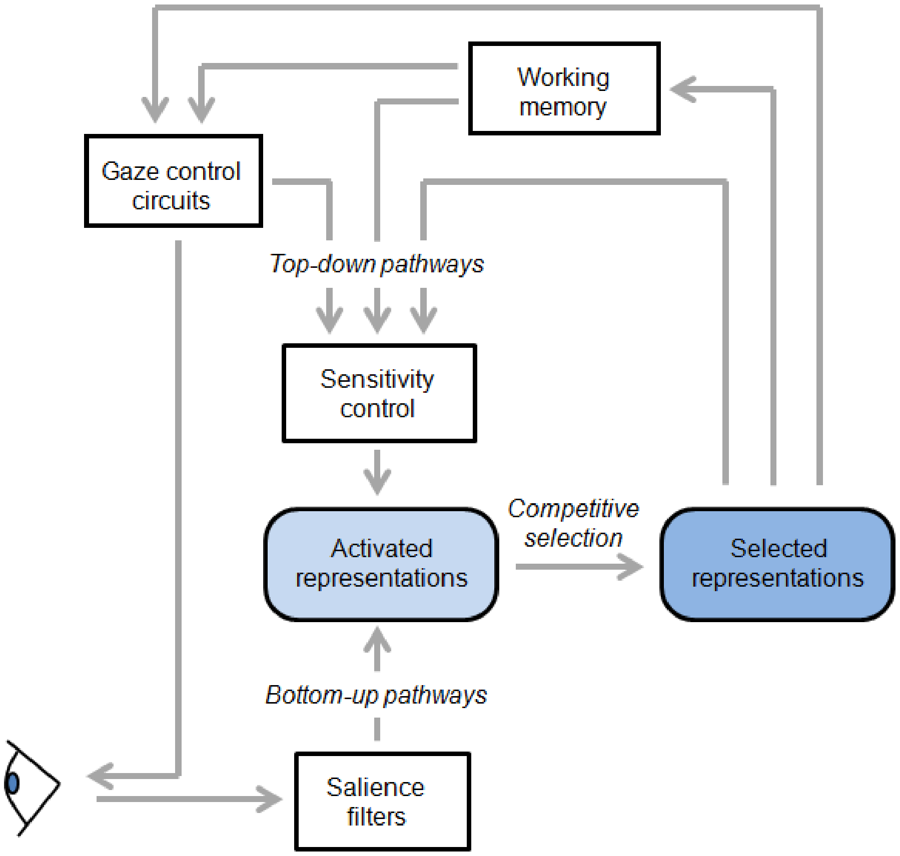


Figure 16. A model of overt visual attention (adapted from Knudsen (2007)). Bottom-up signals originating from sensory pathways and top down signals originating from higher level processes influence the strength of activated representations which then compete for selection. Once selected, representations can influence gaze control circuitry which determines the location and duration of eye fixations in the visual field. In addition, representations may enter working memory and feedback to sensitivity control which acts to bias the sensitivity of networks underlying currently activated representations.

Thus, stage 1 may have exerted an influence on attentional control by directly increasing the activation strength of target-relevant representations or by facilitating the activation of such representations, for example by lowering thresholds at which such representations become activated by signals from bottom-up or top-down pathways. In the attention and memory literature there is evidence that memory for previously encoded items can affect the allocation of overt attention to those items when encountered for a second time. Chanon and Hopfinger (2008) initially presented participants with images of real-world objects in a task involving deep encoding where they were asked to make semantic judgements about each object. Subsequently, scenes were presented that either contained old objects encountered during the previous encoding phase or new objects and participants were required to study the scene in preparation for a memory test. It was found that participants spent more time fixating old objects compared to new objects even when they were explicitly encouraged to memorise the scene by allocating more attention to objects that they had not seen before, suggesting that the increase in dwell time for old items was an unintentional behaviour. Chanon and Hopfinger (2008) explain this finding by suggesting that more information was retrieved when old objects were fixated which resulted in greater pertinence values being assigned to these items and this consequently affected attentional disengagement.

In Knudsen's (2007) scheme, this could be interpreted as an example of top-down bias signals affecting the sensitivity of the networks underlying activated representations for old objects, thus increasing their probability of being selected which subsequently influenced gaze control. In the current thesis, the results of Experiment 1 suggest that targets were fixated for longer in the old condition (notwithstanding problems with randomisation discussed previously). Perhaps then, this result was due to the generation of top down biasing signals that attached pertinence values to target-relevant representations which subsequently affected the disengagement of attention from target items in a similar manner to the effects reported in the Chanon and Hopfinger (2008) study.

However, this account is not consistent with the results of Experiment 4 which showed that less time was spent fixating targets in the old condition. This finding suggests that target stimuli may have been subject to more efficient processing during the study phase such that less time was needed to encode these items into short term memory. A multiple regression analysis on the data from Experiment 4 revealed a significant positive relationship between an item's total dwell time and normative data on the time taken to name the item, suggesting that participants were naming items as they viewed the array. Theories of picture naming propose the involvement of three main stages of processing; object recognition, object comprehension and lexicalisation (Barry et al., 1997; Humphreys, Price & Riddoch, 1999). Firstly, sensory input is thought to activate an object's stored structural description based on how well the sensory input matches the stored description. This corresponds to the object recognition stage. The stored structural description then makes contact with the object's associated semantic representations which allows the object to be comprehended. The final stage involves the use of semantic representations to activate information about the object's name. Thus, if less time was spent fixating targets in the old condition, stage 1 could have functioned to facilitate one or more of these processing stages. The multiple regression analysis showed that naming time was not a significant predictor of the precognitive effect, suggesting that the locus of stage 1 influence was not at the level of lexicalisation. However, there were suggestive relationships between the precognitive effect and both item familiarity and complexity. That is, less time was spent fixating target items as they became more familiar and more visually complex. This suggests that the precognitive effect perhaps had an influence at processing stages earlier than lexicalisation.

One possibility is that stage 1 exerted an influence at the stage of object recognition by decreasing recognition thresholds. There is considerable evidence that repeated stimuli are easier to perceive and recognise, an effect known as perceptual priming (Wiggs & Martin, 1998). This is typically observed as faster or more accurate performance on perceptual identification tasks when stimuli are repeatedly presented. There is also evidence that perceptual priming involves implicit memory and can be dissociated from declarative memory since amnesic patients who display impairments in episodic memory tasks nonetheless display robust perceptual priming effects (Hamann & Squire, 1997). Perhaps then, the precognitive effect observed in Experiment 4 may be considered an instance of 'retro-active' perceptual priming in that target items were processed more efficiently during the study phase of old trials compared to new trials, suggesting that their repeated presentation had a time-reversed effect. Experiments in cognitive neuroscience have also shown that the activity of individual neurons and cortical networks is reduced when a stimulus is repeatedly presented (Desimone, 1996; Henson, 2003). This phenomenon, termed repetition suppression, has been proposed as a neural basis for perceptual priming effects (Gotts, Chow & Martin, 2012; Wiggs & Martin, 1998) mainly on account of the similarities between the two types of effect and their co-occurrence. For example, both perceptual priming and repetition suppression are stimulus-specific, show cumulative effects with successive repetitions (Jiang, Haxby, Martin, Ungerleider, & Parasuraman, 2000; Miller, Gochin & Gross, 1991; Ostergaard, 1998; Wiggs, Martin & Sunderland, 1997) and are long lasting (Cave, 1997; van Turennout, Bielamowicz & Martin, 2003). Several neural-based theories have been proposed in an effort to link repetition suppression to the improved performance seen in perceptual priming experiments (see Gots et al., 2012, for a review). The *representational sharpening* model offers a potential mechanism that may be able to account for the suggestive relationship between the precognitive effect and item ratings of visual complexity observed in Experiment 4. This model states that upon initial visual presentation of an object, an activated neural network will contain a certain proportion of neurons coding for features essential for object recognition. These neurons will show a robust response while other neurons coding for non-essential features will show a weak response. Repeated presentation of the stimulus results in both a reduction in the activity of neurons coding for non-essential features in object recognition and a weakening of the connections to and from these neurons to other neurons in the network. The result is a 'sharpened' representation in that a greater proportion of neurons responding to the repeated presentation are those representing features essential for recognition of the object while those neurons that initially gave a weak response tend to drop out of the network (Desimone, 1996). Thus, although less overall neural activity is observed in repetition suppression, the sharpening model provides an explanation of this phenomenon in terms of a representation that is more selective which results in an improved behavioural response.

In order for the representational sharpening model to provide a basis for the relationship between the precognitive effect and item complexity, it must first be assumed that items rated as more visually complex tend to be comprised of a greater proportion of features not essential to object recognition compared to items rated as visually simple. This may be a reasonable assumption to make. For example, a picture of a butterfly from the Snodgrass and Vanderwart (1980) stimulus set is rated as more visually complex than a picture of an envelope (see figure 17). One could imagine that the butterfly can be recognised as such mainly from visual features corresponding to the outline of the wings and body. In other words, the detail contained in the wing pattern may be relatively unimportant in recognising that particular object as a butterfly. On the other hand, the pictorial representation of an envelope taken from the same stimulus set appears to be comprised mainly of essential features necessary to recognise the object as an envelope. If the precognitive effect in Experiment 4 was indeed an instance of 'retro-active' perceptual priming, objects rated as more visually complex may have tended to produce a larger precognitive effect because they were comprised of a greater proportion of features relatively unimportant in object recognition. In terms of repetition suppression, the proportion of weakly responding neurons dropping out of a network representing a complex object would tend to be larger than the proportion dropping out of a network representing a visually simple object. The degree of representational sharpening (or here, ‘retro-active’ representational sharpening) would therefore tend to be greater for complex objects, leading to a greater increase in the efficiency of object recognition and a greater precognitive effect if based on this level of processing. Of course, this argument assumes that objects from the Snodgrass and Vanderwart (1980) picture set that are rated as more visually complex are, in fact, comprised of a relatively greater number of redundant features in object recognition. This assumption, as far as the author of the current thesis is aware, has not been tested in any systematic way. Neither does there appear to be any published studies attempting to test whether there is a positive relationship between repetition priming and ratings of object complexity. Future studies aimed at testing these assumptions would therefore be useful and are discussed below.

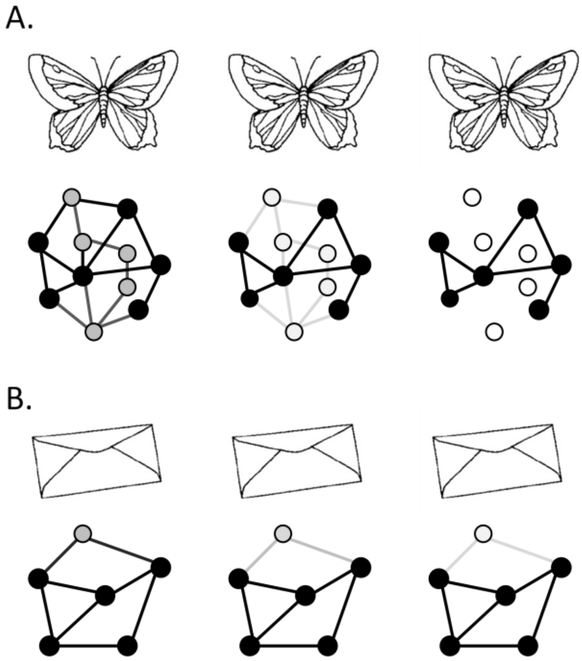


Figure 17. The representational sharpening model of repetition suppression (adapted from Wiggs & Martin, 1998). Black circles represent neurons coding for features essential to object recognition while grey circles represent neurons coding for non-essential features. Repeated presentation of stimuli results in a more selective (sharpened) representation. Hypothesised repetition suppression effects for a complex (A) and simple (B) pictorial object are shown.

The results of the multiple regression analysis on the data from Experiment 4 also suggest a positive relationship between the precognitive effect and ratings of object familiarity. This trend might be explained if familiar objects are semantically processed at a more rapid rate. In the original Snodgrass and Vanderwart (1980) study (and the Barry et al. (1997) study from which the familiarity norms used in the current thesis were obtained), participants were asked to rate each item for familiarity defined as ‘the degree to which you come in contact with or think about the concept’. This measure of familiarity has been suggested as a pictorial analogue of word frequency (Karlsen & Snodgrass, 2004). It is well established that high frequency words are recognised at a faster rate (Forster & Chambers, 1973; Scarborough, Cortese, & Scarborough, 1977). Borowsky and Besner (1993) propose a multistage model of visual word recognition involving the initial activation of representations in the orthographic input lexicon which then make contact with associated representations in the semantic system. According to their model, high frequency words have a more efficient mapping between the orthographic input lexicon and the semantic system resulting in a more rapid activation of semantic representations compared to low frequency words. In other words, a given level of input from the orthographic processing stage has a greater effect on the activation of semantic representations for high frequency words compared to low frequency words. Similarly, more efficient mappings may exist between stored structural descriptions of familiar pictorial objects and their associated semantic representations. If so, this may go some way to explaining why the precognitive effect observed in Experiment 4 was also related to ratings of familiarity. As discussed, the relationship between the precognitive effect and ratings of complexity suggest that the activation of structural descriptions may have been facilitated on the basis of a ‘retro-causal’ perceptual priming effect. These representations may then have made contact with the semantic system with the activation strength of semantic representations being dependent on item familiarity. Thus, a more rapid activation of the semantic system for familiar objects would, according to models of picture naming (Alario et al., 2004; Barry et al., 1997), result in a further reduction in the time taken to name these objects. Since it is likely that participants were sub-vocally naming objects in the current set of experiments, this in turn may have contributed to a greater reduction in the time spent fixating familiar items during the encoding phase compared to less familiar items.

Suggestions for further work

Replication attempts are critical in resolving whether the positive results observed in the current series of experiments are more likely to be explained by chance, known factors such as statistical cueing or genuine precognition. The accumulation of null results in replication attempts would suggest that the current results were due to statistical noise or a fortuitous match between the sequence of trial conditions and participants' response biases. On the other hand, positive replications would serve to strengthen the case for a real effect since it is less likely that a fortuitous match between sequential bias and response bias would occur on repeated replication attempts.

It is recommended that replications focus on Experiment 4 since two replication attempts of Experiment 1 have already been made in the current thesis and both failed to reject the null hypothesis. Strict replications of Experiment 4 that balance the frequency of experimental conditions and item positions will introduce statistical cues into the sequence of trials. Positive results from such replications would therefore be potentially explainable in such terms. However, the data could be examined for the presence of learning effects, as was done here. A cumulative lack of evidence for learning of sequential bias, indicated by the absence of an increase in effect size over the time course of individual participant sessions, would strengthen the case that the effect was not due to such factors. Alternatively, replications may employ randomisation with replacement on all experimental variables in order to rule out explanations based on statistical cueing. However, as discussed, this would introduce noise into the data due to unequal distributions of experimental conditions and item positions. One possible solution to this problem would be to implement randomisation with replacement as the experiment is run, but then randomly remove data *post-hoc* until the remaining dataset contained an equal number of items in each condition and array position for statistical analysis. This would avoid statistical cues being made available to participants and would reduce the amount of noise in the data. However, this technique would most likely result in discarding a large proportion of data and would therefore require a much larger number of participants than used in Experiment 4 in order to ensure that the remaining dataset had sufficient power to detect an effect.

It is also recommended that replication attempts restrict the precognitive hypothesis to planned analyses of TDT while ignoring FFT as it was evident from the results of this thesis that participants displayed a strong tendency to fixate items in a clockwise pattern. Any underlying precognitive effect would therefore need to be relatively strong if it were to influence that particular behaviour in a significant way. Clearly, if there is indeed a genuine precognitive effect in the data from the current thesis, it is small in magnitude. In addition, by limiting planned analyses to TDT, future studies could avoid using Bonferroni corrections for comparisons on both of these variables.

Alternative experimental paradigms may also be explored on the basis of the results from Experiment 4 where, in addition to a main effect, there were suggestive relationships between the precognitive effect and item ratings of complexity and familiarity. As discussed above, these findings suggest that the effect occurred at the level of object recognition and may therefore conform to models of perceptual priming. It is therefore recommended that further studies more directly examine this possibility. For example, a more appropriate paradigm in which to test for this effect may be one that reverses the order of study and test in a standard perceptual priming experiment. In a picture naming study, Mitchell and Brown (1988) presented participants with items from the Snodgrass and Vanderwart (1980) stimulus set and asked them to name each picture as fast as possible. One week later, participants were again asked to name pictures that were either old or new and it was found that old pictures were named approximately 70ms faster. A ‘time-reversed’ version of such an experiment might employ two randomly determined conditions; a repetition condition where naming latency is measured to the initial presentation of a picture and is followed by the repeated presentation of the same picture in a related task; and a no-repetition condition where a picture is presented and named only once. A ‘retro-active’ perceptual priming hypothesis would predict a decrease in the time taken to name pictures on their initial presentation in the repetition condition compared to the no-repetition condition. An experiment of this sort would have the advantage of being able to examine the apparent effect observed in Experiment 4 more directly since the results here suggest the involvement of processes preceding that of the lexicalisation stage of picture naming such as object recognition and semantic activation. However, for theoretical reasons, it is not clear what the optimum inter-stimulus interval should be in such an experiment. Intuitively, a relatively short interval, perhaps on the order of seconds, between first and repeated presentation would seem to offer the best conditions under which observe a ‘retro-active’ priming effect since forward priming effects have been reported to decline somewhat over longer periods of time (Cave, 1997). Considering that the meta-analysis on forced choice precognition (Honorton & Ferrari, 1989) found the largest effect size for studies that selected targets milliseconds after a response had been made, this may be a reasonable assumption to make.

In addition, future work using a standard perceptual priming paradigm (i.e., with cause and effect in the normal temporal order) might investigate whether more complex and more familiar pictures tend to generate a greater perceptual priming effect since, as far as the current author is aware, no such studies have yet been performed. The representational sharpening hypothesis outlined above might predict that objects rated as more visually complex would generate a greater perceptual priming effect because they comprise of relatively more visual features that are not essential for recognising the object compared to less complex objects. Similarly, more familiar objects might generate a greater priming effect because the more efficient mapping between the structural and semantic representations for these objects results in faster processing compared to less familiar objects. One strategy to test for the complexity relationship might be to use item ratings to control for familiarity while creating high complexity and low complexity stimulus groups from the Snodgrass and Vanderwart (1980) stimulus set. Alternatively, pictures could be edited, for example by removing intricate detail, to form ‘complex’ and ‘simple’ versions of the same objects allowing for a within-items analysis that would control for familiarity effects (see figure 18). Similarly, to isolate the effect of familiarity, complexity could be held constant while manipulating levels of familiarity using available norms.

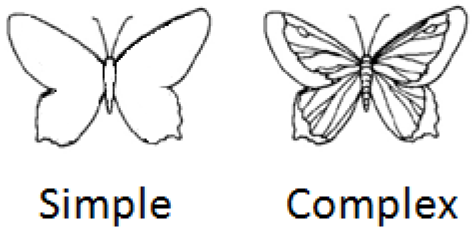


Figure 18. Potential method to manipulate object complexity while controlling for familiarity in tests of ‘forward’ and ‘retro-active’ perceptual priming.

In addition to running this ‘forward’ priming experiment, it would be of interest to observe whether a similar relationship between a ‘retro-active’ priming effect and ratings of item complexity and familiarity emerge, as suggested from the results of Experiment 4 of the current thesis. Confirmation of these relationships in both a ‘forward’ priming and ‘retro-active’ version of the experiment would not only contribute to the priming and repetition suppression literature but would also conceptually replicate the results observed in Experiment 4.

Closing remarks

It is unclear whether the findings of the current thesis offer support for the precognitive hypothesis or whether known factors such as chance or statistical cueing can offer a more plausible explanation for the results. As suggested above, replication attempts may be able to resolve this question. However, if the null is indeed false in the current paradigm then the observed effect appears to suffer from poor replicability. Therefore, further work aimed at testing for precognition should also more directly examine the involvement of potential processing loci suggested by the results of the current thesis. An avenue of inquiry has been suggested to test for potential 'retro-causal' perceptual priming effects and it has been put forward that more visually complex and more familiar objects may produce stronger effects in such a paradigm. Indeed, if precognition represents a genuinely anomalous means to anticipate future events, the concept of 'retro-causality' may currently offer the most useful framework for making sense of these experiences and laboratory phenomena. Perhaps the clearest prediction from the 'retro-causal' account is that precognitive effects should occur during a wide range of perceptual and cognitive task conditions. This appears to be borne out in the range of laboratory methods and task conditions in which precognitive effects have been reported to occur as described in Chapter 2. However, a significant challenge for parapsychology remains in offering an experimental method that is able to produce a robust and replicable precognitive effect. If this challenge can be overcome, the next step would be to establish the limits of this apparent 'retro-causal' influence on perception, cognition and affect.

References

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, 36, 140-155.

Alcock, J. E. (2003). Give the null hypothesis a chance: Reasons to remain doubtful about the existence of psi. In Alcock, J. E., Burns, J., & Freeman, A. (Eds.), *Psi Wars: Getting to Grips with the Paranormal*, (pp. 29-50). Charlottes-ville, VA: Imprint Academic.

Alcock, J. E. (2011, January 6). Back from the future: Parapsychology and the Bem affair. *Skeptical Inquirer*. Retrieved from http://www.csicop.org/specialarticles/show/back_from_the_future

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology*, 53, 221-227.

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, AD. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency and name agreement. *Quarterly Journal of Experimental Psychology*, 50A, 560-585.

Batthyany, A. (2010). Retrocausal habituation and induction of boredom: A successful replication of Bem (2010; Studies 5 and 7). Available at SSRN: http://ssrn.com/abstract=1715954 or http://dx.doi.org/10.2139/ssrn.1715954.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 80, 1-46.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275, 1293-1295.

Bem, D. J. (2003). Precognitive habituation: Replicable evidence for a process of anomalous cognition. In S. Wilson (Chair), *Proceedings of presented papers from the 46th annual convention of the Parapsychology Association*, 2003 (pp. 6-20), Vancouver, Canada.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 1-19.

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, 115, 4-18.

Bem, D. J., Palmer, J., & Broughton, R. S. (2001). Updating the ganzfeld database: A victim of its own success? *Journal of Parapsychology,* 65, 207-218.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, 20, 585-593.

Bierman, D. J. (1988). A test on possible implications of the OT's for ganzfeld research. *European Journal of Parapsychology*, 7, 1-12.

Bierman, D. J. (2010). Consciousness induced restoration of time symmetry (CIRTS): A psychophysical theoretical perspective. Journal of Parapsychology, 74, 273-299.

Bierman, D. J., & Radin, D. I. (1997). Anomalous anticipatory response on randomized future conditions. *Perceptual and Motor Skills*, 84, 689-690.

Bierman, D. J., & Scholte, H. S. (2002, August). *Anomalous anticipatory brain activation preceding exposure of emotional and neutral pictures*. Paper presented at the meeting of the Parapsychological Association, Paris, France.

Blagrove, M., French, C.C., & Jones, G. (2006). Probabilistic reasoning, affirmative bias and belief in precognitive dreams. *Applied Cognitive Psychology*, 20, 65-83.

Blakely, D. P., Wright, T., Dehili, V. M., Boot, W. R., & Brockmole, J. R. (2012). Characterizing the time course and nature of attentional disengagement effects. *Vision Research*, 56, 38-48.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106, 265-289.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 813-840.

Braud, W.G., Wood, R., & Braud, L.W. (1975). Free-response GESP performance during an experimental hypnagogic state induced by visual and acoustic ganzfeld techniques: a replication and extension. *Journal of American Society for Psychic Research*, 69, 105-113.

Broadbent, D. (1958). *Perception and Communication*. London: Pergamon Press.

Broadbent, D. E. (1982). Task combination and selective intake of information. *Acta Psychologia*, 50, 253-290.

Broughton, R. S., (2004). Exploring the reliability of the “presentiment” effect. *Proceedings of presented papers from the 47th annual convention of the Parapsychology Association*, (pp. 15-26), Vienna.

Broughton, R. S., (2006). Memory, emotion and the receptive psi process. *Journal of Parapsychology*, 70, 255-274.

Brugger, P. (1997). Variables that influence the generation of random sequences: an update. *Perceptual and Motor Skills*, 84, 627-661.

Brugger, P., & Taylor, K. I. (2003). ESP: extrasensory perception or effect of subjective probability? *Journal of Consciousness Studies*, 10(6-7), 221-246.

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science*, 8, 322-325.

Cave, C. B., Bost, P. R., & Cobb, R. E. (1996). Effects of color and pattern on implicit and explicit picture memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 22, 639-653.

Camfferman, G. (1987). Time reversed interference: A replication study. *European Journal of Parapsychology*, 7, 13-31.

Castelhano, M. S., Mack, M. L., Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):6, 1-15, http://www.journalofvision.org/content/9/3/6, doi:10.1167/9.3.6.

Chanon, V. W., & Hopfinger, J. B. *(*2008*).* Memory's grip on attention: The influence of item memory on the allocation of attention. *Visual Cognition*, 16, 325-340.

Child, I. (1985). Psychology and anomalous observations: The question of ESP in dreams. *American Psychologist*, 40, 1219-1230.

Cho, R., Nystrom, L., Brown, E., Jones, A., Braver, T., Holmes, P., & Cohen, J. D. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, Affective and Behavioral Neuroscience*, 2, 283-299.

Christie, J., & Klein, R. (1995). Familiarity and attention: Does what we know affect what we notice? *Memory and Cognition*, 2, 547-550.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304–1312.

Cohen, J. (1994). The earth is round (p < 05). *American Psychologist*, 49, 997–1003.

Colwell, J., Schroder, S., & Sladen, D. (2000). The ability to detect unseen staring: a literature review and empirical tests. *British Journal of Psychology*, 9, 71-85.

Cook, T. D., & Groom, C. (2004). The methodological assumptions of social psychology: The mutual dependence of substantive theory and method choice. In C. Sansone, C. C. Morf, & A. T. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 19–44). Thousand Oaks, CA: Sage.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, 104, 163-191.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge, U.K.: Cambridge University Press.

Cramer, J. G. (2006). Reverse causation and the transactional interpretation of quantum mechanics. In Sheehan, D. P. (Ed), *Frontiers of time: Retrocausation – experiment and theory*, AIP Conference Proceedings for 87th Meeting of AAAS Pacific Division, University of San Diego.

Dalkvist, J., & Westerlund, J. (2006). A bias caused by inappropriate averaging in experiments with randomized stimuli. *Journal of Parapsychology*, 70, 233-254.

Dalkvist, J., Westerlund, J., & Bierman, D. J. (2002). A computational expectation bias as revealed by simulations of presentiment experiments. *Paper presented at the Proceedings of the 45th Annual Convention of the Parapsychological Association*, Cary, NC.

Damian, M. F. (2000). Semantic negative priming in picture categorization and naming. *Cognition*, 76, B45-B55.

Davis, J.W. (1978). The stacking effect: Its practical significance in parapsychology. *Journal of Parapsychology*, 42, 67.

De Jong, H. L., Kok, A., & van Rooy, J. C. G. M. (1988). Early and late selection in young and old adults: An event-related potential study. *Psychophysiology*, 25, 657-671.

Dell'Acqua, R., & Job, R. (1998). Is object recognition automatic? Psychonomic Bulletin & Review, 5, 496-503.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 13494–13499.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193-222.

Destrebecqz, A. and Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin and Review*, 8, 343-50.

Deutsch, J. A., & Deutsch, D. (1963). Attention, some theoretical considerations. *Psychological Review*, 70, 80-90.

Dishon-Berkovits, M., & Algom, D. (2000). The Stroop effect: It is not the robust phenomenon that you have thought it to be. *Memory & Cognition*, 28(8), 1437-1449.

Don, N. S., McDonough, B. E., & Warren, C. A. (1998). Event-related brain potential (ERP) indicators of unconscious psi: A replication using subjects unselected for psi. *Journal of Parapsychology*, 62, 127-145.

Duc, A. H., Bays, P., & Husain, M. (2008). Eye movements as a probe of attention. In Kennard, C., & Leigh, R. J. (Eds.), *Progress in Brain Research*, 171 (pp. 403-411). Elsevier.

Dux, P. E., & Harris, I. M. (2007). Viewpoint costs occur during consolidation: Evidence from the attentional blink. *Cognition*, 104, 47-58.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 24, 515-523.

Emeric, E. E., Brown, J. W., Boucher, L., Carpenter, R. H. S., Hanes, D. P., Harris, R., . . . Schall, J. D. (2007). Influence of history on saccade countermanding performance in humans and macaque monkeys. *Vision Research*, 47, 35-49.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, 40, 225-240.

Fecteau, J. H., & Munoz, D. P. (2003). Exploring the consequences of the previous trial. *Nature Reviews Neuroscience*, 4, 435-443.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behaviour,* 12, 627-635.

French, C. C. (2003). Fantastic memories: The relevance of research into eyewitness testimony and false memories for reports of anomalous experiences. *Journal of Consciousness Studies*, 10, 153-174.

French, C. C., & Wilson, K. (2006). Incredible memories —how accurate are reports of anomalous events? *European Journal of Parapsychology*, 21(2), 166-181.

Gatlin, L. L. (1979). A new measure of bias in finite sequences with application to ESP data. *Journal of the American Society for Psychical Research*, 73, 29-43.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17, 295-314.

Gotts, S. J., Chow, C. C, & Martin, A. (2012). Repetition priming and repetition suppression: Multiple mechanisms in need of testing, Cognitive Neuroscience, 3(3-4), 250-259.

Gurney, E., Myers, F. W. H., & Podmore, F. (1886). *Phantasms of the Living. Vol. I*. London: Trubner.

Green, C. E. (1960). Analysis of spontaneous cases. *Proceedings of the Society for Psychical Research*, 53, 97-161.

Hagen, R. L. (1997). In Praise of the Null Hypothesis Statistical Test. *American Psychologist*, 52, 15-24.

Hamann, S. B., & Squire, L. R. (1997). Intact perceptual memory in the absence of conscious memory. *Behavioural Neuroscience*, 111, 850-854.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7, 498-504.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In van Gompel, R. P. G., Fischer, M. H., Murray, W. S., & Hill, R. L. (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford, UK: Elsevier.

Henson, R. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, 70(1), 53-81.

Hirsh, K. W., & Funnell, E. (1995). Those old, familiar things: Age of acquisition, familiarity and lexical access in progressive aphasia. *Journal of Neurolinguistics*, 9, 23-32.

Honorton, C., & Harper, S. (1974). Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *Journal of the American Society for Psychical Research*, 68, 156-168.

Honorton, C. (1977). Psi and internal attention states. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 435-472). New York: Van Nostrand Reinhold.

Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology*, 49, 51-91.

Honorton, C. (1987). Precognition and real time ESP performance in a computer task with an exceptional subject. *Journal of Parapsychology*, 51, 291-321.

Honorton, C., & Ferrari, D. C. (1989). “Future telling”: A meta-analysis of forced-choice precognition experiments, 1935–1987. *Journal of Parapsychology*, 53, 281-308.

Houtkooper, J. M. (2002). Arguing for an observational theory of paranormal phenomena. *Journal of Scientific Exploration*, 16, 171-185.

Humphreys, G. W., Price, C. J., & Riddoch, M. J. (1999). From objects to names: a cognitive neuroscience approach. *Psychological Research,* 62, 118-130.

Hyman, R. (1994). Anomaly or artifact? Comments on Bem and Honorton. *Psychological Bulletin*, 115, 19-24.

Hyman, R. (2010). Meta-analysis that conceals more than it reveals: Comment on Storm et al. (2010). *Psychological Bulletin*, 136, 486-490.

Irwin, H. J. (1979). *Psi and the mind*. Metuchen, NJ: Scarecrow Press.

Itti, L. (2006). Quantitative modelling of perceptual salience at human eye position. *Visual Cognition*, 14, 959-984.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489-1506.

Jacoby, L. L., & Hayman, C. A. G. (1987). Specific visual transfer in word identification. Journal of Experimental Psychology: *Learning, Memory & Cognition*, 13, 456-463.

Jiang, Y., Haxby, J. V., Martin, A., Ungerleider, L. G., & Parasuraman, R. (2000). Complementary neural mechanisms for tracking items in human working memory. *Science*, 287, 643-646.

Johnston, W. A., Hawley, K. J., & Farnham, J. M. (1993). Novel popout: Empirical boundaries and tentative theory. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 140-153.

Karlsen, P. J. & Snodgrass, J. G. (2004). The word-frequency paradox for recall/recognition occurs for pictures. *Psychological Research*, 68, 271-276.

Klintman, H. (1983). Is there a paranormal (precognitive) influence in certain types of perceptual sequences? Part I. *European Journal of Parapsychology*, 5, 19-49.

Klintman, H. (1984). Is there a paranormal (precognitive) influence in certain types of perceptual sequences? Part II. *European Journal of Parapsychology*, 5, 125-140.

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience*, 30, 57-78.

Kohler, S., Moscovitch, M., & Melo, B. (2001). Episodic memory for object location versus episodic memory for object identity: Do they rely on distinct encoding processes? *Memory & Cognition*, 29, 948-959.

Krippner, S., Honorton, C., & Ullman, M. (1971). A precognitive dream study with a single subject. *Journal of the American Society for Psychical Research*, 65, 192-203.

Krippner, S., Honorton, C., & Ullman, M. (1972). A second precognitive dream study with Malcolm Bessent. *Journal of the American Society for Psychical Research*, 66, 269-279.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1995). *International Affective Picture System (IAPS): Technical manual and affective ratings*. Gainesville, FL: Center for Research in Psychophysiology, University of Florida.

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, 15, 371-379.

L' Ecuyer, P. & Simard, R. (2007). TestU01: A C library for empirical testing of random number generators. *ACM Transactions on Mathematical Software*, 33, 22-40.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 13464-13471.

Logie, R. H. (1986). Visuo-spatial processes in working memory. *Quarterly Journal of Experimental Psychology*, 38A, 229-247.

Luke, D. P., Delanoy, D., & Sherwood, S. J. (2008). Psi may look like luck: Perceived luckiness and beliefs about luck in relation to precognition. *Journal of the Society for Psychical Research*, 72(4), 193-207.

Luke, D. P., Roe, C., & Davison, J. (2008). Testing for forced-choice precognition using a hidden task: Two replications. *Journal of Parapsychology*, 72, 133-154.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163-203.

Maljkovic, V., & Nakayama, K. (1996). Priming of popout: II. Role of position. *Perception and Psychophysics*, 58, 977-991.

Maljkovic, V., & Nakayama, K. (2000). Priming of pop-out: III. A short-term implicit memory system beneficial for rapid target selection. *Visual Cognition*, 7, 571-595.

Marsaglia, G. (1985). *A current view of random numbers*. In Computer Science and Statistics, Proceedings of the Sixteenth Symposium on The Interface, North-Holland, Amsterdam,
3-10.

Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modelling and Computer Simulation*, 8(1), 3-30.

May, E. C., Paulinyi, T., & Vassy, Z. (2005). Anomalous anticipatory skin conductance response to acoustic stimuli: Experimental results and speculation about a mechanism. *Journal of Alternative and Complementary Medicine*, 11, 695-702.

McCallum, W. C., Barrett, K., & Pocock, P. V. (1989). Late components of auditory event-related potentials to eight equiprobable stimuli in a target detection task. *Psychophysiology*, 26, 683-694.

McCraty, R., Atkinson, M., & Bradley, R.T. (2004a). Electrophysiological evidence of intuition: Part 1. The surprising role of the heart. *Journal of Alternative and Complementary Medicine*, 10, 133-143.

McCraty, R., Atkinson, M., & Bradley, R.T. (2004b). Electrophysiological evidence of intuition: Part 2. A system-wide process? *Journal of Alternative and Complementary Medicine*, 10, 325-336.

McDonough, B., Don, N., & Warren, C. (2002). Differential event related potentials in targets and decoys in guessing task. *Journal of Parapsychology*, 16, 187-206.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195-244.

Miller, E. K., Gochin, P. M., & Gross, C. G. (1991). Habituation-like decrease in the responses of neurons in inferior temporal cortex of the macaque. *Visual Neuroscience*, 7, 357-362.

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin*, 125(4), 387-391.

Mitchell, D. B. (2006). Nonconscious priming after 17 years. *Psychological Science*, 17, 925-929.

Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 213-222.

Moore, D. (2005). Three in four Americans believe in paranormal. *Princeton: Gallup News Service*, retrieved 20th Aug 2012, from http://www.gallup.com/poll/16915/three-four-americans-believe-paranormal.aspx.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11, 56-60.

Mossbridge, J., Tressoldi, P. E., & Utts, J. (2011). *Physiological anticipation of unpredictable stimuli, a meta-analysis*. Manuscript submitted for publication.

Moulton, S. T., & Kosslyn, S. M. (2008). Using neuroimaging to resolve the psi debate. *Journal of Cognitive Neuroscience*, 20, 182-192.

Näätänen, R. (1982). The N2 component of the evoked potential: A scalp reflection of neuronal mismatch of orienting theory. In Strelau, J., Farley, F., & Gale, A. (Eds.), *Biological Foundations of Personality and Behavior*. Washington: Hemisphere Press.

Norman, D. A. (1968). Towards a theory of memory and attention. *Psychological Review*, 75, 522-536.

Orme, J. (1974). Precognition and time. *Journal of the Society for Psychical Research*, 47, 351-365.

Ostergaard, A. L. (1998). The effects on priming of word frequency, number of repetitions, and delay depend on the magnitude of priming. *Memory & Cognition*, 26, 40-60.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 107-123.

Pashler, H., & Harris, C. (2001). Spontaneous allocation of visual attention: Dominant role of uniqueness. *Psychonomic Bulletin and Review*, 8, 747-752.

Paivio, A., Clark, J. M., Digdon, N., & Bons, T. (1989). Referential processing: Reciprocity and correlates of naming and imaging. *Memory & Cognition*, 17, 163-174.

Palmer, J. (2006). Memory and ESP – A review of the experimental literature. *European Journal of Parapsychology*, 21, 95-121.

Perneger, T. V. (1998). What’s wrong with Bonferroni adjustments. *British Medical Journal*, 316, 1236-1238.

Posner, M. I. (1980). Orienting of visual attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25.

Pratt, J. G., Rhine, J. B., Smith, B.M., Stuart, C.E., & Greenwood, J. A. (1940). *Extra-sensory perception after sixty years: A critical appraisal of the research in extra-sensory perception*. New York: Henry Holt.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28, 369-381.

Radin, D. I. (1997a). *The Conscious Universe*. HarperCollins

Radin, D. I. (1997b). Unconscious perception of future emotions: An experiment in presentiment. *Journal of Scientific Exploration*, 11, 163-180.

Radin, D. I. (2004). Electrodermal presentiments of future emotions. *Journal of Scientific Exploration*, 18, 253-273.

Radin, D. I. (2006). Psychophysiological evidence of possible retrocausal effects in humans. In Sheehan, D. P. (Ed), *Frontiers of time: Retrocausation – experiment and theory*, AIP Conference Proceedings for 87th Meeting of AAAS Pacific Division, University of San Diego.

Radin, D. I., & Borges, A. (2009). Intuition through time: What does the seer see? *Explore*, 5, 200-211

Radin, D., Lobach, E. (2007). Toward understanding the placebo effect: investigating a possible retrocausal factor. *Journal of Alternative and Complementary Medicine*, 13, 733-739.

Radin, D. I., & May, E. C. (2001). Evidence for a retrocausal effect in the human nervous system. In, *Proceedings of Presented Papers, 43rd Annual Convention of the Parapsychological Association*, pp. 249-266, Cary, NC.

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

Rhine, J. B. (1934). *Extra-Sensory Perception*. Boston, MA, US: Bruce Humphries.

Rhine, J. B. (1938). Experiments bearing on the precognition hypothesis: I. Pre-shuffling card calling. *Journal of Parapsychology*, 2, 38-54.

Rhine, L. E. (1961). *Hidden channels of the mind*. New York: Sloane.

Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: three failed attempts to replicate Bem's 'retroactive facilitation of recall' effect. *PLoS ONE*, 7 (3), e33423.

Robinson, E. (2011). Not feeling the future: A failed replication of retroactive facilitation of memory recall. *Journal of the Society for Psychical Research*, 75, 142-147.

Roll, W. G. (1966). ESP and memory. *International Journal of Neuropsychiatry*, 2, 505-521.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, 33, 217-236.

Sartori, L., Massaccesi, S., Martinelli, M., & Tressoldi, P. E. (2004). Physiological correlates of ESP: Heart rate difference between targets and non-targets. *Journal of Parapsychology*, 2, 351-360.

Savva, L., & French, C. C. (2002). Is there time-reversed interference in Stroop-based tasks? *45th Annual Convention of the Parapsychological Association*, Paris, 5-9 August.

Scarborough, D. L., Cortese, C., & Scarborough, H. L. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 1-17.

Scargle, J.D. (2000). Publication bias: The "file drawer" problem in scientific inference. *Journal of Scientific Exploration*, 14(1), 91-106.

Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition*, 7, 377-396.

Sheehan, D. P. (2006). Retrocausation and the thermodynamic arrow of time. In Sheehan, D. P. (Ed), *Frontiers of time: Retrocausation – experiment and theory*, AIP Conference Proceedings for 87th Meeting of AAAS Pacific Division, University of San Diego.

Sherwood, S. J., & Roe, C. A. (2003). A review of dream ESP studies conducted since the Maimonides dream ESP programme. *Journal of Consciousness Studies*, 10, 85-109.

Simmons, J. P., Nelson, L. D., & SimonSohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366.

Sinclair, U. (1962). *Mental Radio* (2nd ed.). Springfield, IL: Charles C. Thomas

Snodgrass, J. G., & Corwin, J. (1988). Perceptual identification thresholds for 150 fragmented pictures from the Snodgrass and Vanderwart picture set. *Perceptual and Motor Skills*, 67, 3-36.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 6, 174-215.

Soetens, E., Boer, L., & Hueting, J. (1985). Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 598-616.

Sommer, W., Leuthold, H., & Soetens, E. (1999). Covert signs of expectancy in serial reaction time tasks revealed by event-related potentials. *Perception and Psychophysics*, 61, 342-353.

Spottiswoode, S. J. P., & May, E. C. (2003). Skin conductance prestimulus response. *Journal of Scientific Exploration*, 17, 617-642.

Srinivas, K. (1995). Representation of rotated objects in explicit and implicit memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 21, 1019-1036.

Srinivas, K. (1996). Size and reflection effects in priming: A test of transfer-appropriate processing. *Memory and Cognition*, 24, 441-452.

Stanford, R. G. (1974). An experimentally testable model for spontaneous psi events: I. Extrasensory events. *Journal of the American Society for Psychical Research,* 68, 34-57.

Stanford, R. G. (1975). Response factors in extrasensory performance. *Journal of Communication*, 25, 153-161.

Stokes, D. (1987). Spontaneous Psi Phenomena. In S. Krippner (Ed.) *Advances in Parapsychological Research*, 8, 6-87, Jefferson, N.C.: McFarland & Co.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010a). Meta-analysis of free response studies, 1992-2008: Assessing the noise reduction model in parapsychology. *Psychological Bulletin*, 4, 471-485.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010b). A meta-analysis with nothing to hide: Reply to Hyman (2010). *Psychological Bulletin*, 4, 491-494.

Tanaka, Y., & Shimojo, S. (1996). Location versus feature: reaction time reveals dissociation between two visual functions. *Vision Research*, 36, 2125-2140.

Tart, C. T. (1976). *Learning to use ESP*. Chicago, IL: University of Chicago Press.

Thalbourne, M. A. (2003). *A glossary of terms used in parapsychology* (2nd ed.). Charlottesville : Puente.

Thalbourne, M. A. (2004). *The common thread between ESP and PK*. New York : Parapsychology Foundation.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: capture of the eyes by new objects. *Psychological Science*, 9, 379-385.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113, 766-786.

Treisman, A. M. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behaviour*, 3, 449-459.

Tressoldi, P.E. (2011). Extraordinary claims require extraordinary evidence: the case of non-local perception, a classical and Bayesian review of evidences. *Frontiers in Psychology*, 2:117. DOI: 10.3389/fpsyg.2011.00117.

Tressoldi, P.E., Martinelli, M., Massaccesi, S., & Sartori, L. (2005). Heart rate differences between targets and non-targets in intuitive tasks. *Human Physiology*, 31, 646-650.

Tressoldi, P. E., Martinelli, M., Scartezzini, L., & Massaccesi, S. (2010). Further evidence of the possibility of exploiting anticipatory physiological signals to assist implicit intuition of random events. *Journal of Scientific Exploration*, 24, 3411-3424.

Tressoldi, P. E., Martinelli, M., Semenzato, L. & Cappato, S. (2011). Let your eyes predict: Prediction accuracy of pupillary responses to random alerting and neutral sounds. *SAGE Open,* DOI: 10.1177/2158244011420451, 1-7.

Tressoldi, P. E., Martinelli, M., Zaccaria, E., & Massaccesi, S. (2009). Implicit intuition: How heart rate can contribute to predict future events. *Journal of the Society for Psychical Research*, 73, 1-13.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

Tyrell, G. N. M. (1946). The "modus operandi" of paranormal cognition. *Proceedings of the Society for Psychical Research,* 48, 65-120.

Ullman, M., & Krippner, S. (1989). *Dream Telepathy* (2^{nd} ed.). Jefferson, NC: McFarland.

van Turennout, M., Bielamowicz, L., & Martin, A. (2003). Modulation of neural activity during object naming: Effects of time and practice. *Cerebral Cortex*, 13, 381-391.

Ungerleider, L. G., & Bell, A.H. (2011). Uncovering the visual "alphabet": advances in our understanding of object perception. *Vision Research*, 57(7), 782-799.

Wackermann, J. (2002). *On cumulative effects averaging artefacts in randomized S-R experimental designs*. Proceedings of Presented Papers: The Parapsychological Association 45th Annual Convention, 293-305.

Wackermann, J., Pütz, P. & Allefeld, C. (2008). Ganzfeld-induced hallucinatory experience, its phenomenology and cerebral electrophysiology. *Cortex*, 44, 1364-1378.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., Rogier, K., & van der Maas, H. (2011). Yes, psychologists must change the way they analyse their data: Clarifications for Bem, Utts, and Johnson (2011). Retrieved from http://www.ruudwetzels.com/articles/ClarificationsForBemUttsJohnson.pdf

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology,* 100, 426-432.

Warcollier, R. (1939). *Experiments in telepathy*. London: George Allen & Unwin.

Warren, C. A., McDonough, B. E., & Don, N. S. (1992a). Event-related brain potential changes in a psi task. *Journal of Parapsychology*, 56, 1-30.

Warren, C. A., McDonough, B. E., & Don, N. S. (1992b). Partial replication of single subject event-related potential effects in a psi task. *Proceedings of Presented Papers: The Parapsychological Association 35th Annual Convention*, 169-181.

Wezelman, R., Gerding, J. L. F., & Verhoeven, I. (1997). Eigensender ganzfeld psi: An experiment in practical philosophy. *European Journal of Parapsychology*, 13, 28-39.

Wiggs, C.L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, 8, 227-233.

Wiggs, C. L., Martin, A., & Sunderland, T. (1997). Monitoring frequency of occurrence without awareness: Evidence from patients with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 19, 235-244.

Yarbus, A. L. (1967), *Eye Movements and Vision*, New York: Plenum.

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, 11, 125-131.

Zusne, L., & Jones, W. H. (1982). *Anomalistic psychology: A study of extraordinary phenomena of behaviour and experience*. Hillsdale, NJ: Lawrence Erlbaum Associates.