Time as a Fourth Spatial Dimension:

Linking Retrospective and Prospective Time to Space.

Bradley Jay Patten

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy in Psychology

The University of Auckland

2014

Abstract

Experiments addressing the relationship between time and space are conducted using separate approaches for retrospective and prospective timing.

Retrospective timing, at the subsecond and suprasecond level, is investigated using a recognition memory paradigm that has previously shown spatial effects. Although space-like effects with time stimuli were not found, there is some promise for the future.

During these analyses a sample size bias is identified in two common measures of signal. Separate formulas are devised for the correction of the sample size error associated with d' and A'. Computer simulations confirm the removal of the bias when the corrections are used.

Experiments on prospective timing use the more established kappa and tau effect paradigm. In four experiments a relationship between perception of time and space, indexed through scan-speed, is seen in the tau effect. By varying the size and duration of the stimuli used, the relationship between time, scan-speed, and space is manipulated. The ability to manipulate this relationship is taken as evidence of a resonance point specific to a given stimulus set. The results of these experiments support theories linking perceptions of time and space, and help account for contradictory findings common in the literature. These findings overturn and simplify accounts that have stood since 1977 by providing a common equation describing the relationship between the tau effect and scan-speed in this paradigm. The failure to produce similar findings with the kappa effect highlights the difference between these effects and provides a key area for future research.

Retrospective and Prospective findings are discussed in light of related research and possible directions for future work are suggested. General suggestions are made for more retrospective timing work as retrospective timing is heavily under-researched. Specific guidance is provided for future prospective timing work. Accounts of the results under scalar timing theory and spectral models are provided. A potential link between retrospective and prospective time is also outlined.

Dedication

This thesis would not have been possible without the willing assistance of over 400 participants (and that's just for the experiments that have made it into the thesis). In recognition of their mostly unpaid (but always underpaid) help I dedicate this thesis to them.

Acknowledgements

I would like to acknowledge the help of my primary supervisor, Associate Professor Jeff Hamm. Jeff has been a consistent source of support throughout my time as a postgraduate student at the University of Auckland. He was also the first lecturer to get me truly excited about psychology, asking questions rather than just learning answers, then provided the means and opportunity to design experiments to find answers myself. Thanks also for his patience with my tendency to become distracted by anything shiny or horrendously over complicated in my results.

I would also like to thank my secondary supervisor Professor Ian Kirk, particularly for his assistance in the first year of my thesis and for not complaining too loudly as the project drifted from teleportation in virtual environments to a time/line length judgement paradigm.

Thanks also to Professor Tracy Taylor-Helmick, whose design I hijack for my own nefarious purposes in the first section of the thesis and whose help was much appreciated during the design and analysis of the directed forgetting experiments. In particular, it was great to have a fresh set of eyes to pull me out of a deep rabbit hole where I otherwise would have continued to chase ghosts of effects that did not exist.

More generally, to my colleagues from the lab and the wider University, thanks for making days in the lab more interesting. Thanks for not running when I pounced on you for yet another experiment and thanks for eating all the baking. Among these anyone from 309 deserves a special mention as do Graeme Bibby and Michelle Burstall. Graeme, you spent more time in the side room than anyone who didn't have to. Michelle, you are the woman with all the answers.

Funding for this PhD was partially provided through the University of Auckland Doctoral Scholarship. Without this scholarship the expense of the program would have been too much to bear. I am grateful for its availability and hope it will remain available for others for years to come.

As for family, biological and in-law, thank you for providing a safe thesis-free space to retreat to. To your credit I cannot remember the last (or first) time you asked when I would be finished or told me to get a haircut and get a real job. Your support during this time has been much appreciated.

Finally I would like to thank my wife, Jessica Patten. Without her emotional (and, let's be honest, financial) support I would not have made it through the thesis in such good spirits, if at all. She would also like me to say she is beautiful and amazing with three exclamation marks.

Table of Contents

Glossary

Kappa effect: Where spatial information impacts on perceptions of time. Specifically longer lines or larger spaces lead to the perception of longer time.

Tau effect: Where temporal information impacts on perceptions of space. Specifically longer times lead to the perception of longer lines.

Time divisions

- Automatic: Timing that occurs without the need for attention. Associated with a left lateralised network.
- Cognitive: Timing that occurs with the need for attention. Associated with a right lateralised network.
- Explicit: Voluntary timing.
- Implicit: Timing that occurs without the need for voluntary control.
- Macrotime: A time label, probably semantic in nature, associated with episodic memory.
- Microtime: The time between events within a single episode. Primarily associated with episodic memory.
- Prospective: Timing of events as they happen. Restricted to events occurring in the present or the future.
- Retrospective: Timing of events that have already occurred after they occurred without the participant knowing that they needed to be timed when they were happening.
- Subsecond: Short times of less than 1 second using a conservative estimate or 3 seconds using a more generous estimate.
- Suprasecond: Times above the subsecond range. Here they are restricted to times of less than 20 seconds.
- Episodic Time: Times of more than 20 seconds. Named due to their apparent reliance on the hippocampus.

Chapter 1. General Introduction

In day to day life, time and space share clear connections. Consider the amount of time it takes to walk across a room; the larger the room, the greater the time. In fact, assuming you travel at a constant speed, a one unit increase in the length of the room will lead to a one unit increase in the time taken to traverse it. This relationship is unavoidable in the world around us and some researchers (e.g. Casasanto & Boroditsky, 2008; Walsh, 2003) suggest this relationship could be unavoidable in the representations of space and time in the human mind as well. If the latter is true, similarities should appear in the behaviour associated with, and the brain regions activated by, space-based and time-based tasks. However, before considering such evidence, it is necessary to define what is meant by time and to consider the scales on which it is thought to exist.

While arbitrary fourth dimensions have been constructed (Aflalo & Graziano, 2008), the true fourth dimension is simultaneously more abstract and more familiar. It is the dimension of time. Although invisible, it is time that allows the viewer to move around an object or environment and it is time that allows for an ongoing perception of events. When it enables the experience of the present it is known as the time of consciousness (Durgin & Sternberg, 2002). However, the concept of time is more frequently paired with the notion of time that has passed or will pass: the idea that it is something that can be quantified or measured. In this case it is often referred to as the concept of duration. In most research duration is seen as a confounding variable and is controlled out and/or used externally as a measure of performance (i.e. response time). However, a great deal of research into the mental representation of time has occurred. In order to gain a good understanding of this research it is useful to first consider the scales on which time is conceived, theories regarding the representation of time, and examples of common findings in behavioural, imaging and neuropsychological work. Once these basics have been covered consideration of the relationship between time and space will be considered more fully in the section titled “The Link”.

Time Scales

While time may seem to be a simple concept it is one whose representation spans two directions and has been found to contain a number of levels or scales. In terms of direction, time can be seen as something that is referring to things that are yet to happen (the future) or things that have already happened (the past). With the exception of certain film titles, we typically look back to the past and forward to the future. It is this difference that forms the first important division in time research. When given a timing task, if a participant is aware that they will have to time an interval that is yet to occur the task is termed a prospective timing task. If participants are only made aware of a requirement to time the interval after it has occurred the task is termed a retrospective timing task. This division of time is associated with critical differences in performance and theories around how timing is achieved. In particular, increasing the number of distracters present during a target interval will increase estimated durations in retrospective tasks but will decrease estimates in prospective tasks (Block & Zakay, 1997; Predebon, 1996).

The longest duration that is associated with a specific system and still thought of as timing is the circadian rhythm. The next largest, while not often considered in *time* research, might be episodic memory. The shortest scales, and those most investigated in time research, have three different classifications. They can be divided into sub and suprasecond timing, implicit and explicit timing, and/or automatic and cognitive timing.

The circadian rhythm is the least contested time range. Mediated by the pineal gland and its production of melatonin, the circadian rhythm governs the basal activity level in the body (Bear, Connors, & Paradiso, 2007). It gives cues for when it is time to eat and sleep and even influences the core temperature and mental alertness (Bear et al., 2007). While it is sensitive to cues from the external environment it is largely under internal control, as demonstrated by experiments where external time cues are removed and the rhythm is largely maintained (Mills, Minors, & Waterhouse, 1974). A simpler demonstration is the case of jetlag. Here, despite an abundance of external cues to the contrary, internal cues can insist it is night-time in the middle of the day and daytime in the middle of the night.

The next time class may be the time of episodic memory. While episodic memory is primarily thought of as a memory system used to represent events (Tulving, 2002) rather than a system that represents time, logically it should require a representation of the order and passing of those events. Recently it has been suggested that time representation in episodic memory may have two levels: macrotime and microtime (Hassabis & Maguire, 2007). Under this account macrotime is believed to be a time label, possibly semantic in nature, supplying information about when a specific episode happened (Hassabis & Maguire, 2007). Combining this division with work suggesting that episodic memory and future planning may share common structures and processes (Addis, Wong, & Schacter, 2007) macrotime would be the scale that allows for the use of time concepts that extend beyond the time ranges covered here, such as multiple days, weeks, centuries, or millennia. In contrast, microtime is believed to supply the order of, or time between, occurrences within an event. For example, microtime would refer to the time between entering a store and finding the item you intend to purchase. In this case macrotime would refer to the day and time at which this episode occurred.

At the smallest durations in timing, in what might be considered the microtime range, it seems to be widely accepted that at least two timescales exist within prospective timing. First there are times in the range of seconds to minutes, so called suprasecond time (Buonomano, 2007; Lewis

& Miall, 2003a). Below this lies the subsecond range (Buonomano, 2007; Lewis & Miall, 2003a). Yet while subsecond implies below and suprasecond implies above one second, the division is not so convenient. Some suggest the change occurs at one or one and a half seconds depending on the individual (Gilden & Marusich, 2009), some two (Hirsh & Sherrick, 1961), and some three seconds (Poppel, 1997). Others even suggest that duration is only one factor in determining the switch (Coull & Nobre, 2008; Lewis & Miall, 2003b). In these cases a functional division is preferred.

One such division is into implicit and explicit timing (Coull & Nobre, 2008). Implicit timing is associated with a left lateralised network and refers to timing processes that are activated involuntarily. Explicit timing is associated with a right lateralised network and refers to timing that requires an overt time response that is typically voluntary. Each can be divided into perceptual and motor subcomponents, which can in turn be thought of as systems for time perception and time production. Implicit perceptual timing describes instances where a region of time is perceptually important or predictable, such as when it is cued in attention tasks (e.g. Coull & Nobre, 1998). Implicit motor timing is believed to be involved in the co-ordination of motor output and is particularly useful for rhythmic productions. Explicit perceptual timing occurs when two time periods must be compared while explicit motor timing occurs when a period of time must be produced or reproduced. In such cases subsecond timing would typically employ implicit timing while suprasecond timing would employ explicit timing.

An alternative but similar division is into automatic and cognitive processes based on the duration of the task, whether movement is involved, and whether the task is predictable (Lewis & Miall, 2003b). Based on a review of the literature it was proposed that automatic timing occurred when tasks involved continuous or predictable movement in the subsecond range. Cognitively controlled tasks occurred when the tasks involved discrete suprasecond intervals not requiring movement. Additionally, automatic processes were associated with areas associated with motor systems and a left lateralised network while cognitive processes were associated with a right lateralised network including the frontal lobes and the parietal lobes. Particular explanatory emphasis was placed on working memory and attention in this regard.

While the alternate functional divisions do have subtle differences between them they can generally be used interchangeably. Their common lateralisation of shorter durations to the left hemisphere and longer durations to the right hemisphere builds confidence in this particular division. Their classification by task type rather than duration also helps to explain the disparate assumptions about when time moves from the subsecond domain to the suprasecond domain. For the purpose of simplicity, subsecond and suprasecond will be the preferred terminology for naming the target time ranges while implicit, explicit, automatic, and cognitive will be used to describe functional roles for timing where appropriate.

One final distinction, and one that essentially allows participants to extend subsecond timing to the suprasecond range, is worth considering: the distinction between timekeeping and chronometric counting. When asked to time intervals it is common for people to count things such as one thousands, hippopotami, or Mississippi's. Such behaviour is termed chronometric counting (Hinton, Harrington, Binder, Durgerian, & Rao, 2004) and is typically not what time researchers are interested in. Comparisons of chronometric counting to time keeping have found counting to be more accurate (Hinton & Rao, 2004) in adults and 8 year olds but not 5 year olds (Liberman & Trope, 2008) and that counting leads to different areas of neural activation (Hinton et al., 2004) and different characteristics in the intervals produced (Hinton & Rao, 2004). Specifically, counting elicits greater activity in the supplementary motor area, the motor cortex regions associated with the mouth, Broca's area, and the right cerebellum (Hinton et al., 2004) and produces less variability than would otherwise occur (Hinton & Rao, 2004). While chronometric counting is potentially problematic for time research it can be avoided through employing distracter tasks or simply asking participants not to count[1] (Hinton et al., 2004; Hinton & Rao, 2004). Unless otherwise specified, time keeping will be the focus of this thesis from this point on.

Key Theories

When it comes to explaining how time is measured and/or perceived at the sub and suprasecond levels there are two key types of theory with some alternatives gathering support. The first type of theory proposes that time is measured using a dedicated internal clock while the second suggests that timing is achieved by a spectrum of cells that have become tuned to timing differences. One alternative group of theories suggest that space and time may be interlinked.

Scalar timing theory. The first and most popular theory is of a unique timing system: the internal clock model. It is currently referred to as the scalar timing theory (STT) (Gibbon, Church, & Meck, 1984) or the attentionally gated model (Zakay & Block, 1997) although references to its older name, the Scalar Expectancy theory (SET) (Gibbon, 1977), still occur. This theory postulates that, within the brain, there is a structure, or multiple structures (Treisman, Cook, Naish, & MacCrone, 1994), producing clock ticks in an oscillatory fashion. Provided a gate, argued to be controlled by attention (Zakay & Block, 1997), is open and a switch initiating timing is closed/'on', these ticks are

[1] While this is a riskier option it has been shown to prevent chronometric counting. This has been shown through the activation of different areas (Hinton et al., 2004) and different behavioural results (Hinton & Rao, 2004) when participants are instructed to chronometrically time and when they are instructed not to.

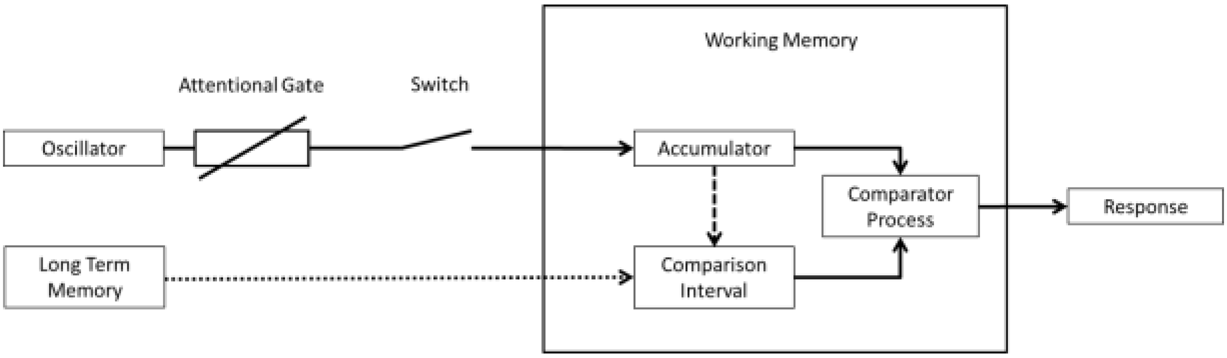

Figure 1.1. A depiction of the internal clock model outlined under Scalar Timing Theory. The oscillator produces ticks. If these ticks are attended to the attentional gate is said to be open. If the attentional gate is open and the switch is closed, ticks can pass to the accumulator. The number of ticks in the accumulator gives an indication of the amount of time that has passed. In some instances accumulated time is compared to a comparison interval. This comparison interval may come from long term memory or a previously timed interval held in working memory.

collected in an accumulator. If the attentional gate is closed or the switch is open/'off' the ticks are ignored. Accumulated ticks are then compared to another interval, recorded in working memory or recovered from long-term memory, to determine how much time has passed. The model is illustrated in Figure 1.1.

Critically STT is able to account for common effects seen in timing studies such as the scalar property/Weber's law from which it draws part of its name (Gibbon et al., 1984). The scalar property describes the tendency for the spread of the error to increase as the length of the timed interval increases. The scalar aspect seems to come from the oscillator itself (Wearden & Bray, 2001) with some research suggesting that the time between ticks increases as the interval lengthens (van Rijn & Taatgen, 2008).

The attentionally gated STT account can also explain why the presence of more distracters leads to a reduction in perceived time during a prospective task. Because the distracters capture attention, the attentional gate closes when a distracter is presented. Therefore some ticks are not accumulated, which in turn leads to a drop in perceived time (Predebon, 1996).

While the theory itself does not suggest brain regions associated with each subcomponent, areas of the brain have begun to be associated with the theory's functions (see Meck, 1996 for an early example). The number of clocks is also up for debate. Some suggest multiple clocks (Tversky, Kugelmass, & Winter, 1991), such as one for visual tasks and one for auditory tasks (Hirsh & Sherrick, 1961) or one for all tasks with others for modality specific demands (Bueti, Bahrami, & Walsh, 2008). Some suggest multiple clocks even within individual tasks (Hirsh & Sherrick, 1961) while others suggest a single clock for a single domain (van Rijn & Taatgen, 2008).

Spectral models. The second group of theories are collectively referred to as spectral models (such as Hopfield & Brody, 2001). These models suggest that no unique system is necessary. Instead time can be inferred from the properties of task related systems. As an example, a different set of neurons would be expected to be active 1 second into a task than those active at the beginning. Proponents of this form of theory argue that such things can be modelled with neural networks. Furthermore such models are argued to account for psychophysical results that the STT does not.

In the specific case of Karmarkar and Buonomano (2007) the authors referred to findings in the subsecond range. They found that when participants were required to respond to intervals of 100 to 200ms responses became less accurate when preceded by a distracter interval. Furthermore they were far less accurate when the distracter was unpredictable. Finally if two target intervals were presented in rapid succession at the same auditory frequency interference was observed, yet when two target intervals were presented at different auditory frequencies no interference was observed. Under a spectral theory difficulty in accurate time perception arises in each of these situations as the starting state of the system is different to the expected zero position. Under a strict STT account none of this should matter as the switch could remain closed during the distracter interval or the accumulator could be reset at the start of the target interval. This should nullify the effect of the distracter interval or, at the very least, variability in the distracter interval. Most importantly though, if distracters did impact on STT there is no clear reason to expect them to impact on intervals presented in the same audio frequency but not different audio frequency bands. If a common clock is being used that is independent of the stimulus itself errors observed in one frequency should be observed in the other frequency as well. Taken to the extreme such a spectral model of timing suggests that internal clocks are not necessary for, and are not used, in timing. Non-dedicated systems can convey the information themselves.

However, as a potential split from the overarching notion of the spectral theories, two spectral theories locate time perception in specific brain structures. The first locates timing in the basal ganglia (Lustig, Matell, & Meck, 2005) while the second suggests a role for the hippocampus (Grossberg & Merrill, 1992). In fact the hippocampal theory was one of the initial spectral theories and gave them their name by suggesting that time was monitored by a *spectrum* of time sensitive cells. However, despite being associated with a specific structure, it could be argued that these models operate independently of a true oscillator and, most likely, involve systems with a primary purpose other than timing. In this way they remain somewhat separate from clock based accounts.

Critically, while the findings of Karmarkar and Buonomano (2007) demonstrate some separation from STT accounts they were not reproduced with 1000ms intervals, intervals that could be argued to be in the suprasecond range. Instead findings at the higher range supported a clock like model such as STT. Karmarkar and Buonomano (2007) suggested that the distracter related deficits identified in their experiments most likely extend to about 500ms. However, no follow up experiments were used to verify this particular time point.

Time as space theories. An alternative is the idea that the representation of time is equivalent to or reliant upon that of space. It is a notion proposed at least as far back as 1890 (James, 1890) and one that receives occasional support within the research community (recent examples being Bueti & Walsh, 2009; Casasanto & Boroditsky, 2008; Oliveri et al., 2009; Walsh, 2003). While some advocates suggest that time is dependent on space (Casasanto & Boroditsky, 2008) others suggest that time and space are dependent on a common metric (Bueti & Walsh, 2009; Walsh, 2003).

The notion of time being dependent on space can grow out of linguistic analysis (Casasanto & Boroditsky, 2008; Ulrich & Maienborn, 2010). When the language used to talk about time is compared to that used to talk about space similarities are uncovered. For example, people talk about meeting *on* a date, *at* a time, or *moving* meetings *forward* or *backward*. As there are fewer instances of time being used to describe space, light-*years* being one example, Casasanto and Boroditsky (2008) suggested that space was used to understand and conceptualise time more than time was used to conceptualise space. Based on this imbalance in metaphors Casasanto and Boroditsky (2008) proposed that an asymmetrical relationship between space and time should exist. In essence, changes in space would have a larger influence on time than changes in time would have on space.

However it is also arguable that space does not hold a privileged position over time. This notion is expressed in ATOM (Bueti & Walsh, 2009; Walsh, 2003). ATOM stands for a theory of magnitude and suggests that time, space, and numbers are all just measures of magnitude, each of which is important for action. Whereas the previous linguistic account argues for an asymmetrical or unidirectional relationship between time and space, where space will influence time more than time will influence space, ATOM allows for, but does not necessarily require, a symmetrical or bidirectional relationship.

Under any of these theories quantities of time, whether represented as space or magnitude, would still need a representation. As space and magnitude are essentially measures of how much of something exists this appears to favour an accumulation based model rather than a spectral model. In this sense the system proposed in STT could still be useful; time as space theories might simply redefine what the accumulator is and what it is accumulating. The key difference would be that time would not be identified as *time* in the accumulator. Instead time would be represented through space or as some generic unit of magnitude.

The role of working memory. As working memory is often drawn into descriptions of how timing is achieved at an experimental as well as a theoretical level some consideration of how working memory might apply to timing models is appropriate at this stage. Critically, the standard model of working memory (Baddeley, 2003) is not typically described as having a temporal aspect. Instead working memory is made up of the central executive, the phonological loop, the visuospatial sketchpad and a potential fourth element: the episodic buffer (Baddeley, 2000). Theories of time perception appealing to working memory as a core component, and experiments describing time tasks as working memory tasks, must rely on the re-purposing of one or more of these systems for use in timing or the existence of an unspecified time related component. Consideration of potential roles for each of these elements will be presented briefly in turn.

The central executive provides the first candidate for a time based role and has been implicated in such a role before (Brown, 1997). While it is said to have a role in executive tasks, such as randomisation (Field & Groeger, 2004), it is generally thought of as more of a control element than a representation element (Baddeley, 2000). If time were an executive task this would associate its representation with central activation of the frontal lobes (Baddeley, 2003).

The phonological loop provides a clearer option. In addition to an ability to hold sounds of approximately 2 seconds in duration (Baddeley, Thomson, & Buchanan, 1975) its left hemisphere location (Baddeley, 2003) makes it compatible with left lateralised accounts of subsecond time. Furthermore, verbal working memory tasks are associated with activation in the supplementary motor area, Broca's area, and Wernicke's area and are believed to be the result of sub-vocal verbalisation (Baddeley, 2003). Similarities between this system and that identified by Hinton et al. (2004) make a compelling case for use of the phonological loop in chronometric counting tasks.

The visuospatial sketchpad, with its spatial associations, would be a good fit for ATOM and time as space models. In such cases the sketchpad could act as the accumulator. Additionally the associated right hemisphere activations see it fit well with the Lustig et al. (2005) spectral account with activation typically observed in the frontal lobes and the right parietal cortex (Baddeley, 2003). Such lateralisation also fits the proposed cognitive (Lewis & Miall, 2003b) and explicit (Coull & Nobre, 2008) divisions of timing. In these cases the combination of right parietal activity and frontal lobe activity would suggest the co-activation of the visuospatial sketchpad and the central executive. This is a scenario that is logical from a theoretical perspective and in line with activations typically seen in spatial working memory tasks (Baddeley, 2003).

The final consideration must be the possibility of a time specific component. Recently such an additional component has been suggested: the episodic buffer (Baddeley, 2000). The primary proposed purpose of this buffer is in providing an overflow space, one that can be used when either the phonological loop or the visuospatial sketchpad exceed their capacity, and/or the provision of an association space for tasks that require the combination of spatial and verbal information. Supporting the case for the existence of the episodic buffer, evidence supporting an anatomical corollary, the hippocampus, has begun to be uncovered (Berlingeri et al., 2008; Rudner, Fransson, Ingvar, Nyberg, & Ronnberg, 2007).

Retrospective models. Each of the models considered so far have focussed on prospective timing. Models focussing on retrospective timing are less common and less developed, perhaps owing to the contention that retrospective timing occurs without involving a timer (Ivry & Hazeltine, 1992). In reality only one model of retrospective time has taken hold, one that centres on qualities of the durations to be remembered.

The storage size theory, originally proposed by Ornstein (1969) and later refined by Block and Zakay (1997) to a theory based on memory complexity, is the key retrospective timing theory. Under this theory retrospective memory for time is based on the amount of information in memory: as the number of events making up an interval increases, or as the complexity of the events within an interval increases, so too does the reported duration. Generally speaking this is supported by findings that demonstrate that more events, or more complex events, lead to longer estimates in retrospective timing tasks regardless of actual event duration (Block & Zakay, 1997; Ornstein, 1969). While some moderating factors have been identified (Block & Zakay, 1997) the structure of the model itself remains undefined.

Model conclusions. At this stage it may be worth noting that the models are not mutually exclusive. When considering the spectral models and the STT model one could think of the oscillator as a spectral system. Furthermore, the fact that there is currently no specific neural system associated with temporal working memory could lend support to the idea of the visuospatial sketchpad being used to represent time, thereby introducing the link between space and time. As for retrospective models the lack of clear structure leaves open the possibility for overlap with prospective models. One candidate for overlap would be the idea that retrospective timing could use prospective processes during the replay and/or reconstruction of episodes. However, the differences between prospective and retrospective tasks also allows for the possibility of entirely different systems.

Research Examples and Relevance

In considering the processing of time, common behavioural, imaging, and neuropsychological findings need to be considered in order to determine if the divisions and models proposed are accurate and which structures are associated with which functions. However, before this can be undertaken, some key issues with time research need to be considered.

First, the implications of the differences between the methodologies used to study retrospective and prospective time need to be considered. Due to the nature of their experimental design retrospective timing tasks are far less common than prospective timing tasks. The key element of retrospective timing design that is problematic is the tendency for their tasks to have a single trial (such as Block & Zakay, 1997; Ogden, Wearden, Gallagher, & Montgomery, 2011; Ornstein, 1969). For example, subjects are often asked to perform a simple task, such as an uppercase vs. lowercase letter judgement task, and are asked to estimate the duration of the task on its completion. Once this single estimate has been provided the retrospective timing element of the experiment is over. Although it makes retrospective timing experiments ungainly and leads to high variability, this single trial approach is necessary as it means that participants are unlikely to become aware that the experiment is investigating time and shift their attention to time. This is the very definition of prospective timing and the antithesis of retrospective timing. For this reason the majority of time research is performed on prospective timing. As such, the findings covered will focus on prospective time unless otherwise specified.

Secondly, as debate exists around the point of division between subsecond and suprasecond times their suitability for classifying the results of different studies must be considered. While division based on the durations commonly paired with each time class is tempting, it has already been noted that candidates for the point of division is unclear. The alternative would be to base the distinction on the activity of brain regions. However this is impractical. First, not all studies utilise imaging techniques. Second, basing the distinction between sub and suprasecond time on proposed splits naturally enforces those splits through the use of circular logic. Instead the evidence should be considered on its own merits. Through this agnostic approach, divisions between sub and suprasecond time, if they exist at all, should be free to emerge from the evidence on their own rather than being imposed by the structure of the investigation.

Behavioural.

Detecting time. In some respects the most interesting division of time relates to the point at which time can be said to be perceived at all. This point of time detection is associated with two levels of perceptual awareness. The first, investigated with Temporal Onset Judgement (TOJ) tasks, is the minimum separation required to perceive two stimuli as being separate stimuli rather than a single constant stimulus. In the auditory domain the threshold is a 2ms separation (Hirsh, 1959); in the visual domain the threshold is a 44ms separation (Fortin & Breton, 1995).

Although two separate events can be reported in the tasks above, the order of those events is unknown. The threshold at which items can be placed in their correct temporal order (Hirsh, 1959; Poppel, 1997), sometimes called the temporal order threshold (TOT) (Ulbrich, Churan, Fink, &

Wittmann, 2009), is the next important point in time detection. While this threshold has been placed as low as 20ms (Hirsh, 1959) it has been shown to creep higher, changing with the complexity of the stimuli (Ulbrich et al., 2009) and the focus of attention (Spence, Shore, & Klein, 2001).

Measuring time. Weber's law, or the scalar property as it is often called in time research, refers to the common finding that when asked to estimate the length of an interval the variability around the estimated interval increases with increasing length of the target interval (Gibbon et al., 1984). As such, time estimates spread outwards like a fan when plotted on a graph. Although consistently reported in humans and animals at sub and suprasecond time scales (Gibbon et al., 1984) the source of these fluctuations is debated. While some have suggested that the memory store could be responsible, a study design that excluded the use of long term memory still produced results showing the scalar variance pattern (Wearden & Bray, 2001). While the authors suggested that the variation could potentially occur within working memory their results do not rule out the oscillator itself as the source. While there is at least some support for the oscillator producing such variance (van Rijn & Taatgen, 2008) the nature of the clock model makes narrowing down the correct answer difficult. Further complicating matters, spectral models can also account for such findings.

Yet this is not the only difference that occurs across time scales. Vierordt's law describes the common finding that shorter time intervals tend to be overestimated while longer intervals tend to be underestimated (Lejeune & Wearden, 2009). While initially proposed based upon data from a single participant it has since been replicated in a number of studies using multiple participants and is found in all timing task types including verbal report (Lejeune & Wearden, 2009). Interestingly a point of indifference is said to exist between the two ends of the scale at which point time estimations become accurate. The location of this point is dependent on the width of the scale in the auditory domain but tends to be around 3 seconds for the visual domain. Critically the variation of the time of this point of indifference suggests that a fixed division between sub and suprasecond time arising from a set oscillatory rate (Poppel, 1997) may be unjustified.

The clocks. While key theories outline the structure of a single clock this does not mean that they imply or require the existence of no more than one clock (Gorea, 2011). Support for multiple clocks come from findings of spatially localised distortions in timing (Johnston, Arnold, & Nishida, 2006), existence of discrepancies between auditory and visual timing (Burr, Banks, & Morrone, 2009; Goldston.S & Lhamon, 1974; Grondin, Bisson, Gagnon, Gamache, & Matteau, 2009), and the suggestion of different systems for sub and suprasecond times (Coull & Nobre, 2008; Lewis & Miall, 2003b).

However some research suggests that timing is not achieved by multiple clocks when performed within a single modality (van Rijn & Taatgen, 2008). This research found that when an interval to be produced was preceded by production of another interval, the second interval took longer to produce. For example, if the interval to be produced was 2 seconds long in both cases participants would always report they were producing a 2 second interval. However their second 2 second interval would take longer to produce than their first 2 second interval. This implied a slowing of the clock tick rate. Such a finding may contribute to explaining Weber's law and Vierodt's law. In the case of Weber's law the two tick values either side of the correct response are further apart at longer intervals than shorter ones leading to greater variability. In the case of Vierodt's law shorter intervals would be associated with faster tick rates and longer intervals slower tick rates. This would mean that shorter intervals would be less prone to timing disturbances throughout experiments while longer intervals would be more susceptible to changes.

It is also worth noting that the tick rate, the experienced rate of time, can be increased or decreased by external stimuli. Click trains have been shown to increase the rate of the clock (Wearden, Philpott, & Win, 1999) while boredom inducing experimental designs in humans, and extreme cold in animals, have been shown to contribute to a slowing of the clock (Wearden, 2008). Boredom related changes to the clock tick rate may explain the common finding of time lengthening in experimental designs (Brown, 1997) as the experiment should become less interesting as it continues.

Attention. Such slowing of the clock with increasing boredom could be attributable to an actual decrease in the oscillation rate of the pacemaker (Wearden, 2008) or to decreased attention. Decreased attention has been shown to produce underestimations of the time periods experienced (Predebon, 1996) and is included in clock models. Furthermore, attention has been shown to have a differential effect on prospective and retrospective timing (Predebon, 1996). If attention to items in a display is increased in prospective tasks, such as through increasing a task's difficulty, the prospective time estimate decreases (Hicks, Miller, & Kinsbourne, 1976). The retrospective estimate however is likely to remain unchanged. Furthermore, if the number of items in an interval is increased *and* these items must be attended to, the prospective estimate will decrease while the retrospective estimate will increase. This is explained in terms of the processes used to estimate time. In prospective timing the participant is assumed to time the interval as it occurs, as such increasing the number of items seen reduces the amount of attention available for time, causing the observer to drop ticks and the perceived time to decrease. In the case of retrospective judgements the participants must reconstruct the interval. In doing so, participants refer to the number of items that occurred. As this number increases so too does their estimate of the time that has passed (Predebon, 1996).

However, split attention and a tendency toward distraction do not always lead to decreased time perception. In the case of people with Attention Deficit Hyperactivity Disorder (ADHD) or high impulsivity, time can actually be perceived as moving slower than it is. In experiments investigating time, such participants verbally report intervals as longer than they are, but under-produce intervals when compared to control participants (Wittmann & Paulus, 2008). This is equivalent to saying that ticks are accumulated more quickly (Wittmann & Paulus, 2008). In line with this, a rhythm production task showed deficits for participants with high impulsivity and/or ADHD relative to control participants. Control participants were able to maintain rhythms as slow as 40 beats per minute. Consistent with the idea that people with ADHD accumulate ticks more quickly those with higher impulsivity and/or ADHD encountered difficulty at 60 beats per minute (Gilden & Marusich, 2009). In each case dropping below the threshold leads to a qualitative change in rhythmic productions: production became less accurate and highly variable. Interestingly, it also provides potential evidence of where the border between sub and suprasecond time intervals might be: in the 1 to 1.5 second range corresponding to a potential lower cap of 60-40 beats per minute.

Imaging and neuropsychology. Pure behavioural timing designs suffer from difficulty in pinning down which element of the clock is used by a task and have difficulty distinguishing between potential theories of timing. This is evident in the above cases where there is speculation as to which element is involved in which phenomenon. This is further exacerbated when the underlying structures involved cannot be determined. Such limitations can be at least partially overcome by imaging and neuropsychological work.

Electroencephalography (EEG) is the simplest first step from behavioural designs. While the spatial resolution is comparatively low it retains a focus on reaction times, moving the site of observation from button press responses to neural responses. EEG offers the advantage of investigating how brain activity changes over time during the task. Ultimately though, the regions used must also be considered. At the simplest level neuropsychological data can be used to gain an understanding of which areas are required for successful timing. However, while lesion evidence is helpful, evidence from members of the general population gives a better indication of how healthy brains function. Here functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), and Transcranial Magnetic Stimulation (TMS) have the advantage.

EEG. While EEG studies initially sought to associate temporal encoding with alpha activity this proved unsuccessful (Block, 1990). Instead, more recent studies have found associations with Event Related Potential (ERP) components. In particular, associations have been found with a

Contingent Negative Variation (CNV), a p150, a p300, and patterns associated with oddball detection.

CNV. In time research fronto-central negativities have been reported in a number of studies (Gibbons & Stahl, 2008; Macar & Vidal, 2004; Tarantino et al., 2010). In one illustrative study by Tarantino et al. (2010) the CNV was shown to have an association with time. Participants were shown 2 durations per trial and asked to indicate whether the 2 durations were the same or different. The first was always 1000ms while the second could be 800, 1000, or 1200ms. For the 800 and 1000ms conditions the CNV was practically identical; however in the 1200ms condition there was a clear difference. Once the target interval of 1000ms had passed the EEG trace became far more positive than in the other conditions.

This switch to positivity supported the idea that the target interval was represented in an accumulator in working memory. Under such an account the accumulator collects ticks up to the point where the target interval has been exceeded. Here, the increase in negativity is associated with an increase in accumulated clock ticks. Once the target interval has been exceeded there is no reason to continue to collect ticks. Presumably this leads to the accumulator being emptied, resulting in the sharp reversal.

Importantly, these results are in line with findings from a previous review (Macar & Vidal, 2004). Across a number of studies the size of the CNV was seen to be associated with increases in time stored in the accumulator and increases in effort in encoding time. Importantly, the positive shift was thought to reflect the retrieval of memory and/or the reaching of a decision point in a duration based task. Each of these components has a core analogue in STT, increasing confidence in the theory and this collection of findings.

P150. Tarantino et al. (2010) also showed a p150 to be increasingly positive and move increasingly rightward on the scalp as durations increased despite corrections being made for underlying differences in the CNV between conditions. While EEG based claims about the areas involved in processing must be interpreted with extreme caution, this was interpreted as evidence of involvement of the right parietal cortex. This is in line with the idea that time relies on working memory, perhaps spatial working memory, and longer intervals tending to activate a more right lateralised network. Interestingly the authors specifically mention ATOM (Walsh, 2003) and Lewis and Miall (2003b, 2006).

P300. An alternative means of assessing EEG components related to time is to base comparisons on the qualities of the participants themselves. In one such study (Gibbons & Stahl, 2008) participants were identified as good timers or bad timers and differences between the groups were examined. Each trial began with the presentation of a 2000ms standard interval then, when the participants were ready, participants reproduced the 2000ms interval. Differences between good and bad timers were only found in recordings related to the standard interval, not the reproduction. One key finding was that good timers, those whose reproductions were less variable, had a larger p300 in response to the standard interval than bad timers. This was interpreted as demonstrating a role for attention in timing in line with behavioural work covered earlier. Specifically, participants were thought to be checking their representation of the standard interval at the start of each trial. Those who did check their representation produced a larger p300, believed to index their expectation, and performed more accurately.

Oddball detection. One final class of EEG response identified in Macar and Vidal's (2004) review are those related to the detection of oddballs, namely the mismatch negativity and the omitted stimulus potential. The mismatch negativity, as it applies to time, occurs when there is a disruption in a consistent rhythm. It is thought to originate in the auditory cortex and, unlike the effects covered so far, occurs automatically. It is so automatic that it can be found in comatose patients; in these cases it can give an indication of the likelihood of their recovery (Macar & Vidal, 2004).

The omitted stimulus potential has two subcomponents, a negative modality specific component followed by a positive component that peaks at 300ms, and is not modality specific. Critically the omitted stimulus potential shows evidence of automatic and cognitive control. In a key study (Bullock, Karamursel, Achimowicz, McClune, & Basareroglu, 1994) when stimuli were presented at over 2Hz omitted stimulus potentials could be elicited automatically. With low frequency stimuli (less than 0.3-1Hz) omitted stimulus potentials were only elicited when attention was directed to the target stimulus. This fits with the idea that duration is a key component in the split between prospective timing systems and also fits with the finding that perception of rhythm, associated with automatic/implicit/subsecond timing, breaks down once a certain delay point is reached (Gilden & Marusich, 2009). In the case of the Bullock et al. (1994), stimuli required attention once the pacing dropped to 60 beats per minute or less. This is compatible with the lower range identified by Gilden and Marusich (2009)[2].

Identifying regions: The contribution of fMRI/TMS/neuropsychology. While EEG can provide good temporal resolution, ironically spatial imaging and localisation techniques currently provide more information about temporal processing. Knowledge of regions associated with

[2] It is worth remembering that the Gilden and Marusich (2009) task did require subjects to attend to the rhythm they were trying to produce. At pacings slower than 60-40bpm production became less accurate but did continue.

particular tasks can be obtained through a number of techniques including fMRI, TMS, and examination of neuropsychological evidence. Here the methodology used will take a back seat to the regions uncovered in an effort to give a cohesive summary. One commonly identified site is the basal ganglia but more traditional hotspots have also been identified including the cerebellum and the cortex while a spectral model considered earlier points toward the hippocampus. More generally functional asymmetries should also be present if implicit and explicit timing really are lateralised.

The basal ganglia. The basal ganglia have proven to be a common site of activation in timing tasks. In one well controlled study by Hinton and Meck (2004) participants timed a suprasecond interval of 11 seconds. To control for potential task confounds, such as the need to produce a response and the absence/presence of sensory stimuli, a number of tasks were used. These were a combined timing and motor task, a timing-only task, a motor-only task, and control task. After successfully replicating 48 example durations of 11 seconds participants were required to produce 11 second intervals in an MRI scanner in the absence of examples or feedback. In one trial block the start of timing was cued by the onset of a stimulus, in another trial bock it was cued by the offset of the stimulus. By averaging across these Hinton and Meck (2004) hoped to remove the influence of the stimulus. Additionally two stimulus types were used: one auditory and one visual. The idea was that time specific activations should be present regardless of which stimulus type was used. In the timing only task participants noted to themselves when they thought the 11 second target was approaching, they then noted to themselves a time after the 11 second criterion such that the 11 second target would lie directly between their two notations. In the combined timing and motor condition the participants squeezed a response ball at the two time points rather than noting them to themselves. In the motor-only task the participants squeezed the ball in response to a stimulus. Finally in the control task the participants did nothing. By comparing across these conditions the authors were able to find time specific activation in the basal ganglia along with activation in the frontal cortex and the thalamus.

Imaging evidence for involvement of the basal ganglia is increasing (Beudel, Renken, Leenders, & de Jong, 2009; Ivry & Spencer, 2004). Furthermore timing deficits have been observed where changes have occurred in the dopaminergic system of which the basal ganglia are a part. Notable examples being patients with Parkinson’s disease (Koch et al., 2008; Malapani et al., 1998; Perbal et al., 2005), who under-produce dopamine and overestimate durations, and Huntington’s disease (Paulsen et al., 2004), who overproduce dopamine and under estimate durations[3].

[3] Potential associations with ADHD and dopamine were also highlighted by Gilden and Marusich (2009) in their rhythm production work. However, in this case, associations are theoretical only.

Importantly, work with animals supports these findings (Meck, 1996). Drugs that increase the amount of dopamine available in the brains of rats increase the perceived rate of time. Those that decrease the available dopamine decrease the perceived rate of time.

The involvement of the basal ganglia in timing helps to strengthen a specific spectral timing account: The Striatal Beat Frequency theory (SBF) (Lustig et al., 2005). This theory suggests that the basal ganglia detect coincidence in the cortical oscillation activity that moves through the striatal loops. Through conditioning, the loops become attuned to a specific level of coincidence between the activated areas and increase their firing rate as that level of coincidence, associated with a specific time point, approaches. In this way timing behaviour emerges from the basal ganglia without them being truly involved in timing per-se (Lustig et al., 2005). Importantly, while the consistent region of activation is problematic for a spectral account in that it seems to suggest that there may be a clock structure, it is not incompatible with a spectral account. Here, the basal ganglia are said to be involved in the task through their association with working memory rather than a dedicated clock per-se. As such the basal ganglia are not time specific. Instead, time is represented as a by-product of a system that has purposes beyond time.

However, the finding of a single region impacting on timing behaviour over a number of tasks is consistent with STT. Under such an account the basal ganglia would form part of the clock. While the specific role that the basal ganglia would occupy under a clock model is unclear some candidates emerge. Work with Parkinson's (Koch et al., 2008; Malapani et al., 1998; Perbal et al., 2005), Huntington's (Paulsen et al., 2004), and ADHD populations (Gilden & Marusich, 2009) point to the basal ganglia playing a role in the oscillator, the accumulator, or both. Other work also suggests roles in the storage of candidate intervals (Coull, Nazarian, & Vidal, 2008), which could just be a special case of the accumulator, and decision making (Ivry & Spencer, 2004) which is presented as a specialised case of an accumulator type model. Yet the most distinctive answer may come from animals rather than humans. Lesion studies have indicated that different regions of the basal ganglia have associations with different roles in timing (Meck, 1996). On the basis of this work the caudate and putamen have been associated with the accumulator while the substantia nigra has been associated with the oscillator.

The cerebellum. Another area commonly implicated in timing is the cerebellum. Key support for its involvement comes from work with participants with cerebellar lesions. Such studies have shown deficits in timing intervals of 550 and 950ms (Spencer, Ivry, & Zelaznik, 2005), rhythm production tasks (Schlerf, Spencer, Zelaznik, & Ivry, 2007), and discrete rather than continuous movements (Spencer, Zelaznik, Diedrichsen, & Ivry, 2003) when compared to control participants. Importantly though, cerebellar deficits in timing drop markedly by 4 seconds and have been shown to be absent beyond 8 seconds (Gooch, Wiener, Wencil, & Coslett, 2010). Taken as a whole this evidence suggests that the cerebellum is likely to be primarily associated with subsecond timing but not suprasecond timing.

The hippocampus. The simplest evidence supporting a role for the hippocampus comes from timing studies conducted with HM, a man whose hippocampus was resected bilaterally, and work with Traumatic Brain Injury (TBI) patients. HM was found to be accurate on intervals of less than 20 seconds but was inaccurate on intervals of greater than 20 seconds (Richards, 1973). This appears consistent with a study using TBI patients (Schmitter-Edgecombe & Rueda, 2008). Relative to controls these patients were able to produce intervals of less than 30 seconds well. However their performance worsened over this threshold. Furthermore this performance drop was associated with the severity of episodic memory deficits: the more severe the deficit, the poorer their timing performance. This seems to suggest that the hippocampus may be required for the representation of longer intervals rather than shorter ones, perhaps relating to the role of the hippocampus in episodic memory (Tulving, 2002). Perhaps even indicating that episodic memory and/or the episodic buffer may take over for intervals of 20-30 seconds and over.

The role of the hippocampus is further clarified by work with patients who have had sections of their hippocampus resected. Specifically, unilateral resections demonstrated that resection of the right hippocampus leads to timing deficits while resection of the left either results in no differences compared to controls or less severe deficits when compared to those with right hemisphere resections (Vidalaki, Ho, Bradshaw, & Szabadi, 1999). This suggests a right bias for the representation of time within the hippocampus.

While such work seems to suggest the functional role of the hippocampus would be in ongoing processing others have suggested it could instead relate to retrieval (Melgire et al., 2005). Compatible with such a role, the hippocampus was recently found to be active in a study that used two reference intervals in place of a single reference interval (Harrington et al., 2004). Imaging was performed as participants timed a candidate interval; their task was to indicate which reference interval the candidate interval was closest to. The areas of activity were similar to those of previous timing tasks except for the point where the candidate interval approached the point of bisection between the two reference intervals. At this point an increase in hippocampal activity was observed, presumably reflecting the retrieval of memory for the longer reference interval.

However, regardless of its specific role, the hippocampus appears to be less important than the frontal lobes in prospective timing tasks (McFarland & Glisky, 2009). Older adults were compared to younger control participants and amongst each other. Among each other they were scored on the basis of function scores for the frontal lobes and medial temporal lobe. Not surprisingly, lower scores in either category were associated with poorer performance on the prospective timing task. However the medial temporal lobe based deficit was only observable in the participants who had high frontal lobe functioning scores. That is to say that if the participants had low frontal lobe function scores their medial temporal lobe function scores had no impact on their results. This can be seen to be in line with either function suggested above. Namely, the medial temporal lobe could act as a spill over space or could be involved in recall. As a result the medial temporal lobes would fill a support role for the frontal lobes. As such, damage to the frontal lobes would have a large effect on timing whereas damage to the medial temporal lobe would have a reduced impact.

The frontal lobes. The results of McFarland and Glisky (2009) beg the question, what is the role of the frontal lobes? While a TBI study has suggested that the frontal lobes are associated with the speed at which time is perceived to pass (Binkofski & Block, 1996) the frontal lobes are more commonly associated with a role in working memory or attention and representations of suprasecond, rather than subsecond time (Koch, Oliveri, Carlesimo, & Caltagirone, 2002; Mangels, Ivry, & Shimizu, 1998). Typically activation is observed in the dorsolateral prefrontal cortex (DLPFC), is biased to the right, and increases with task demands (Lewis & Miall, 2006). While the DLPFC is sometimes associated with times in the subsecond range (Lewis & Miall, 2006) it is more commonly and more strongly associated with times in the suprasecond range (Lewis & Miall, 2003a, 2006). A finding that is backed up by lesion and TMS evidence showing reduced frontal lobe functioning is typically associated with deficits in suprasecond but not subsecond timing (Koch et al., 2002; Koch, Oliveri, Torriero, & Caltagirone, 2003; Mangels et al., 1998).

At a theoretical level the frontal lobes are typically thought to be associated with the central executive (Baddeley, 2003). This ties in nicely with the idea that the frontal lobes have a role in attention as the central executive can be thought of as allocating resources, or attention, to tasks of interest. In combining this with a STT account the frontal lobes may play the role of the attentional switch causing the collection or neglect of clock ticks that may then pass into an accumulator.

In truth though, time work with activation in the frontal lobes tends to show activation in a number of areas (Lewis & Miall, 2003b). While this is compatible with the idea of the frontal lobes being related to the central executive and/or attention, focussing on the frontal lobes in isolation is largely unproductive. In order to understand their role it is better to consider their place in a system.

Timing systems and lateralisation. Consideration of systems lends itself to the consideration of a bigger picture. While lateralisation of specific regions in timing function have been observed in some of the experiments covered so far, the question of whether or not lateralisation exists at a larger scale becomes interesting, especially given Miller's (1996) proposition that brief intervals may be encoded in the left hemisphere and longer intervals the right hemisphere. This proposition is particularly intriguing given the existence of recent evidence based theories speculating about right and left lateralised representations of suprasecond and subsecond time (Coull & Nobre, 2008; Lewis & Miall, 2003b).

The right hemisphere. Given the right hemisphere's association with space and the proposed association of space with time, it seems logical to suppose that the right hemisphere may house time. This appears to be the case in the studies that identify lateralised activity covered so far and support general findings that report a right hemisphere benefit for time measurement (Battellil, Walsh, Pascual-Leone, & Cavanagh, 2008; Coslett, Shenton, Dyer, & Wiener, 2009; Lewis & Miall, 2006). There is additional support as well. At the general level Coslett et al. (2009) investigated time perception at the suprasecond level in a number of participants with various lesions. Participants with lesions in the right hemisphere were more impaired on time judgement tasks than those with lesions in the left hemisphere. Furthermore, lesions to the right parietal lobe were particularly troublesome. This particular aspect of the results is consistent with evidence that shows timing deficits occur when TMS is applied over the right but not the left parietal lobe (VanRullen, Pascual-Leone, & Battelli, 2008).

Under such a system account the core components appear to be the DLPFC and the right parietal lobe. While the right parietal lobe is often associated with attentional deficits (Danckert et al., 2007), this system has also been associated with the visuo-spatial sketchpad (Baddeley, 2003), a so-called self-projection network (Buckner & Carroll, 2006), and the proposal of a 'when' pathway (Battellil et al., 2008). Most importantly for the current purposes though, this right lateralised system is preferentially activated by suprasecond rather than subsecond time (Lewis & Miall, 2003b).

The left hemisphere. While the right hemisphere is predominantly associated with suprasecond times the left hemisphere has been shown to have an association with subsecond times. This is particularly evident in chronometric counting tasks (Gruber, Kleinschmidt, Binkofski, Steinmetz, & von Cramon, 2000; Hinton et al., 2004), collision judgement tasks (Assmus, Marshall, Noth, Zilles, & Fink, 2005; Assmus et al., 2003), temporal attention (Coull & Nobre, 1998), time judgement tasks (Lewis & Miall, 2003a, 2006; Nicholls, 1996), and sequence detection tasks (Nicholls, 1996)[4]. In particular, activation is typically seen in the supramarginal gyrus, or more

[4] Nicholls (1996) reviews a number of findings based on experimental evidence or work with people with lesions to the left hemisphere. While most of the work reviewed demonstrates a left hemisphere benefit for subsecond times, some studies show a benefit for the right hemisphere. Additionally, one study shows a left hemisphere benefit for suprasecond times. At face value these results are inconsistent with the right and

generally in the left parietal lobe, and is modulated by task difficulty (Assmus et al., 2005; Assmus et al., 2003).

Of particular interest is the temporal attention task employed by Coull and Nobre (1998). Here a temporal task was compared to a spatial task. The intervals ranged from 300-1500ms and fell comfortably within the subsecond range. Whether the spatial and temporal tasks were performed separately or concurrently, fMRI showed separate areas were activated by spatial and temporal stimuli. Furthermore, when tasks required attention to both space and time, the spatial attention task activated areas associated with suprasecond time while the temporal attention task activated left lateralised structures. This was argued to demonstrate that these two systems may be separate.

Both hemispheres. However, despite this lateralisation some common activity exists. In particular, subsecond times activate pre-motor areas bilaterally (Lewis & Miall, 2003b), believed to be a product of a reliance on motor timing for brief durations, and difficult timing tasks have shown a tendency to recruit structures from both hemispheres (Lewis & Miall, 2006). While a structural division would help to simplify matters, bilateral activation shows that the divisions in time are made on a more functional basis in line with the splits based on implicit vs. explicit timing (Coull & Nobre, 2008) and automatic vs. cognitive timing (Lewis & Miall, 2003b) outlined earlier.

Imaging Summary. Overall, the areas of activity observed so far seem fairly compatible with dominant theories of time perception. Notably though the consistency in the regions found to be active is compatible with an STT account. The cerebellar and basal ganglia activation outlined earlier suggests these structures could house the pacemaker/oscillator of the STT theory. The accumulator appears to involve working memory and, potentially, the hippocampus. Working memory is implicated through the involvement of the DLPFC (Koch et al., 2003; Lewis & Miall, 2006) and the right inferior parietal cortex. The hippocampus is implicated at longer time intervals, perhaps through the episodic buffer coming into play when working memory's standard capacity is exceeded (Baddeley, 2000). However, while the work covered so far does provide a good overview of general time perception the question of the grounds for a relationship with space remains.

---

left hemisphere divisions outlined here. However, it should be noted that these studies use presentation of auditory stimuli to the left or right ear to determine whether a benefit for right or left hemisphere processing exists. The right and left ear project to both hemispheres (Bear et al., 2007), reducing confidence in the ability of the studies to separate right and left hemisphere processing. When this is combined with the fact that these findings were rare in the Nicholls (1996) review, contradict the majority of finding in the wider literature, and divisions between times served by the left and right hemispheres are based on factors other than duration, including task demands (Coull & Nobre, 2008; Lewis & Miall, 2003a, 2003b, 2006), concern over these contradictory findings is reduced.

The Link

Potential overlaps between time and space are becoming a popular area of research (Bonato, Zorzi, & Umilta, 2012; Walsh, 2003). Over time, interesting research has come from a number of sources including case studies, experiments with children, linguistic analysis, and behavioural experiments. While case study evidence is interesting, and will be considered here, linguistic analysis and experiments with healthy controls provide clearer and more widely applicable evidence. In particular, some of the most compelling evidence in support of a time space relationship comes from the simplest behavioural research itself.

Neuropsychological evidence. If time and space are linked it is reasonable to suggest that they should also use the same brain regions. It then follows that lesions impairing spatial processing should also impair time processing. In support of this it has been noted that lesions impairing only time or only space are rare (Critchley, 1953). Furthermore, much like space, time is preferentially disrupted by damage to the right hemisphere (Coslett et al., 2009). More specifically a stereotypically spatial lesion also shows temporal effects.

Neglect is associated with lesions to the right inferior parietal lobe (Danckert et al., 2007). It leads to neglect of the left side of space and can be object or environment centred. Popular examples of neglect involve people ignoring a side of their body, failing to eat from one side of the plate, failing to draw the left half of an image, and being biased to bisect lines far to the right of centre. What is less frequently reported, and apparently less frequently investigated, is their difficulty in measuring time observed in comparison to controls and people with right hemisphere damage that is not associated with the parietal lobe (Danckert et al., 2007). At a basic level patients have shown difficulty representing ideas related to the past, typically thought to occupy the left of space, compared to the future, typically thought to occupy the right of space (Saj, Fuhrman, Vuilleumier, & Boroditsky, 2014). Deficits have also been shown for both the subsecond and suprasecond estimates and are consistent with either an increased clock speed or damaged memory of interval lengths (Danckert et al., 2007). Furthermore TMS to the right parietal lobe has replicated these deficits (Oliveri et al., 2009) and helped to confirm the association between timing and this area. However the degree of impairment neglect patients showed in time judgements was not associated with the degree of spatial neglect symptoms that patients were showing (Danckert et al., 2007). This either suggests that time and space are represented by separate systems in a similar area or suggests that time perception is relatively more or less robust to damage than spatial perception.

Children. While most adults seem to have no issue separating time and space young children have been seen to make no distinction between time and space in their use of language (Piaget, 1970). Across a number of experiments children under the age of 7, those typically classified as preoperational, were seen to use spatial concepts to make inferences about how much time had passed. Specifically, longer distances were estimated to have taken more time regardless of the actual duration and fast runners were estimated to take as much time to cover a set distance as slow runners.

More recent work (Lourenco & Longo, 2010) supports these findings by showing that 9 month old infants transfer learned associations between spatial stimuli and duration. Specifically, if an arbitrary stimulus is paired with a spatially larger line, infants habituate to its pairing with a longer duration in a subsequent task more quickly than if it is paired with a shorter duration. This link between time and space is also demonstrated more directly by work showing that 9 month old infants habituate to stimuli where spatial and temporal length match (i.e. both are long or short) than those that mismatch (i.e. one is long and the other is short) (Srinivasan & Carey, 2010). Importantly, generalisation was shown from space to time but not from space to the amplitude of sounds (Srinivasan & Carey, 2010). This specificity increases confidence in the potential for a special relationship between time and space.

The general population. If ATOM, or any theory suggesting that time and space are related, is correct overlaps between time and space should be widely evident in research literature. Furthermore, such overlaps should not be restricted to a special subset of the population, such as those with right hemisphere lesions or children; overlaps between perception of time and space should affect everyone. While some general overlaps, such as the right hemisphere dominance for the representation of space and suprasecond time, have already been mentioned many more exist. In order to give a fair account of the evidence in support of ATOM and general time as space theories a number of areas of overlap need to be considered. In doing so it makes sense to start with the broadest example: language.

Language. While not an experimental measure, language does begin to give some hints to the way time may be conceived. In particular, the frequent use of metaphors that rely on space to convey information about time (e.g. moving meetings *around* and meeting *at* a time *on* a specific day) compared to the allegedly infrequent use of space as time metaphors (e.g. light-years) has been used as evidence to suggest that our understanding of time is built on top of our understanding of space (Casasanto & Boroditsky, 2008; Moore, 2006). While such an account is clearly open to criticism, as the existence of metaphor is common across domains and has the very practical ability to allow abstract ideas to be discussed using concrete language, some support does exist. Although the discrepancy between the quantity of time as space vs. space as time metaphors has not been quantified, the proposed direction of the imbalance is in line with what would be expected if time were to be constructed on top of representations of space. An imbalance favouring space over time is also compatible with Piaget's (1970) finding that children understand space before they understand time and report time as if it is space when requested. Additionally, a common tendency to develop spatial metaphors for time can be seen in the presence of such linguistic relationships in English, Chinese, Wolof, Japanese and other languages (Moore, 2006). Furthermore, the way in which language is experienced has been seen to impact on another task commonly used to demonstrate the link between time and space: spatial mapping tasks.

Standard spatial mapping. Spatial mapping tasks are simple experiments that can be conducted with members of the general population. Being a behavioural investigation they have the advantage of producing results that are more quantitative than arguments based on language based metaphor alone. However, similar to language based metaphor arguments, all spatial mapping tasks operate off the assumption that time has been mapped onto a spatial representation, typically one where time progresses from left to right.

One example is a study by Vallesi, Binns, and Shallice (2008). Operating under a belief that time may be represented according to space, they looked to the literature for examples of instances where other concepts were spatially mapped. Sequence presentation was identified as a spatially mapped concept. In sequence presentation studies (e.g. Dehaene, Bossini, & Giraux, 1993) participants are required to respond to a sequence of numbers. They are faster at responding to smaller numbers with their left-most response option while larger numbers are preferentially responded to with the right-most response. If time is represented in a similar fashion, one side of space should show a preferential response for short intervals and the other a preferential response for long intervals.

In the Vallesi et al. (2008) study, participants were first presented with a reference interval. Participants were then presented with a number of additional intervals that were longer or shorter than the reference stimulus. They had to indicate whether the candidate interval was longer or shorter than the reference by pressing one of two buttons counterbalanced across participants. In initial experiments one button was assigned to the left hand and one to the right. As expected a response preference was found. Specifically, shorter intervals were responded to more quickly with the left hand while longer intervals were responded to more quickly with the right hand. In order to determine whether this was due to a response preference for the hands or a response preference based on the spatial layout of the buttons this was repeated with participants crossing their hands. If the preference was hand based the pattern should remain hand specific. If it was space based the hand relationship should reverse. The hand relationship did reverse suggesting a space based representation. Finally, to rule out the possibility of some difference between the hemispheres the task was isolated to a single hand. Once again the left most response was fastest for shorter intervals while the right most response was fastest for longer intervals.

The results of Vallesi et al. (2008) were consistent with the idea that participants were representing time from left to right. Vallesi et al. (2008) note that this is consistent with the way in which time is usually presented on graphs and timelines as well as being consistent with the way that their participants read and write. In fact, in line with both of these findings, Vallesi et al. (2008) point out that similar tasks requiring participants to construct spatial representations of concepts including time produce results consistent with reading direction when conducted in other languages (Tversky, 2003). Specifically, Arabic participants produce temporal relationships from right-to-left, English speaking participants produce temporal relationships from left-to-right (Tversky et al., 1991), and Mandarin speakers have been shown to produce temporal relationships from top-to-bottom (Boroditsky, 2001). This cultural variation, combined with the finding that this relationship can be altered to fall in line with that of another culture through training (Boroditsky, 2001), would seem to suggest that the direction of this relationship is not innate or fixed.

Furthermore, the association is not restricted to a single task type. Shifts in spatial attention can also lead to differences in time estimation. Frassinetti, Magnani, and Oliveri (2009) used prismatic adaption, a technique used in the treatment of neglect that essentially forces spatial attention one way or the other, with time bisection and reproduction tasks. By having participants reproduce or bisect time intervals before and after adaptation then comparing the results an effect of shifting attention was found. Time estimates made when attention was shifted to the left were shorter than those made when attention was shifted to the right. Similar results have also been found in healthy participants whose spatial attention, measured through a line bisection task, is naturally biased to one side (Zach & Brugger, 2008) as well as through shifting spatial attention without the use of prismatic adaptation (Di Bono et al., 2012; Vicario et al., 2008). Another experiment showed the reverse scenario: time itself could be used to shift spatial attention in a similar fashion (Weger & Pratt, 2008).

Associations with space, time, and numbers. Critically, as ATOM is a theory of magnitude rather than a theory of an association between time and space only, relationships between time, space, and number should all be observable. This is the case (Casarotti, Michielin, Zorzi, & Umilta, 2007). Specifically, each domain is found to influence number processing and be influenced by number processing in turn: smaller numbers lead to smaller time perceptions while larger numbers lead to larger time perceptions.

Another finding that is predicted by ATOM is that of a single master clock for space, time, and numbers (Droit-Volet, 2010). Here click trains commonly used to speed up perceived time were also used for space and number judgements. In both cases shifts similar to those observed in time were produced. Finally ATOM also predicts the results of the Lourenco and Longo (2010) study with 9 month old infants. While the association between time and space has already been mentioned, Lourenco and Longo (2010) found an additional overlap between each of these and number was also found. Critically ATOM is the only theory covered that can account for these results.

Garner interference. Garner interference tasks (Garner, 1974) are used to investigate the relationship between two dimensions, such as height and width (Cant, Large, McCall, & Goodale, 2008) or time and location (Dutta & Nairne, 1993), in a given stimulus. The tasks themselves have two key divisions that are presented in separate blocks: the baseline condition and the filtering condition. In any given block participants are required to respond to changes in one of the dimensions while ignoring variation in the other dimension. In the baseline condition stimuli only vary on the dimension of interest. In the filtering condition stimuli vary on the dimension of interest as well as the irrelevant dimension. Garner interference refers to a drop in performance in the filtering condition relative to the baseline condition. This drop occurs when the two dimensions are integral, meaning that their representations are linked, but not when they are separable from each other. Critically this has been observed with time and space (Dutta & Nairne, 1993). In this study 2 stimuli were presented on each trial. The stimuli were a square and a circle. They could appear at one of two time points, 2 seconds into the trial or 3 seconds into the trial, and one of two locations, above or below a fixation cross. After each trial subjects were shown a square or a circle. They were then asked to indicate whether it was first/second in the sequence or above/below the fixation cross. Interference was found in both cases. This suggests that time and space are integral, non-separable representations.

While this is compatible with ATOM it should be noted that it is unclear whether participants are responding to sequence or duration in this experiment. If subjects are responding to sequence it is arguable that they may be spatialising the time component in the same way that sequence is spatialised when reading: earlier objects on the left and later objects on the right. If spatialisation is occurring the integral relationship with space is what would be predicted: space would be interfering with itself.

Kappa and tau effects. While spatial mapping and Garner interference tasks can show a general relationship between time and space it is difficult to show any direct impact of space on time, or vice versa, as responses are typically restricted to one of two options. A more direct means of testing for a relationship between time and space would involve the reproduction of each dimension. Representations of space and time stimuli of known size would then be free to interact with each other and graded effect sizes could be recorded. This has been achieved with studies on kappa and tau effects (Casasanto & Boroditsky, 2008; Jones & Huang, 1982; Sarrazin, Giraudo, Pailhous, & Bootsma, 2004). In such studies each stimulus contains spatial and temporal information. Commonly this information is provided through the sequential presentation of two stimuli, referred to here as dots. The appearance of the first dot marks the beginning of the temporal interval and one end of the spatial length. This first dot typically disappears and a second dot appears at the end of the interval and at the other end of the spatial length. Judgements, typically involving reproduction, of the length of space or time displayed are then required. Kappa effects are said to occur when varying the space presented influences judgements of time. Tau effects are said to occur when varying the length of time presented influences judgements of space. Generally an increase in the irrelevant dimension leads to an increase in the estimate of the relevant dimension although reverse kappa and tau effects do occur in some participants (Collyer, 1977).

While this points to a relationship between space and time, potentially consistent with the ATOM account, it has been noted that such relationships are likely to be based on inferences of imputed velocity (Jones & Huang, 1982). In the case of the tau effect a given spatial length paired with a brief time is associated with a faster speed than one paired with a long time. In the case of the kappa effect a given duration paired with a long line is associated with a slower speed than those paired with a short line. This perception of speed then influences the estimate of interest. Distances covered at faster speeds (shorter times) are perceived as shorter than those covered at slower speeds (longer times). Similarly, times covered at faster speeds (longer distances) are perceived as longer than those associated with slower speeds (shorter distances).

Most interestingly, while this type of experiment is typically restricted to the presentation of a line or a sequence of dots on a screen, comparable results have been found with time productions being made in the presence of more general differences of magnitude including shape (Xuan, Zhang, He, & Chen, 2007), a scaled model of a lounge (Delong, 1981), and a scaled mental image of a clock (Zach & Brugger, 2008). Shapes, models, and mental images of smaller scale were associated with briefer intervals being produced, essentially showing a kappa effect. The implication is that this particular paradigm is valid across different stimulus classes, thereby increasing confidence in its design and implications.

Subsecond versus suprasecond time. Before continuing it is worth noting that, unless otherwise specified, each of the areas of overlap has not considered differences between subsecond and suprasecond intervals. While the right lateralised network associated with suprasecond times makes it a clear candidate for space/time overlap, space/time overlap is also seen in the subsecond range (Coull & Nobre, 2008; Jones & Huang, 1982). In light of findings suggesting that subsecond timing tends to be associated with a left hemisphere network (Lewis & Miall, 2003b) this may be somewhat surprising. However it is important to remember that the left hemisphere vs. right hemisphere split is based on more than just the duration of the intervals in the task (Lewis & Miall, 2003b). In particular it is worth noting that each of the tasks covered favour an explicit timing or cognitive timing (Lewis & Miall, 2003b) approach in their completion. In support of this idea a study that showed time activated a left lateralised network also showed that time was represented independently of space (Coull & Nobre, 2008). Overall then it seems that the current idea of a shared representation between time and space should be restricted to one that focuses on cognitive, if not strictly suprasecond, timing.

Summary – The link. Overall the link between time and space appears to have a reasonable basis. While some discrepancies have been reported (Coull & Nobre, 1998) they are not incompatible with an account that considers a shared representation, especially if the shared representation is restricted to cases of cognitive timing. Under an ATOM or time as space account the association with the right hemisphere, and the large contribution of the right parietal lobe, provide a reasonable neurological substrate for the presence of a relationship between space and time. The impact of time on spatial attention, and the presence of tau effects, helps to establish the potential for a two-way relationship.

However, gaps remain. The most glaring being in retrospective timing where there are no clear examples of studies investigating a relationship between space and time. As has been mentioned in earlier sections this is likely to be a practical limitation in that it is more difficult to get an adequate number of trials in a retrospective timing task than a prospective timing task. When this is combined with the relatively low number of studies looking for associations between space and time the existence of such a gap is unsurprising.

In terms of prospective timing, research appears healthy with well-established paradigms producing reliable results. Among these paradigms the kappa-tau paradigm appears to address the nature of the relationship between time and space the most directly.

In the chapters that follow attempts will be made to investigate retrospective and prospective timing. In both cases the idea of an association between time and space will be tested. In the case of retrospective timing, methods from a directed forgetting study will be adapted to this purpose. In the prospective timing section kappa and tau effects will be investigated. A more targeted rationale will be provided for each approach at the start of the associated section.

Chapter 2. Directed Forgetting and Subsecond Times

Retrospective timing is the area of time research in the clearest need of methodological development. While a search on Web of Science for “time perception” returns over 1500 results, searching for the term “retrospective timing” within that drops the field to five results. While it is likely that many factors contribute to the current lack of retrospective timing research two likely contributors stand out. The first is the lack of a well-defined framework for understanding what retrospective time is and how it might operate. The second is the lack of a practical method for the investigation of retrospective time itself.

The most telling aspect of the lack of the framework is that two of the most recent studies addressing retrospective timing (El Haj, Moroni, Samson, Fasotti, & Allain, 2013; Ogden et al., 2011) point to a lack of work in the area and only cite two sources when supplying a theory of retrospective time. The first is a book by Ornstein (1969), the second, and the only one common to both studies, is a theoretical paper and meta-analysis of general time research from Block and Zakay (1997). While both suggest that the amount of information in memory is key to retrospective timing, with Ornstein (1969) focussing on the number of occurrences and Block and Zakay (1997) focussing on the complexity of the memory itself, neither provide the level of detail encountered in models of prospective timing. While this could be related to the idea that retrospective time is a “...timing without a timer...” (Ivry & Hazeltine, 1992, p. 184) approach, other factors are worth considering.

Perhaps the greatest deterrent to research, and therefore the greatest contributor to the lack of developed models, is the impracticality of current retrospective timing designs. As retrospective timing is about times that have occurred in the past, retrospective timing tasks require participants to be unaware that time is an important element of a given task until the task is over. This means that participants are essentially unable to run in more than one trial as, after they are asked about the time associated with the first trial, they become aware of the importance of time in subsequent trials. Once participants become aware of the importance of time they are likely to attend to time, thereby transforming the task into a prospective timing task (Block & Zakay, 1997). While experiments asking participants to provide retrospective estimates of multiple sub-events in a retrospective trial have occurred (i.e. Boltz, 1994; Grondin & Plourde, 2007) they are the exception. The most common form involves participants providing a single response to a single event, often in the range of minutes (i.e. Bisson, Tobin, & Grondin, 2012; Branas-Garza, Espinosa-Fernandez, & Serrano-del-Rosal, 2007; El Haj et al., 2013; Ogden et al., 2011; Ornstein, 1969).

This form of design, and its shortcomings, is neatly demonstrated in a recent study by Ogden et al. (2011). In this study 58 participants were divided into 3 conditions: placebo, low dose alcohol, and high dose alcohol. Participants were given their alcohol or placebo (tonic water) dose and then asked to complete a number of tasks. One of these tasks required participants to classify words as being presented in uppercase or lowercase letters. At the end of this case judgement task they were asked to estimate the duration of the task. This provided one estimate per participant. They then went on to perform two prospective timing tasks consisting of 72 and 68 trials respectively[5]. The study found that alcohol had an effect on prospective timing but not retrospective timing. As a result the researchers concluded, in line with Ivry and Hazeltine (1992), that retrospective timing is timing without a clock.

However there are issues with this finding. The first of these, high variability, is indicative of the retrospective field as a whole. Given that a standard retrospective timing task, including the one used in Ogden et al. (2011), relies on a single estimate of time per participant the true mean estimation of the individual participants cannot be established. Furthermore, measures of variation in estimates within a single participant cannot be calculated at all. This leaves observations open to influence by outliers, creating uncertainty at the participant level that goes on to contribute to uncertainty at the group level. For the Ogden et al. (2011) data, with a total of 58 observations, a strong claim about the true population mean is difficult to have faith in even assuming no alcohol dose effect exists. On the other hand if an alcohol dose effect does exist group means should be considered separately. In such a case, with approximately 19 observations per group, confidence in the estimated mean drops further.

The issue of high variability is further illustrated in Ogden et al. (2011) by the fact that there is a consistent trend in the retrospective data that does not reach significance in their chosen analysis[6]. If the trend were real it would suggest that increasing doses of alcohol are correlated with decreasing retrospective estimates of duration. While it does contradict the effect of alcohol seen in their prospective timing tasks, contradictions between retrospective and prospective timing have been observed before (Predebon, 1996) and would not be too problematic to address.

Finally there is the issue of the duration itself. In this study the target duration was over 4 minutes (255 seconds) long. This goes beyond the standard suprasecond range and is well beyond the prospective times used in the prospective tasks in their experiment (which, coincidentally, fall predominantly in the subsecond range). This long duration approach is fairly common in

---

[5] Arguably the imbalance between sampling in the retrospective and prospective tasks could be remedied by only analysing the first trial of each prospective task.

[6] The authors chose to test for a difference between groups using ANOVA. An alternative would have been to test for a correlation between alcohol dose and group mean time estimate. In this case the relationship does reach significance ($r<-.99$, $p=.012$) despite having only 1 degree of freedom.

retrospective work (for example Bisson et al., 2012; Branas-Garza et al., 2007; El Haj et al., 2013; Unrugneervoort, Kaiser, Coenen, & Vanluijtelaar, 1991).

In order to overcome these limitations a new method of investigating retrospective timing would be beneficial. Ideally it should focus on times in a range that is comparable to typical prospective experiments. Furthermore, for the present thesis, it would be preferable if this new approach could include a spatial dimension as well in order to address the notion that time and space might be related. The key to identifying this may lie in examining a key component of retrospective timing identified by Block and Zakay (1997) and Ornstein (1969) among others: its reliance on memory.

Episodic memories are said to consist of three core components; information on what, when, and where (Tulving, 2002)[7]. In the context of exploring theories regarding relationships between time and space episodic memory is an encouraging start given that time and space are respectively included in the when and where components. Under this conceptualisation the remaining component, what, provides the main aspect of the memory with when and where potentially acting as information tags or cues.

Crucially any new task must be able to include a temporal component without drawing attention to it. Preferably it would be a task that has shown an ability to reveal associations with something that also has not had attention explicitly drawn to it. Given the current interest in relationships between time and space, and the notion that time and space both act as information tags in episodic memory, it would be preferable if a spatial version of such a task could be found. One potential candidate is a directed forgetting task.

Experiment 1A

Directed forgetting tasks come in two main forms: the list method and the item method (Basden, Basden, & Gargano, 1993). Each involves the presentation of a list of stimuli and the presentation of cues to remember or forget associated stimuli during an initial study phase. In list method directed forgetting the stimuli are divided into two lists with the instruction to remember or forget occurring after each list is presented. In item method directed forgetting a remember or forget cue is presented after each stimulus. In each case the success of forgetting is tested through a memory test where participants are typically asked to recall or recognise as many stimuli as they can, independent of the memory instruction that was given during the study phase (Basden et al.,

[7] While Tulving (2002) does list subjective re-experience as a fourth component it was added after the what, where, and when components. Unlike the other three, subjective re-experience is more of a quality, the sensation of being there in the memory, and cannot be truly objectively assessed.

1993). In recall tests participants typically write down as many words as they can remember without prompts. In recognition tests participants are presented with one word at a time and asked to indicate whether the word was present in the study phase of the experiment (an old word) or was not present in the study phase of the experiment (a new word). Foil words, not previously shown in the study phase, are intermixed with words that were present in the study phase in these cases. If participants have successfully forgotten the forget-cued words their performance on the memory test should be worse for forget-cued words than remember-cued words. Critically this effect is demonstrated for recognition and recall tests with the item method but only with recall tests for the list method. In the list method a recognition test is said to rescue the forget-cued information (Basden et al., 1993).

In a sense the rescue effect described above reinstates the 'what' in the what, where, when puzzle. In fact, in the case described above, it is the reinstatement of the very piece of information that is required. While the focus of such tasks on these elements alone has led to some criticism (Tulving, 2002) evidence supporting a benefit for the reinstatement of another piece of this puzzle has also been found.

Memory benefits for the reinstatement of where information are well documented. While a famous example is that of the SCUBA study (Godden & Baddeley, 1975), in which location information aided recall, a study more directly relevant to directed forgetting tasks has been performed. Small but significant benefits of spatial context reinstatement were observed using item method directed forgetting and a recognition test in a study that intended to test for an entirely different effect (Hourihan, Goldberg, & Taylor, 2007).

In an attempt to examine inhibition of return effects on remembering and forgetting during an item method directed forgetting task, Hourihan et al. (2007) varied the presentation location of items on screen. This was done by independently randomising word presentation to one of four locations during the study phase and the test phase causing some words to appear in the same location at study and at test. When memory performance for words appearing at the same location at study and test was compared to performance for words that appeared in different locations at study and test, an increase in the accuracy of responses was found for forget-cued words in the same location condition relative to the different location condition. This suggested a memory benefit compatible with a reinstatement of where information for forget-cued words. Interestingly the effect was not present in the remember-cued words.

While the absence of an effect in the remember-cued words may seem odd at first it is important to remember that the rescue effect described earlier in the list method (Basden et al., 1993) also only occurred in the forget-cued lists. Furthermore, forget-cued words are believed to be less thoroughly encoded in item-method directed forgetting tasks than remember-cued words (Basden et al., 1993; Hourihan et al., 2007). Specifically, forget-cued words are thought to have gone through less elaborative rehearsal and, as a result, are paired with fewer memory cues (Basden et al., 1993; Hourihan et al., 2007). As such the benefit associated with the presentation of an associated memory cue or contextual element should be relatively larger for forget-cued words than remember words. While a multitude of other explanations exist this relative benefit account has been deemed to be the most likely (Hourihan et al., 2007). Most importantly, in terms of informing the current design, such an account suggests that the forget condition may be particularly important for revealing the benefits of reinstating spatial context.

As this study meets the criteria mentioned earlier (a task with multiple trials that shows an association with space in memory without specific instructions to attend to space) it is a good candidate for repurposing toward the investigation of retrospective timing. In the present case the presence of a time effect could be tested for by replacing the spatial locations used in Hourihan et al. (2007) with temporal durations (or temporal *locations*). That is, by associating words with an absolute location in time (e.g. 500ms after the onset of the fixation cross) rather than a location in space (e.g. above the fixation cross), memory benefits for forgot cued items should be observed when time locations match at study and test compared to when they do not match[8].

While this may seem a simple task the first hurdle is deciding which kind of time should be investigated first. As has been noted, time in research is far less uniform than time in experience. In addition to the split between prospective and retrospective time, its divisions into micro and macrotime (Hassabis & Maguire, 2007), automatic and cognitive timing (Lewis & Miall, 2003b), implicit and explicit (Coull & Nobre, 2008) and suprasecond and subsecond (Gibbon, Malapani, Dale, & Gallistel, 1997) timing muddy the waters.

While most of these divisions typically apply to prospective time a case can be made for them existing in retrospective time as well (Block & Zakay, 1997; Brown, 1985). At the very least the idea for the time between events being represented does exist with divisions between microtime and macrotime (Hassabis & Maguire, 2007). While microtime, being the time between events in memory, should include subsecond and suprasecond time the safest initial candidate is likely to be subsecond time. While it is linked to a left lateralised network in prospective time (Coull & Nobre,

[8] While spatial terms will be used for temporal concepts (e.g. temporal location) it should be noted that this does not pre-suppose the existence of a link between time and space. Instead, it is done to emphasize the similarity between the current temporal experiments and the spatial experiments of Hourihan et al. (2007). Throughout these experiments temporal location will be manipulated by varying the duration of the fixation cross.

2008; Lewis & Miall, 2003b), arguably making it less like space and therefore less likely to replicate the spatial effect seen in Hourihan et al. (2007), it is also associated with ideas of automatic (Lewis & Miall, 2003b) or implicit (Coull & Nobre, 2008) processing. As participants will not be instructed to time the intervals at any point in the experiment this makes an automatic or implicit timing system the most likely to be relevant.

If the task does use the subsecond system this should limit the time lengths that can be used. While the point of sub/suprasecond division has been suggested to occur at least as late as 3 seconds (Gibbon et al., 1997; Ulbrich, Churan, Fink, & Wittmann, 2007) this study will take a conservative upper limit of 1500ms based on qualitative differences observed at this point in an automatic timing study (Gilden & Marusich, 2009). The lower limit will be set at 500ms as it is above the region of common associated stimulus effects (Hackley & Valle-Inclan, 2003). Finally the time between 500 and 1500ms will be divided into 300ms blocks: 500, 800, 1100, and 1400ms. Such segregation exceeds the durations at which stimuli in a similar time range have been found to be confused with one another (Kristofferson, 1980). Additionally, a linear segregation of time has been shown to lend itself to balanced representations of time (Wearden & Ferrara, 1995).

To increase the chances of finding what is likely to be a small effect, especially given the length of the intervals being used and the fact that there will be no explicit direction to attend to time, attempts will be made to reduce barriers to the detection of real differences. The first will be through using signal strength as a primary variable of interest in addition to accuracy and reaction time. The second will be the use of a selection criterion.

Unlike accuracy and reaction time, signal measures attempt to remove underlying response biases in the results by reconceptualising the data. Where accuracy and reaction time measures typically focus on the trials where subjects provide a correct response, such as classifying an old word as an old word, signal measures contrast the tendency to provide correct responses with the tendency to provide incorrect responses, the tendency to classify a new word as an old word.

In any experiment where participants are asked to indicate whether a stimulus is present or not using yes or no answers there are four possible outcomes in every trial. In the current experiment present or absent equates to words that were presented at study (old words) and words that were not presented at study (new words). If a participant responds ‘old’ when the stimulus was presented at study it is a hit. If they respond ‘old’ when the stimulus was not presented at study it is a false alarm. If they respond ‘new’ when the stimulus was not presented at study it is a correct rejection. If they respond ‘new’ when the stimulus was presented at study they have missed that target. These categories are clarified in Table 2.1. By considering the relationship between the hits

Table 2.1
Classification of signal related terms in a new/old word recognition test.

		Stimulus Present at Study	
		Yes	No
Response	Old	Hit	False Alarm
	New	Miss	Correct Rejection

and the false alarms a measurement that should correct for a participant's tendency, or bias, to say yes or no is obtained.

Two such measures of signal are d' and A'. While d' is a popular option the number of trials in some conditions will often lead to the production of false alarm and hit rates of 0 and 1. These values are problematic for d' as d' takes on a value of infinity in these cases. Although corrections for the infinite value issue are possible (Hautus, 1995) it is preferable to use a measure that does not produce such issues with these values.

Owing to the way it is calculated, A' does not create infinite values when hit rates or false alarms take on values of 0 or 1. While other issues have been noted with A' these tend to focus on whether or not the claim that A' is a non-parametric measure of signal is valid (Rotello, Masson, & Verde, 2008) and/or are issues common to other signal measures (Rotello et al., 2008; Verde, MacMillan, & Rotello, 2006). As non-parametricity is not important to the current design these criticisms are of lesser concern. Furthermore, as A' has been recommended for use in old/new response tasks such as recognition memory paradigms (Donaldson, 1992) and has been recommended over an alternative when *N* is small (Verde et al., 2006) it will be used here.

Finally, only subjects showing a significant difference between the number of remember-cued and forget-cued words recognised at test will be included. This will be determined by chi-square analysis of the words they recognised. There are two key reasons to include this criterion. The first is that this is a way to check if participants are correctly following the instructions of the task. If they are performing the task as instructed there should be a difference in performance between the remember-cued and forget-cued conditions. The second is that Hourihan et al. (2007) only showed the effect in the forget-cued words. This implies that, in order to show the effect at all, participants must actually treat the forget-cued words differently to remember-cued words and their performance must be different between the memory cue conditions. Including the proposed criterion will ensure that this is the case.

Methods.

Participants. A total of 44 volunteers were recruited from the University of Auckland. All participants reported normal or corrected to normal vision and were able to hear and discriminate between a 260Hz and an 1170Hz tone. Participant handedness was measured using the Edinburgh handedness inventory (Oldfield, 1971). Following participation, accuracy scores for remember-cued and forget-cued words were screened for the presence of directed forgetting effects using chi-square analysis[9]. Participants who failed to show a difference in accuracy between memory instruction conditions were excluded from the analysis and replaced. All but 12 passed this criteria, leaving 32 participants (N_{males}=11, M_{age}=22.29, $Range_{age}$=18-27, 9 Right Handed, 2 left handed; $N_{females}$=21, M_{age}=22.29, $Range_{age}$=17-44, 20 Right Handed, 1 Left Handed). Use of participants in this experiment, and every subsequent experiment in this thesis, was approved by the University of Auckland Human Participants Ethics Committee.

Materials. The experiment was constructed in E-Studio and was run using E-run. The program was controlled by a computer with a Pentium II 450MHz processor, 128MB RAM, a 4MB video card, and a sound card. The computer was calibrated to supply millisecond accurate timing with this software. Visual stimuli (words and fixation crosses) were displayed on a 17 inch VGA CRT monitor set to a display of 1024x768 pixels with a refresh rate of 60Hz. The words (see Appendix A) were taken from Hourihan et al. (2007). All words were presented in the centre of the screen in black lowercase letters on a white background in size 18 Calibri font. The shortest word was 3 letters long (fly) and extended 0.7 degrees of visual angle. The longest word was 10 letters long (blackboard) and extended 3 degrees of visual angle. The original division into two lists of 64 items was retained. One list was used as a study list for each participant, the other as a foil list. Auditory stimuli, a 260Hz tone and an 1170Hz tone, were delivered via generic supra-aural headphones at a comfortable volume for participants. Responses were collected from the computer's keyboard. The "Z" and "/" keys were used.

The assignment of lists to study and foil conditions, tones to remember and forget-cue conditions, and response keys to old and new responses were counterbalanced across participants. Recency and primacy effects were controlled by having four buffer trials at the start and end of each study list (Hourihan et al., 2007). These four trials each contained a unique word that was not on the

---

[9] Each participant's combined accuracy scores for remember and forget words were used to predict participant specific expected values across remember and forget trials. Expected values assumed no difference between accuracy in remember and forget trials. Participant's with p>.05 were rejected and replaced.

study/ foil lists and were the same for every participant. Buffer trials were always followed by a remember-cue and were excluded from analysis. Their presentation order was randomised.

During the study trials a word, fixation cross duration, and memory-cue grouping was randomly selected from a list of possible combinations until all of the study words had been presented. As outlined in the introduction, time delay periods of 500, 800, 1100, and 1400ms were used. Each time location was paired with 16 study words (eight remember-cued and eight forget-cued)

During the test trials participants were sequentially presented with words from the study list and the foil list in a randomised order. When words were presented from the foil list the fixation cross duration was randomly assigned as 500, 800, 1100, or 1400ms with each duration being presented 16 times. When the words were presented from the study list each word was paired with a fixation cross duration period under two constraints. First, each fixation cross duration would be presented 16 times. Second, for each fixation cross duration at study each fixation cross duration would be presented equally as often at test. That is to say if the eight remember-cued words were paired with a fixation cross duration of 800ms, two would be tested with a 500ms fixation cross duration, two with an 800ms fixation cross duration, two with an 1100ms fixation cross duration, and two with a 1400ms fixation cross duration.

These times were then flagged according to the difference between duration of the fixation cross at study and the duration of the fixation cross at test (study-test). These temporal distance values could be -900ms, -600ms, -300ms, 0ms, 300ms, 600ms, or 900ms. Where the fixation cross presented at test was shorter than that presented at study the value was negative, where it was longer the value was positive.

Procedure. During the experiment participants sat 57cm from the computer monitor. They were told that the experiment involved a study phase which would be followed by a recognition memory test. Participants were instructed to remember items followed by a remember-cue and forget items followed by a forget-cue.

During a familiarisation phase participants were exposed to 5 remember-cue tones and 5 forget-cue tones in a randomised presentation order. During this phase remember-cue tones were paired with visual presentation of “High = Remember the Word” or “Low = Remember the word” as appropriate. Forget-cue tones were paired with “High = Forget the Word” or “Low = Forget the word” as appropriate. After presentation participants initiated the study phase of the experiment by pressing any key.

During the study phase participants were presented with four buffer trials, then 64 study trials, then a final 4 buffer trials. Trials began with a 9 second blank screen followed by a fixation cross whose duration was randomly assigned to one of the 4 time delays. The word was then presented for 1 second then the memory cue tone was presented during a 500ms blank screen. On half the trials word presentation was followed by a cue to remember, on the other half, a cue to forget. The final trial was followed by an additional 9 second blank screen. This combination of times preserved the inter-trial delay of the original experiment (Hourihan et al., 2007).

After the study phase participants were given instructions about the test phase. They were instructed to indicate, as quickly and accurately as possible, whether or not a word had been on the study list using the “ Z” and “/” keys. The assignment of these keys was counterbalanced across participants and participants received key assignment information specific to their condition. They were instructed to disregard whether the study list words had been followed by a forget-cue or a remember-cue (i.e. they were asked to remember the remember-cued words and the forget-cued words).

The 128 study and foil words were presented individually. Trials began with a 500ms blank screen followed by a fixation cross whose duration was randomly assigned to one of the 4 time delays under the constraints mentioned in the materials section. A test word from the foil or study list followed. This word stayed on the screen until a response was made. The procedure is summarised in Figure 2.1.

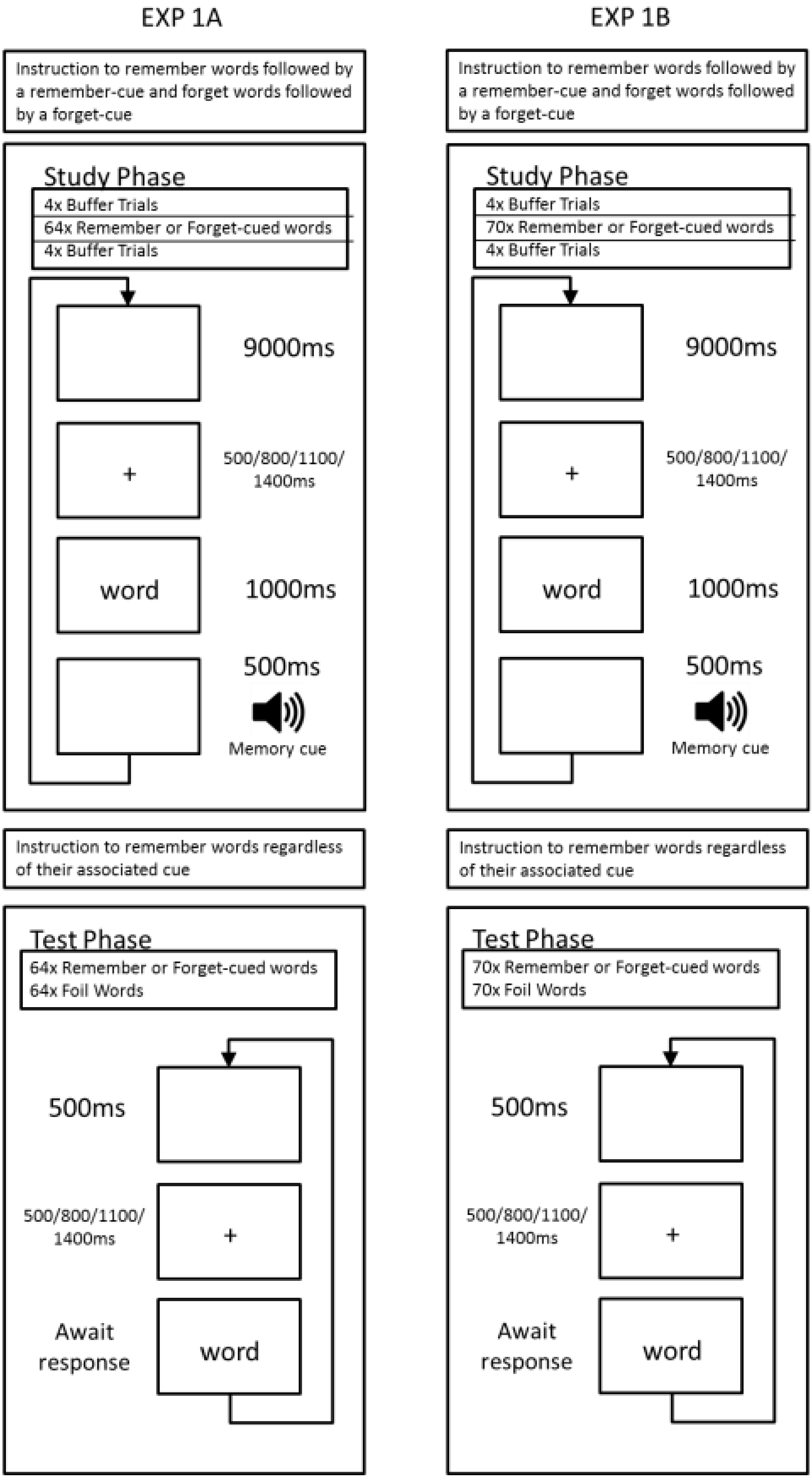

Figure 2.1. Trial procedure for experiments 1A and 1B. Stimuli are for illustrative purposes only and are not drawn to scale.

Results. Initial comparisons of reaction times for correct responses and accuracy across time and memory conditions were considered using separate $2_{(\text{memory cue})} \times 7_{(\text{temporal distance})}$ repeated measures ANOVAs. A significant main effect of memory was found for accuracy ($F_{(1,31)}$=161.60, p<.001) and reaction time ($F_{(1,9)}$=5.78, p=.040) but no significant main effect of or interaction with time was found in either case. This suggests that remember-cued words were responded to more quickly than forget-cued words and that remember-cued words were more likely to be identified as old words. In the case of the reaction time data the data of 22 participants was excluded due to accuracy values of 0 in one or more conditions. Data for accuracy and reaction time is shown as a function of temporal distance in Figure 2.2.

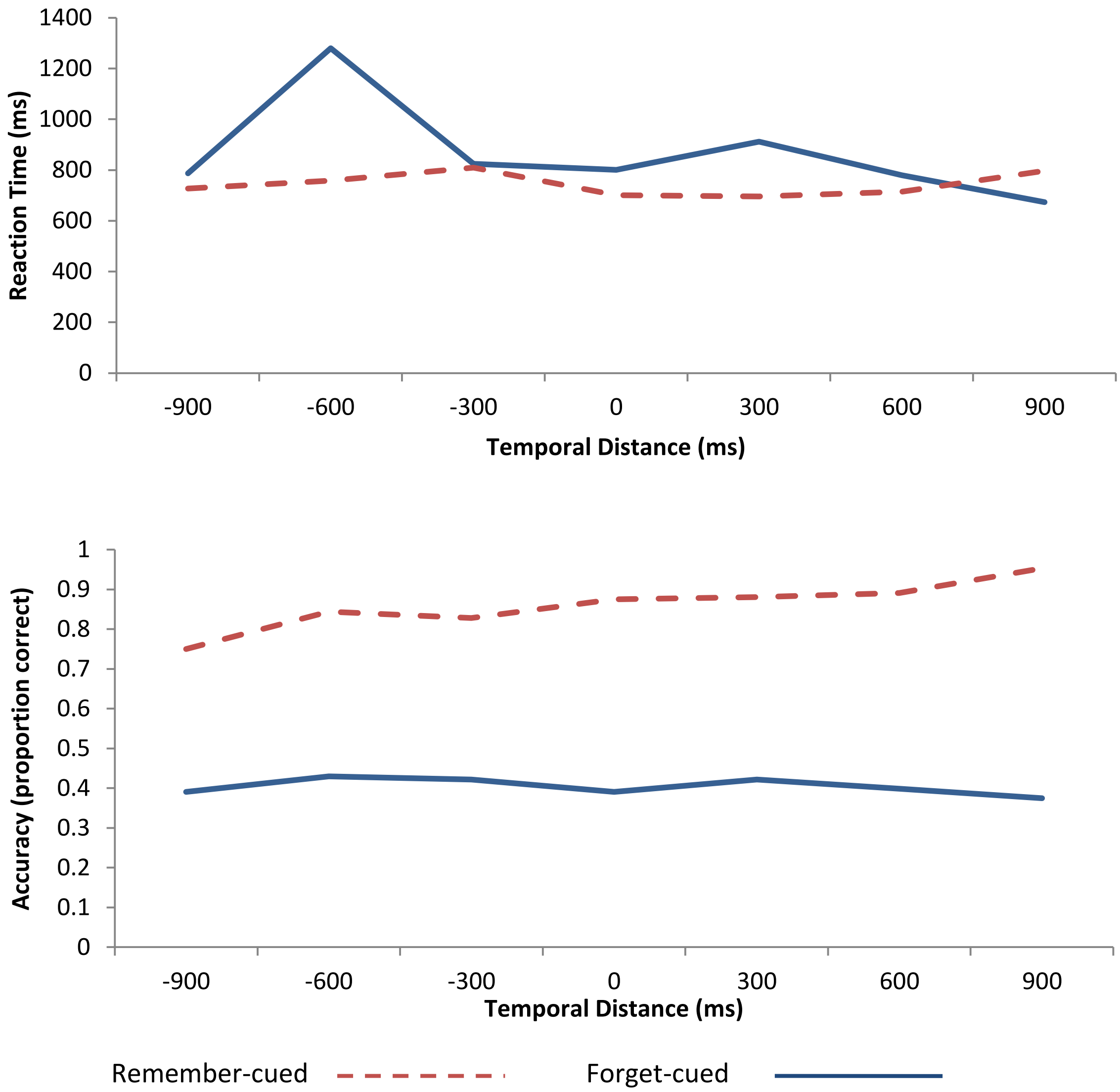

Figure 2.2. Reaction time and accuracy data for remember-cued (dashed red line) and forget-cued (solid blue line) words. The X axis represents the difference in time between the duration of the fixation cross at study and test. The false alarm rate in the foil condition was .12. The reaction time to foils was 962ms.

Signal strength was examined using A' (Donaldson, 1992)[10]. In order to provide a suitable noise comparison for each temporal distance at test the foils were treated in two different ways. Either the mean foil value for an individual participant was used at each condition from -900 to 900 or weighted contributions (see Table 2.2 for weightings) were used[11]. As similar patterns and significance values were produced regardless of the foil method used the values reported here relate to the mean foil values. Mean foil values were preferred as they involve less data manipulation. The same foil values were used for remember-cued and forget-cued words[12].

A $2_{(\text{memory cue})}$ x $7_{(\text{temporal distance})}$ repeated measures ANOVA performed on the signal data revealed a significant main effect of memory condition ($F_{(1,31)}= 68.78$, $p<.001$) suggesting that remember-cued words are recognised more easily than forget-cued words. A borderline significant interaction of memory condition and time delay was also produced ($F_{(6,186)}= 1.483$, $p>.05$,

Table 2.2
Temporal distance conditions with the associated fixation cross durations at test shown beneath them.

-900	-600	-300	0	300	600	900
500	500	500	500	800	1100	1400
	800	800	800	1100	1400	
		1100	1100	1400		
			1400			

[10] As A' compares noise to signal and only one measure of noise was taken the same noise data is used for remember and forget trials. This is believed to be valid as noise should be the same for both conditions. Furthermore the factor of interest is how signal changes across time difference windows for forget data rather than the comparison between remember and forget data. Such a method has also been used elsewhere (Defeyter, Russo, & McPartlin, 2009). The use of a bias free measure also removes the possibility of a speed accuracy trade off.

[11] As shown in Table 2.1 the distribution of the test delay periods for the study words varies across time differences. In order to make a corresponding foil comparison the contribution of each foil delay period was weighted to create a simulated foil -900 to 900 spread. Taking the -900 and 0 windows as examples the foil value in the -900 window was created by using only the 500ms foils. The foil value for the 0 time difference was created by using the average of all the foil values. This was done as an attempt to reduce the contribution of fluctuations related to test delay variations.

[12] While some readers may be concerned about the use of the same foil rate in both memory cue conditions, such use is justified. First, participants are asked to identify whether or not a word was present, not which list it was present on. Second if there is a difference in noise between remember and forget words it is more likely to penalise the forget words than the remember words. Third, if anything the noise is likely to have a better estimate than signal in this case. Fourth, if the data is reconceptualised with items correctly identified as new as signal this theoretical issue essentially disappears. Finally, such an approach is used in MacMillan and Creelman (2005).

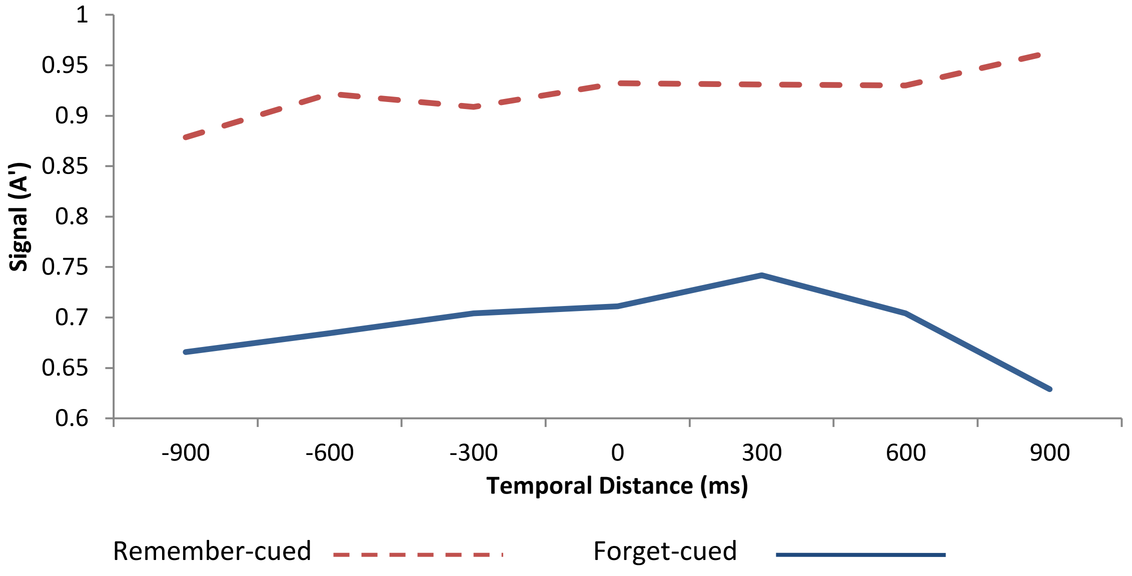

Figure 2.3. Signal strength (A') data for remember-cued (dashed red line) and forget-cued (solid blue line) words. The X axis represents the difference in time between the duration of the fixation cross at study and test.

Greenhouse Geisser = .75) suggesting that the effect of time difference may depend on the memory condition. Signal strength is shown as a function of temporal distance in Figure 2.3.

Planned contrasts revealed significant quadratic trends were for the main effect of time distance ($F_{(1,31)}$= 4.73, p=.037) and the interaction between memory and time distance ($F_{(1,31)}$= 4.56, p=.041). The quadratic main effect for time difference suggests that test times that are more different from study times are associated with lower signal values than those that are more similar to study times when collapsing across memory conditions. The interaction suggests that the effect of time difference depends on the memory condition. Here the difference between memory conditions decreases as the time difference gets closer to 0.

As planned contrasts revealed a significant interaction, and as separate predictions were made for remember-cued and forget-cued words, simple effects analyses were carried out. While an ANOVA of the forget-cued words did not produce a significant difference ($F_{(6,186)}$= 1.26, p>.05, Greenhouse Geisser = .75) planned contrasts did reveal a significant quadratic trend ($F_{(1,31)}$= 5.15, p=.030) accounting for 73% of the total variance. The quadratic trend suggests that signal strength is lowest when the time at study is the most different from the time at test and increases as it becomes more similar to the time at study. Furthermore if all of the remaining variance were loaded onto one orthogonal contrast the *F* value obtained would not reach significance ($F_{(1.31)}$=2.00, $F_{(critical)}$= 3.89) suggesting that this is the only significant trend that could be present in the forget data[13].

As the analysis above contains a linear term that the analysis itself suggests is unnecessary a quadratic trend without a linear term was fitted to the data. This was achieved by squaring the x values (the time differences) and using them to predict variance in the group averages. The slope of the resulting line provided the quadratic coefficient and is shown in Figure 2.4. The intercept provided the constant for the equation. This prediction accounted for 76% of the total variance in the means. Equations from this analysis are reported below. Time difference (t) is represented in seconds, deciseconds, and milliseconds to prevent the multiplier having too many zeroes. This trend would suggest maximal memory benefit for forget words at a time difference of 0.

$$\begin{aligned} &\text{Seconds:} && \text{Signal} = -0.09t^2 + 0.73 \\ &\text{Deciseconds:} && \text{Signal} = -9.4\text{x}10^{-3}t^2 + 0.73 \\ &\text{Milliseconds:} && \text{Signal} = -9.4\text{x}10^{-8}t^2 + 0.73 \end{aligned} \tag{2.1}$$

An ANOVA of the remember-cued words revealed a significant effect of time difference ($F_{(6,186)}$= 5.73, p<.001 Greenhouse Geisser = .64) with the contrasts showing a significant linear trend ($F_{(1,31)}$= 17.015, p<.001) suggesting that signal strength increased from temporal distances of -900 to 900 as shown in Figure 2.3.

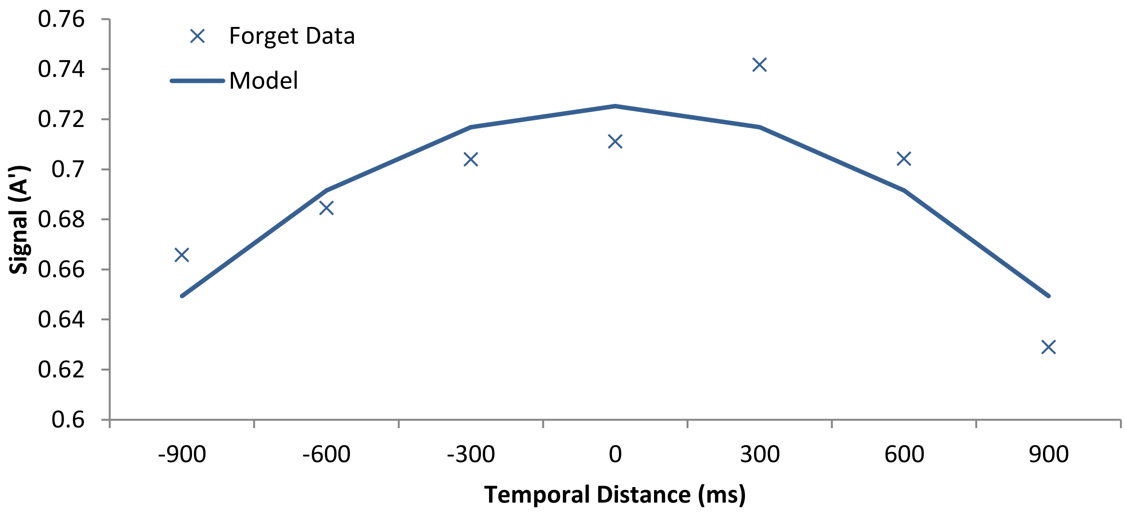

Figure 2.4: Signal strength (A’) for the forget-cued words plotted against time difference between study and test. A proposed quadratic model is presented as a comparison.

13 These assertions are made by using the pooled error rates and rules around orthogonal contrasts.

Analysis of the linear trend in remember-cued words suggested that it accounted for 78% of the total variance observed. If all of the remaining variance were loaded into a single contrast significant results would still be attainable suggesting other factors may make a valid contribution.

Overall these findings suggest that words that participants are cued to remember are associated with a higher signal value than those that they are cued to forget. Furthermore forget-cued words presented at similar times during study and test have higher signal values than those presented at different times during study and test. Signal for Remember-cued words increases as the difference between study and test times becomes more positive. Finally, the quadratic term suggests that the relationship between time at study and time at test could be symmetrical about 0 for forget-cued words.[14]

Discussion. Congruent with earlier spatial work (Hourihan et al., 2007), the current study demonstrated a difference between remember-cued and forget-cued words in terms of recognition strength and recognition patterns. In particular, forget-cued words presented in the same temporal location during the study and test phases were associated with a stronger signal than those presented at different temporal locations while remember-cued words were more recognisable than forget words.

In this case the study design meant that the distance between study and test locations was a graded difference in time rather than an absolute difference in space. This allowed for the plotting of the effect of memory benefit as a function of time distance rather than as a simple same different comparison. A significant quadratic trend was observed suggesting that, for the forget-cued words, when time was more similar at study and test the benefit was greater than when time was less similar at study and test.

While the trend in the forget-cued data could be the result of time acting as a memory cue an alternative interpretation would be that the trend is somehow related to the distribution of the time distances shown in Table 2.2. The case for the trend in Table 2.2 producing the observed data pattern is strengthened by the fact that the significant effect of temporal location is only found in the A’ data. Unlike reaction time and accuracy, the A’ data is the result of an additional calculation/manipulation being performed on the data. In this way a potential source for an artificial effect has already been identified: the A’ calculation itself. The contribution of the peaked distribution is explored in experiment 1B.

[14] These patterns do not hold when participants who did not show a directed forgetting effect are included in the analysis. Furthermore, when the excluded participants are considered on their own no significant effects are found. This can be argued to support their exclusions from the current analyses.

Experiment 1B

This experiment replicates experiment 1A with changes to the distribution of temporal distances. Temporal distances (e.g. -900, -600, etc) rather than time combinations (pairing all times at study with all times at test) will now be emphasised. This will result in a shift from the peaked distribution seen in experiment 1A to a flat distribution of temporal distances. If the effect is due to the difference between the time a word is shown on screen at study and the time it is shown on screen at test it should replicate. If it is due to the differing number of trials that results from the distribution of temporal distances used it should not replicate.

Methods.

Participants. 50 volunteers were recruited from the University of Auckland. All participants reported normal or corrected to normal vision and were able to hear and discriminate between a 260Hz and an 1170Hz tone. Participant handedness was measured using the Edinburgh handedness inventory (Oldfield, 1971). Following participation remember and forget scores were screened for the presence of directed forgetting effects using chi-square analysis[15]. If participants failed either of these they were excluded from the analysis. All but 18 passed this criteria leaving 32 participants (N_{males}=8, M_{age}=22.88, $\text{Range}_{\text{age}}$=19-35, 8 Right Handed; N_{females}=24, M_{age}=22.96, $\text{Range}_{\text{age}}$=18-37, 22 Right Handed, 2 Left Handed).

Materials. In order to shift from a peaked distribution of temporal distance to a flat distribution while keeping the average time at study the same the number of words in each memory condition was increased by 4. The number of words in the foil condition was correspondingly increased by 8. This gave a total of 36 possible time location combinations. As there are 7 temporal distances only 35 of these combinations were used. For each individual participant an 800-800 or 1100-1100 pairing was dropped from all conditions. While it would have been possible to balance the dropping within participants by assigning one combination to the remember-cued words and one to the forget-cued words this would have reduced the validity of comparisons made between remember-cued and forget-cued words. The new balancing is shown in Table 2.3.

The increase from lists of 32 items per condition to lists of 36 items per condition required an additional 16 items to be introduced. This was achieved by using the version of the MRC psycholinguistic database made available by UWA online (Coltheart, 1981; Wilson, 1988) to determine the scores of familiarity, concreteness, and frequency for each word from experiment 1A

---

[15] Each participant's combined accuracy scores for remember and forget words were used to predict participant specific expected values across remember and forget trials. Expected values assumed no difference between accuracy in remember and forget trials. Participant's with p>.05 were rejected and replaced.

Table 2.3
Flat distribution temporal distance conditions with the associated fixation cross durations at test shown beneath them.

-900	-600	-300	0	300	600	900
500	500	800	500	800	1100	1400
500	500	800	800	1100	1100	1400
500	500	800	1100	1100	1400	1400
500	800	800	1400	1100	1400	1400
500	800	1100	800* 1100*	1100	1400	1400

*Only one is presented to each participant. Presentation is balanced across participants

where available. Words that fell between the maximum and minimum scores produced for each criteria were then randomly drawn from the MRC psycholinguistic database and considered for inclusion as additional words in experiment 1B. If the drawn word was already on the list from experiment 1A or was emotionally charged (e.g. bomb) another word was drawn. The additional words are attached beneath the original words in Appendix A.

Procedures. With the exception of trial number and trial distribution outlined above, all procedures remained the same as for experiment 1A.

Results. As for experiment 1A accuracy and reaction times for correct responses were considered first using separate $2_{(\text{memory cue})}\text{x}7_{(\text{temporal distance})}$ repeated measures ANOVAS. As for experiment 1A a significant main effect of memory condition was seen in the accuracy ($F_{(1,31)}$=286.39, p<.001) and reaction time ($F_{(1,14)}$=27.70, p<.001) data with no other comparisons reaching significance. Plots of the accuracy and reaction time data are shown in Figure 2.5. In the case of the reaction time data the data of 17 participants was excluded due to accuracy scores of 0 in one or more conditions.

Signal strength was examined using A' (Donaldson, 1992)[16]. In order to provide a suitable noise comparison for each time difference at test the foils were treated in 2 different ways. Either the mean foil value for an individual participant was used at each condition from -900 to 900 or weighted contributions were used. When weighted contributions were used the weighting from Table 2.3 were used as appropriate for each participant. As similar patterns and significance values were produced regardless of technique used, and as the mean foil value involves less data manipulation, the values reported here relate to the mean foil value.

[16] As A' compares noise to signal and only one measure of noise was taken the same noise data is used for remember and forget trials. This is believed to be valid as noise should be the same for both conditions. Furthermore the factor of interest is how signal changes across time difference windows for forget data rather than the comparison between remember and forget data.

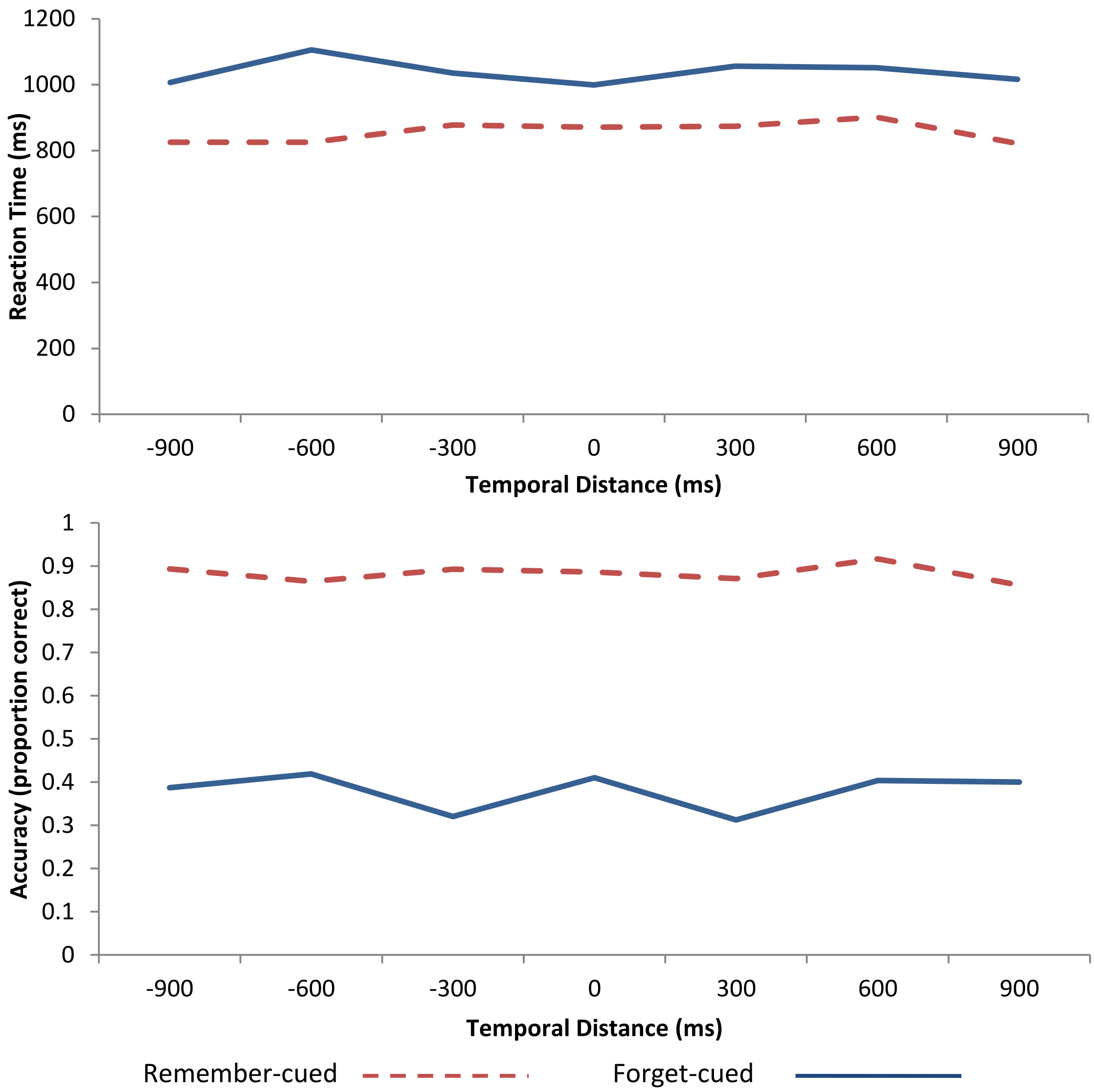

Figure 2.5. Reaction time and accuracy data for remember-cued (dashed red line) and forget-cued (solid blue line) words in the flat distribution design. The X axis represents the difference in time between the duration of the fixation cross at study and test. The false alarm rate in the foil condition was .07. The reaction time to foils was 908ms.

A $2_{(\text{memory})}$x$7_{(\text{temporal distance})}$ ANOVA found a significant main effect of memory condition ($F_{(1,31)}$= 60.82, p<.001). No other significant results were found. The resulting data pattern is shown in Figure 2.6.

Planned contrasts replicating those in the initial experiment showed a significant quadratic trend for the interaction of time and memory condition ($F_{(1,31)}$= 6.60, p=.015). Planned contrasts were then conducted on forget and remember data individually. No significant results were found.

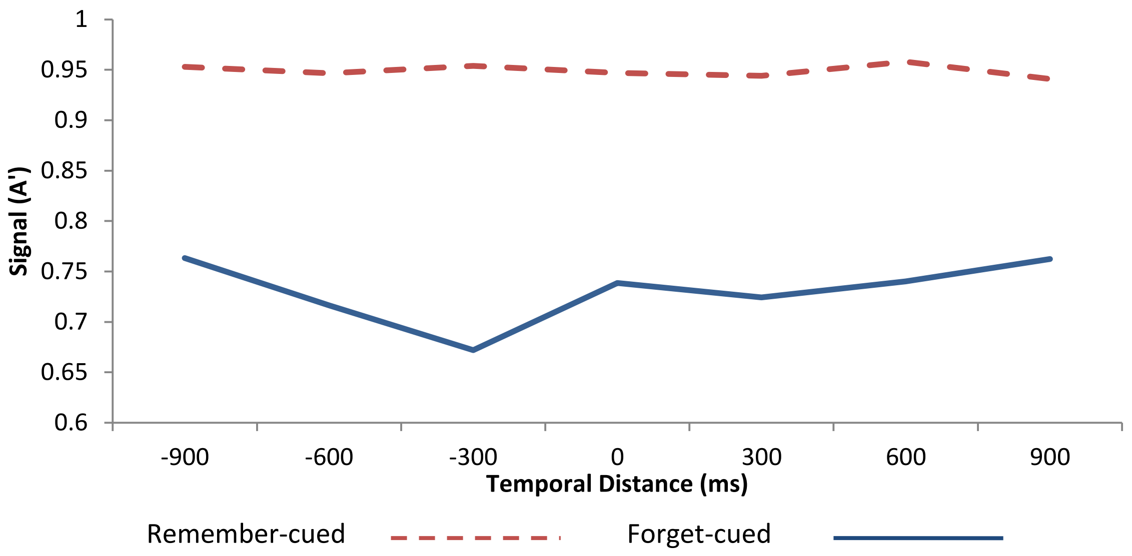

Figure 2.6. Signal strength (A’) data for remember-cued (dashed red line) and forget-cued (solid blue line) words in the flat distribution design. The X axis represents the difference in time between the duration of the fixation cross at study and test.

Discussion. Counter to expectations the results failed to replicate those of experiment 1A. The key significant results found in the ANOVA and contrasts for the balanced design showed a difference between memory conditions. While a planned contrast did show a significant quadratic interaction between memory condition and time condition, no effect of time condition alone was found. Furthermore visual inspection of Figure 2.6 shows that the significant quadratic interaction, if it shows anything other than variability in the data, contradicts the original findings of experiment 1A. Specifically, the current results suggest that the difference between remember-cued and forget-cued word performance is larger when the time at study and test match than when they do not match. This implies a cost for matching temporal locations for forget-cued words rather than a benefit.

General Discussion

While experiment 1A did produce the expected pattern of results experiment 1B did not. As such it seems likely that the initial results were due to the study design rather than the differences in time location. In the first case the signal is seen to increase as the time at test approaches the time at study while in the second case the signal decreases as the time at test approaches the time at study.

The first, and most likely explanation, is that the original curve was the result of the peaked distribution of temporal distances. Key support is provided from the failure to see any significant effect of time in the raw accuracy or reaction time data and the fact that difference between time differences was only apparent when A’ was used. While A’ is supposed to reduce bias it has been critiqued for shortcomings not directly relevant to the current analysis including showing parametric like results despite claiming to be a non-parametric measure of signal (Verde et al., 2006). The existence of these critiques, combined with the fact that calculation of A’ transforms the input data, give cause for concern. In other words it is possible that the data pattern observed in experiment 1A is an artefact produced by the calculation of A’. If this is the case it seems that the severity of the artefact varies with the number of trials in the relevant condition.

Conclusion

Times in the subsecond range have failed to reliably produce an effect that is comparable to that of spatial cues in a directed forgetting task. Initially promising results were shown to be linked to imbalances in the distribution of temporal distances rather than time itself. Notably initial results only proved promising when considering signal measured with A’. These issues should be considered before moving to work using suprasecond intervals.

Chapter 3. Signal Corrections

Experiment 1A and 1B produced contradictory results when analysing the signal measure A'. Critically the pattern of results produced in the forget-cued condition in experiment 1A and 1B matched the pattern of their underlying distributions. Specifically, in the case of experiment 1A lower signal values were seen in conditions with smaller sample sizes (fewer trials) and higher signal values were seen in conditions with larger sample sizes (more trials)[17]. This raised the possibility that the pattern of results observed was due to an unreported sample size bias in the calculation of the signal value A' rather than an underlying change in the signal itself. In addressing this issue two measures of signal will be considered: d' (Tanner, 1954) and A' (Pollack & Norman, 1964).

The Signal Measures

A common measure of signal is d'. By measuring the *z* distance between the distribution of hits and the distribution of false alarms d' produces a standardised measure of signal strength. The hit and false alarm distributions are modelled on a normal distribution. As such the *z* distance between the centres of the distributions can be estimated by converting the hit and false alarm rates to *z* scores and subtracting the *z* score of the false alarms from the *z* score for the hits. This can be expressed in an equation of the form provided below (Verde et al., 2006):

$$d' = z(H) - z(FA) \tag{3.1}$$

However the reliance on *z* scores does impose some limitations. In addition to assuming that the distribution of hits and false alarms are normally distributed and have equal variance, the use of *z* scores means that hit and false alarm rates of 0 or 1 cannot be used as the z score becomes infinite. While this made d' less desirable than A' in experiments 1A and 1B, it is possible to avoid infinities if a correction is used. While a number of corrections have been suggested the log-linear correction (Snodgrass & Corwin, 1988) is simple and has been shown to be preferable to alternatives (Hautus, 1995). This correction works by adding 0.5 to the raw count of trials in each possible condition (hits, misses, false alarms, correct rejections), in turn adding 1 to the overall number of stimulus present and stimulus absent trials for every participant. The correction is applied to all participants, regardless of whether values of 0 or 1 are present in a particular participant. The proportions of hits and false alarms are then recomputed using these corrected values. While this does result in an underestimate of signal, this underestimation becomes diluted as the number of

[17] While it may be considered to be more correct to talk about the number of trials it is easy to become confused between the total number of trials and a specific trial number. As such sample size will be the preferred term here.

samples in a given condition increases (Hautus, 1995). This change in dilution effectively describes a sample size bias.

In contrast A' is not based on *z* scores nor does it have the same issue with hit or false alarm values of 0 or 1. It is for this reason that A' was used in preference to d' in the previous chapter. Instead A' is calculated from a unit square whereby the signal value is calculated as the proportion of the area under fitted lines that approximate an ROC curve. The lines are produced by plotting a line from 0,0 to 1,F and from 1,1 to H,0 where F is the false alarm rate and H is the hit rate. Plotting these lines divides the unit square into the areas shown in Figure 3.1.The A' value can then be calculated by the equation below:

$$A' = I_1 + I_2 + 0.5x(A_1+A_2). \tag{3.2}$$

Alternatively, the calculation of the relevant area under the lines can be summarised by a conditional equation (Grier, 1971). When the hits are greater than the false alarms, the standard equation is used.

$$A'=0.5+[(H-FA)(1+H-FA)]/[4H(1-FA)] \tag{3.3}$$

When the false alarm rate is greater than the hit rate, the equation estimating the area under the curve also changes.

$$A'=0.5-[(FA-H)(1+FA-H)]/[4FA(1-H)] \tag{3.4}$$

When the hit rate and false alarm rate are equal, chance performance occurs and A'=0.5.

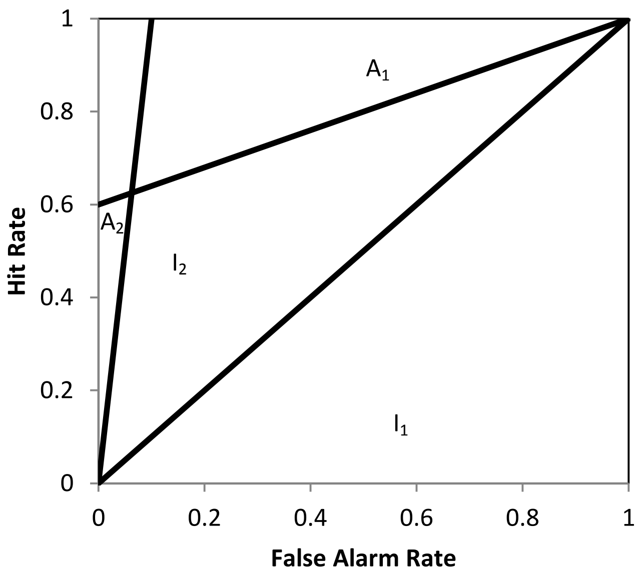

Figure 3.1. A plot of the calculation of A' when the hit rate is 0.6 and the false alarm rate is 0.1. In order to calculate A' the area in region I_1 and I_2 is added to half of that in A_1 and A_2.

Issues With Sample Size

While a number of issues with d' and A' have been identified elsewhere (Rotello et al., 2008; Verde et al., 2006) the issue of varying sample size is not typically considered. When it is considered it is described as beneficial (Hautus, 1995) or comparisons are made between vastly different samples sizes: 16, 64, and 256 in Verde et al. (2006). In each case the issue of sample size is treated as a peripheral issue of limited importance and no corrections are proposed. However, as some study designs consider the role of uncommon distracters or targets and as the peaked distribution design in the previous chapter requires unequal sample sizes between conditions, the sample size issue warrants further consideration. Simulated experiments were conducted for this purpose.

Methods.

Participants. 10,000 simulated participants were generated in each run.

Materials. Simulations were conducted using MatLab R2010a running on a computer operating on Windows 7 64 Bit with a 3.4 GHz Intel i7 processor and 8.00GB of RAM.

Procedure. Simulated d' and A' values were produced by using fixed underlying hit rates varying between 0 and 1 in steps of 0.1. This was done for sample sizes from 1-100 in 1 sample steps.

Generating simulated participant data. The observed number of hits for each simulated participant at each sample size (1-100) was determined by a random draw from a binomial distribution. The mean hit rate of the binomial distribution was set at the underlying hit rate (defined above). The resulting number of hits was then divided by the number of samples to convert the number of hits produced to a subject-specific hit rate. For example, when the underlying hit rate is 0.9 and the sample size is 1, approximately 90% of participants should have a hit rate of 1 and 10% should have a hit rate of 0. When the underlying hit rate is 0.9 and the sample size is 2, approximately 81% of participants should have a hit rate of 1, 18% should have a hit rate of 0.5, and 1% should have a hit rate of 0. For simplicity the false alarm rate of all simulated participants was fixed at 10 of 100 samples. This gives a false alarm rate of 0.1 in all cases except where a log-linear correction is applied. In the case of a log-linear correction the false alarm rate is 10.5 of 101 samples or 0.104.

Corrections. In order to correct for the occurrence of extreme values when calculating d' two methods were used. In the first method simulated participants with observed hit rate values of 0 or 1 were dropped from the simulation without replacement. As such no sample sizes of 1 were included in this version of the correction. In the second method the log-linear correction, as outlined in Hautus (1995), was employed. As no corrections are required for A' the standard A' equation was used in all conditions.

Once all 10,000 simulated participants had completed this process, separate average signal values were calculated at each sample size for each d' correction method and A'.

Results. As shown in Figure 3.2, sample size has an effect on d' and A'. While the ideal scenario would see the calculations produce a perfectly flat line regardless of sample size this has not occurred in any case. When subjects with hit rates of 0 or 1 are excluded, d' scores move outward from 1.28, toward the value associated with 100 trials. When the log-linear correction is employed the pattern is less exaggerated but still present. In the case of A' the issue appears to be more consistent. Scores are underestimated more at lower sample sizes and at lower hit rates[18]. Interestingly, this underestimation is strong enough to push the A' value below 0.5 (chance performance) when the sample size and hit rate are small.

Discussion. As expected the corrected d' and uncorrected A' calculations produce differences between the expected and observed mean signal values. Specifically, as the number of samples decreases the separation between the observed and the expected signal value increases.

While this pattern in the log-linear corrected d' has previously been commented on in terms of a tendency to produce an underestimation of the true signal value and viewed as a strength (Hautus, 1995) it could become problematic in situations where unequal sample sizes are used as an experimental necessity. Furthermore it is worth noting that when the underlying hit rate is below 0.5 the use of either d' correction leads to an overestimation of signal rather than an underestimation. Unlike d' the pattern in the A' data cannot be due to participant exclusion or any additional corrections as neither have occurred. Instead it must be due to the nature of the calculation itself.

Given this sample size related bias in signal strength calculation, when the hit rate is over 0.5, type 1 error rates should increase for d' and A' when the condition predicted to have the smaller signal also has the fewest trials. This is particularly problematic for the results shown in the previous chapter; the matched time condition had the most trials and was expected to produce the highest signal value. If the underlying hit rate is below 0.5 this problem remains for A' but is reversed for d', in these cases the condition with fewer trials is associated with an erroneously high d' value. As such a means of correcting or predicting each of these errors is desirable.

Correcting Log-Linear Corrected d'

In the case of d' a simple modification to the log-linear correction appears tempting. Instead of adding the same correction value to all conditions, a proportional correction might be added

[18] This pattern reverses if the false alarm rate exceeds the hit rate. This is because the equation itself becomes reversed.

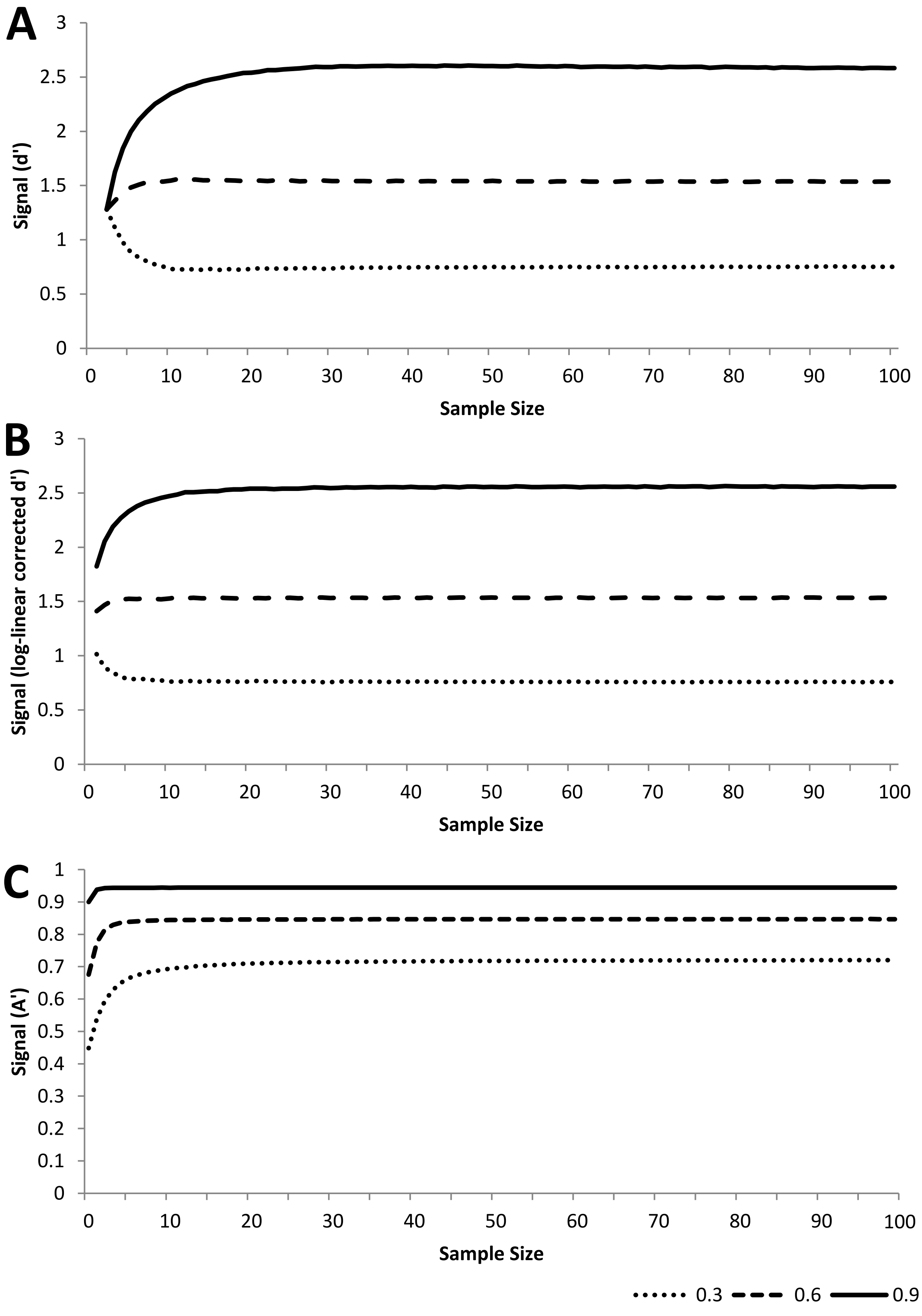

Figure 3.2. Simulated signal values. Panel A shows simulated d' values when hit or false alarm scores of 0 or 1 are excluded. Panel B shows simulated d' values when trials with hit or false alarm rates of 0 or 1 are included and the correction is applied. Panel C shows simulated A' scores. In all cases the false alarm rate is held at 0.1 and hit rate is set at 0.3, 0.6, and 0.9.

instead. As such if a condition with 100 samples was being compared to a condition with 50 samples 0.5 could be added to all cells in the 100 sample condition and 0.25 could be added to all cells in the 50 sample condition.

Unfortunately, as shown in Figure 3.3, when this adjusted correction is used in the simulation in place of the standard log-linear correction this does not return the between trial number comparison to a flat line. In fact such a correction only works when the hits are at 0%, 50%, or 100%[19]. This is probably due to the underlying normal distribution used to produce the *z* scores that are used to calculate d'. As such a correction that incorporates the underlying normal distribution is required. The first step in creating a functional correction for the differences in d' associated with changes in sample size is in identifying the correction value that is required for a given hit rate. Once these values are known a model predicting these values can be created and tested.

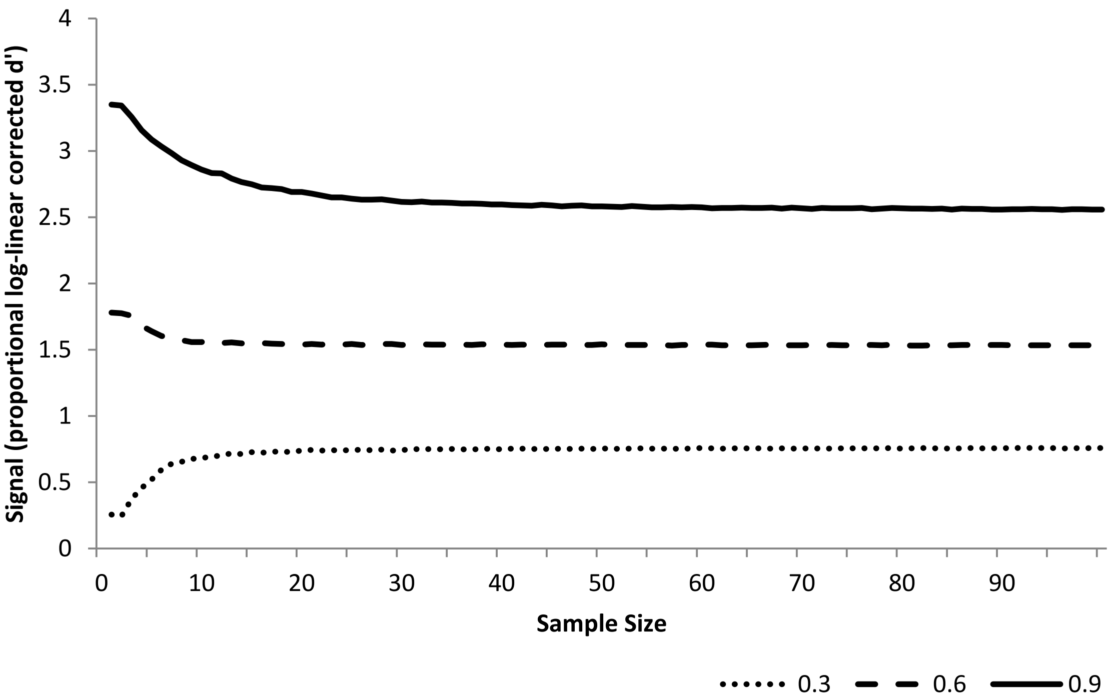

Figure 3.3. The effect of the proportional correction on the calculation of d'. Hit rate is indicated by the legend in the figure. The false alarm rate is 0.1 in all cases.

[19] Recall that the false alarm rate and the false alarm sample size are fixed in the simulations. If the sample size for the false alarms was not fixed the same restriction would apply to them.

Testing will be achieved by running large numbers of simulated experiments across combinations of sample size from 1-100 in 1 sample steps. This testing will produce large amounts of data. In order to simplify its presentation heatmaps will be constructed. These heatmaps will be 100x100 grids and will identify sample size combinations where the number of simulated experiments producing significant results is significantly higher than what would be predicted by chance. In all cases the top left position of the grid will represent 1 sample paired with 1 sample (1,1); the bottom left corner will represent 1,100; the top right corner 100,1; the bottom right corner 100,100.

Methods.

Participants. When finding the correction values 10,000 simulated participants were used for each sample size. When testing the correction values each simulated run included 30 simulated participants[20].

Materials. The same materials were used as when investigating issues with sample size.

Procedure. Simulated data was produced for underlying hit rates ranging from 0 to 1 in steps of .05. For each hit rate the full range of sample sizes from 1 to 100 were produced. Data for simulated participants was calculated in the same way as in the original d' and A' simulations.

Finding the adjusted correction value. The adjusted correction factor required for each sample size was determined by running 10,000 single participant simulations at each sample size. In each simulation the adjustment used was altered such that it created a better approximation of the value required to match a target d' value. The target was set by adding 0.5 to each cell in a 1,000 trial condition where the underlying hit rate was used. Candidate adjustments were defined by an upper value and a lower value. In the first attempt the upper value was 1 and the lower value was 0. In all cases the adjustment was determined by taking the mean of the upper value and lower value. This meant the first adjustment tried was always 0.5. If the d' value produced was smaller than the target value the current adjustment value became the new upper value and the process was repeated. If the d' value produced was larger than the target d' value the current adjustment value became the new lower value and the process was repeated. This continued until an exact match to the target d' value was produced or 200 repetitions had occurred. The resulting adjustment values for a given sample size were averaged across each of the 10,000 simulated participants then plotted as a function of sample size for differing underlying hit rates.

[20] Thirty simulated participants was chosen as it is a reasonable number of participants and 30 or more observations approximate a normal distribution fairly well (Agresti & Franklin, 2007).

An equation modelling the variation in adjustment values was deduced through a combination of visual inspection and brute force techniques. A gamma distribution was used. Alpha, the shaping factor, was set as 1 where beta, the scaling factor, was determined by least squares regression fitting[21] with an exponential function found to provide the best fit.

Testing the adjusted correction values. Signal values calculated using the adjusted corrections were then compared to values calculated using the log-linear correction in terms of their tendency to produce false positives (type 1 error) and their power. In each case 30 new simulated participants were used in each of 100 independent runs. When testing for the tendency to produce type 1 error the underlying hit rate remained the same between sample sizes. When testing for changes in power the underlying hit rate differed by 0.05 between sample sizes.

Tests for type 1 error and power were achieved by running 100 simulations for each sample size pairing. Underlying hit rates varied from 0 to 1 in steps of .05. Separate hit rates for each sample size were generated for each participant as outlined earlier. To mimic real world application the hit rate used to calculate the appropriate adjustment for each participant was calculated by dividing their total number of hits by their total number of samples. The false alarm rate of each participant was also set by a binomial distribution. When setting the false alarm rate the probability of a hit in the binomial distribution was always 0.1 and the number of Bernoulli trials was always 100.

In order to summarise the data only the number of simulated runs that produced a significant difference in a paired sample t-test was recorded in each sample size combination. These counts were then converted to 100x100 pixel heatmaps; each pixel showed a different combination of sample sizes.

When testing for type 1 error sample size combinations producing a type 1 error rate significantly higher than 5% (as determined by chi-square) were deemed to be problematic. These combinations are shown as grey pixels (p<.05) or white pixels (p<.01). Combinations that do not reach this threshold are shown as black.

When testing for changes in power the number of simulations returning a significant distribution is represented on a linear grey-scale. Pure black shows that less than 25% of the simulations returned a significant difference for that sample size combination. Dark grey shows 25-50%, light grey 50-75%, and pure white 75-100%. The symmetry of each plot was quantified by checking whether sample size combination x,y fell in the same colour band as sample size

[21] Initially the alpha and beta values were both free to vary during least squares fitting. While allowing both to vary did improve the model fit slightly the relationship between trial number and the beta and alpha values became less stable. As subsequent testing showed no practical benefit associated with the additional complication a more user friendly solution has been adopted.

combination y,x. The total number of same colour band pairings was then divided by the total number of pairings to give a proportion. Uncorrected symmetry scores were compared with corrected symmetry scores using chi-square analysis ($\chi^2_{(1)\ \mathrm{Critical}} = 2.71$). The analysis compared the total number of matching pairs vs. non-matching pairs between correction conditions.

Results.

Finding the adjusted correction values. Examples of the output of the simulation procedure used to identify the correction values are shown in Figure 3.4.

The correction value required was found to increase with sample size and proximity to a hit rate of 0.5. The correction values were able to be predicted by a cumulative gamma function.

$$\text{Adjustment}=F(\text{samples};\ \alpha,\ \beta)\times w \tag{3.5}$$

In all cases F represents the cumulative density function of the Gamma distribution, samples is the number of trials in the sample size of interest, α has a value of 1, β is a value specific to the observed hit rate, and w is a weighting factor. As the log-linear correction adds 0.5 to all cells the weighting factor used here is 0.5. If a different value were added to the cells that value would be used as the weighting value. The β is given by

$$\beta = 1.6411\times(0.5-|h-0.5|)^{-0.684} \tag{3.6}$$

Where h is the total hit-rate across conditions. In total then the equation for the adjustment is:

$$\text{Adjustment}=F(\text{samples};\ 1,\ 1.6411\times(0.5-|h-0.5|)^{-0.684})\times 0.5$$

Or using Excel functions: Adjustment = GammaDist(samples,1,1.641*(0.5-abs(h-0.5))^-0.684)*0.5

(3.7)

This makes the adjustment a function of the hit rate and the number of trials.

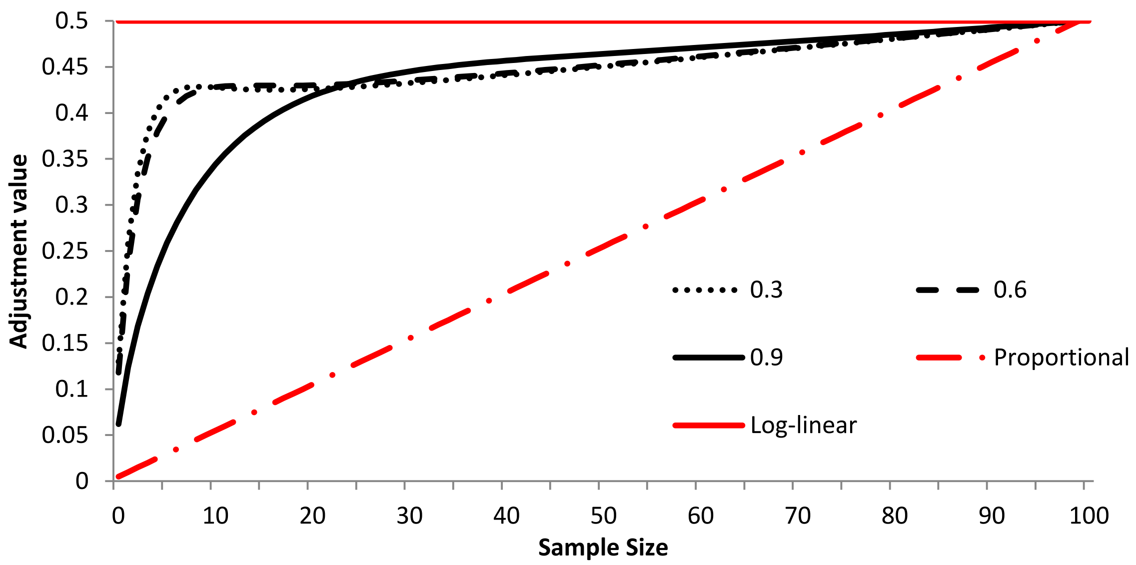

Figure 3.4. Correction values for hit rates of 0.3, 0.6, and 0.9. The proportional correction suggested earlier and the log-linear correction suggested in Hautus (1995) are supplied in red for reference.

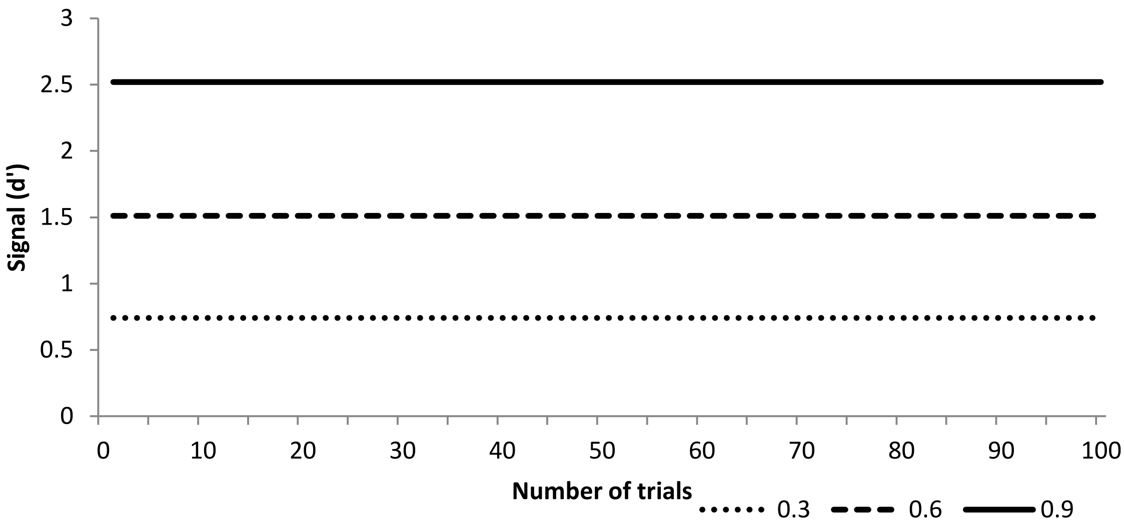

Figure 3.5. Simulated d' values when trials with hit or false alarm rates of 0 or 1 are included and the proposed adjusted correction is applied. The false alarm rate was held at 0.1 and hit rate is set at 0.3, 0.6, and 0.9 as indicated by the legend.

Testing the adjusted correction values. As shown in Figure 3.5, simulations demonstrate that the correction returns the mean d’ values to a straight line, indicating a constant estimate of signal regardless of sample size. Further simulations, shown in Figure 3.6, comparing the corrected values to the uncorrected values show that the uncorrected values inflate type 1 error at both high and low hit rates. Figure 3.6 also shows that the problem gets worse as the difference in trial number between conditions increases. However, unadjusted correction values can be used with hit and false alarm rates between .35 and .65 as long as there is more than 1 sample in one of the conditions. Similarly when both conditions have a sample size over 30 there is little benefit to using the correction even in the worst conditions.

Figure 3.7 shows how the correction alters power. The general pattern can be seen to move from an asymmetrical distribution to a symmetrical distribution. The asymmetrical power distribution reflects the tendency for conditions with fewer trials to produce smaller signal values when the correct adjustment is not used. This leads to asymmetry in two key ways. The first is the production of type 2 error. This occurs where the condition with fewer trials should have a higher signal but this difference is masked due to underestimation. The second is the production of type 1 errors. There are two additional subtypes. The first occurs when the condition with fewer trials should also produce a lower signal score. In this case the bias in the measure increases the probability of finding a significant difference. The second occurs when the condition that should have a higher signal value has very few trials. In this case the difference in signal produced by artefact in the log-linear correction reliably exceeds that produced by the actual difference in signal. In this case a significant difference is detected but the difference is in the wrong direction.

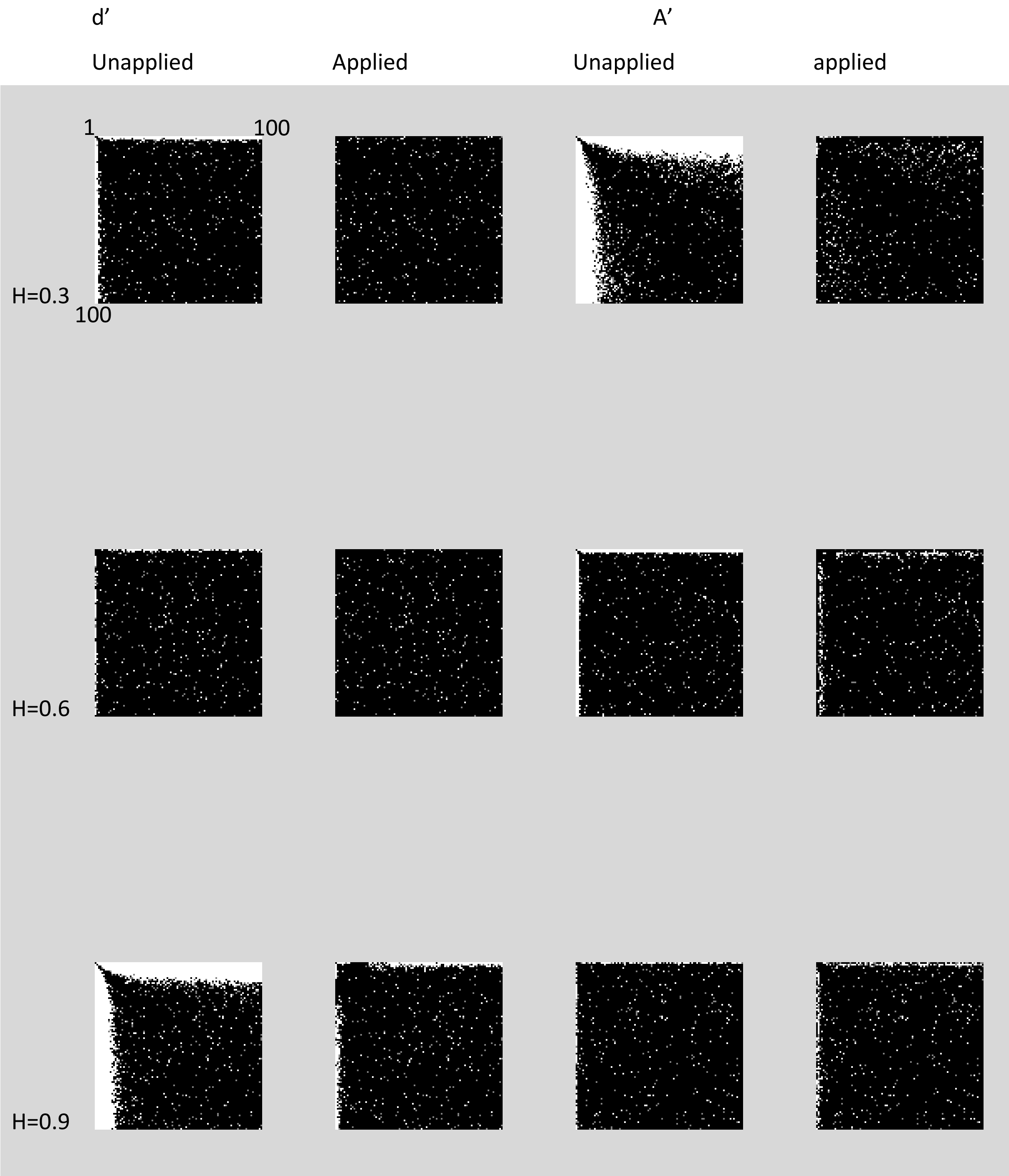

Figure 3.6. Heatmaps for type 1 errors in situations where no difference should exist between conditions. The underlying hit rate of each row (H) is shown on the left. Black areas represent areas where error rates are not significantly higher than 5%. Grey areas represent combinations of trials where the error rate is significantly higher than 5%. White areas represent combinations of trials where the error rate is significantly higher than 1%.

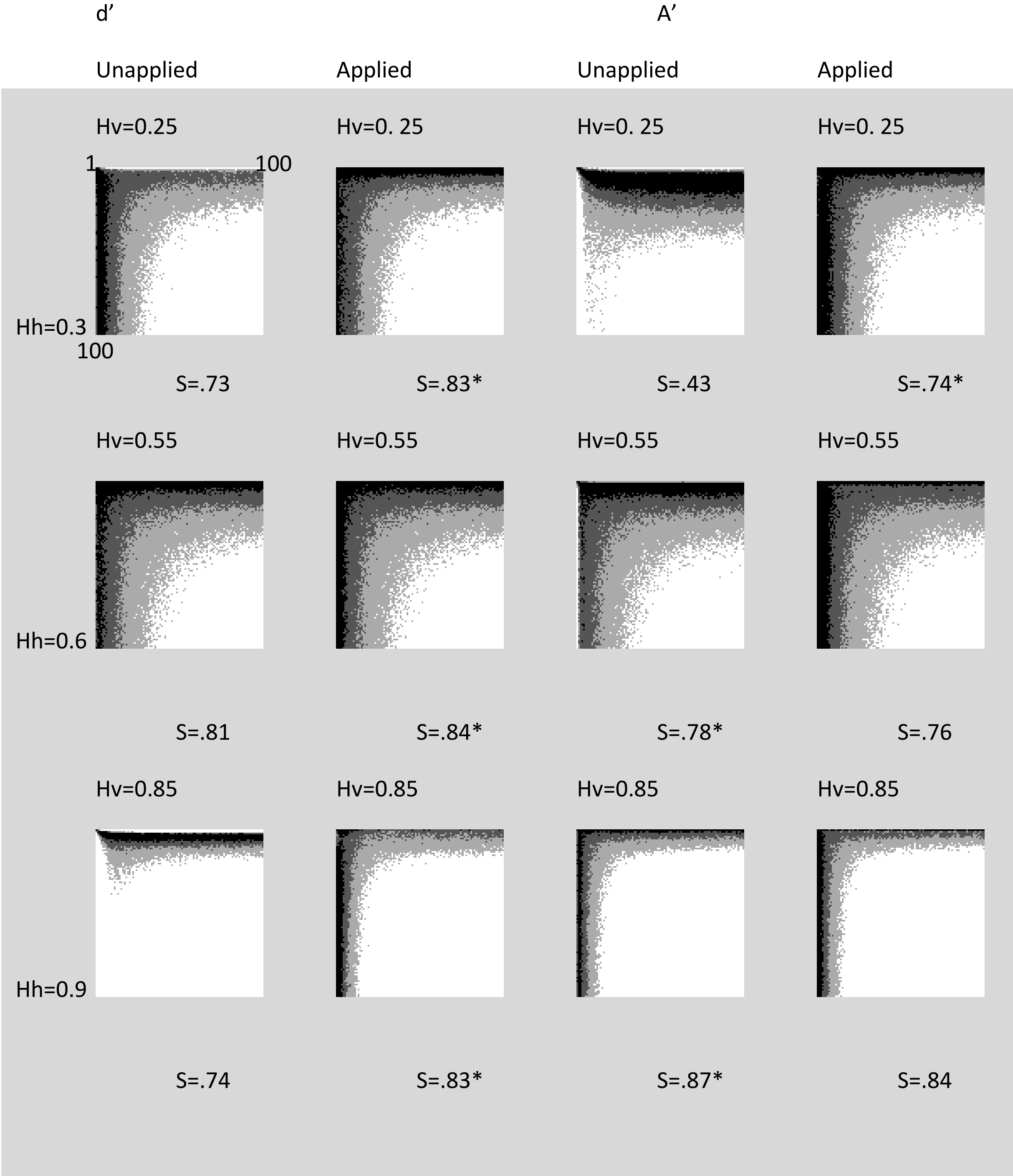

Figure 3.7. Heatmaps where an underlying difference between conditions exists. Sample sizes associated with horizontal rows have the hit rate (Hh) shown on the left of the corresponding heatmap. Sample sizes associated with vertical columns have the hit rate (Hv) shown above the corresponding heatmap. Pure black shows that less than 25% of the simulations returned a significant difference for that sample size combination. Dark grey shows 25-50%, light grey 50-75%, and pure white 75-100%. Symmetry scores (S) are shown beneath each heatmap. An asterisk indicates which condition has the significantly higher symmetry score as calculated by chi-square. In all cases *p*<.001.

Discussion. The simulations confirmed that the size of the correction value required to maintain a constant signal value between conditions with different sample sizes depended on the underlying hit rate and the number of trials in each condition. An equation modelling this relationship was found and the corrections it produced were found to decrease type 1 error rates and reduce asymmetries in power. These benefits are particularly apparent when there is a large discrepancy between the number of trials in each condition and/or as the underlying hit rate moves further from 0.5, particularly at sample sizes under 30.

As the correction and the d’ calculation are both strongly related to the assumed underlying normal distribution the model produced can be used independently on hits and false alarms. This suggests that the initial technique of adding the same value to the hits and false alarms may have led to the production of erroneous results in some cases.

Above 30 trials there does not appear to be any real practical significance to adjusting the value used. This is likely to be due to samples with 30 or more observations approximating a normal distribution fairly well (Agresti & Franklin, 2007).

While d’ was not used in the previous chapter, as hit rates of 100% and 0% were likely to occur reasonably frequently, the analysis undertaken here suggests that d’ could now be used on those values without introducing unnecessary error. However issues remain. As shown in Figure 3.6 type 1 errors remain significantly higher than expected when the underlying hit rate is particularly high and at least one small sample size is involved. While this is not ideal the corrected adjustment still produces a marked reduction in type 1 error with a typical drop being along the lines of 1000 false positives for the uncorrected adjustment vs. 100 for the corrected adjustment. In these scenarios significant results should be interpreted with caution. Ideally these situations should be avoided through experimental design. Where this is not possible a very conservative critical value should be selected and the change in the meaning of the value should be noted during discussion of the result.

Predicting A’

While the correction factor employed in d’ provides a clear culprit for the trial related drift no such correction is used in calculating A’. The lack of alternative sources suggests that the error must be produced by the A’ equation itself. In considering how the A’ calculation could produce the bias observed in Figure 3.2 it is most useful to start with a simple example before moving to more complicated situations. In doing so a means of predicting the A’ values associated with a given hit rate and sample size becomes apparent. The use of this prediction in hypothesis testing will be explored.

The simplest place to start is with the example of an experiment with 1 sample where the stimulus is present and many samples where the stimulus is absent. Consider a hypothetical example where there is a true hit rate of 0.6 and a false alarm rate of 0.1[22]. In this case the expected true A' value is 0.847. However, in the case of any single participant the observed hit rate can only be 1 or 0 while the false alarm rate is assumed to be well estimated at 0.1. As a result the observed A' value for any single participant can only be 0.975 or 0.225. As the odds of a hit are 0.6 this means the observed mean A' value would be 0.6 x 0.975 + (1-0.6) x 0.225 = 0.675 instead of 0.847. This imposed imbalance is what leads to the underestimation of the true A' value seen in the earlier simulations and is the product of the conditional form of the A' equation.

Expanding this to two samples the range of possible observed hit rates would be 1, 0.5, and 0 with the ratio of contribution being determined by a probability tree. The number of branches associated with each probability value in the tree can be modelled using Pascal's triangle. However while manual calculation is simple at small sample sizes the calculation of a whole triangle quickly becomes unwieldy. As such an alternative means of calculating the observed number of branches for each probability value is given below. Here N is the sample size and H is the number of hits in that sample.

$$\text{Branches} = N!/(N-H)!/H! \tag{3.8}$$

In addition to this weighting factor the actual probability of observing these values must also be calculated. Here h is the true hit rate, N is sample size, and H is the number of hits in that sample.

$$\text{Weighting} = h^{H} \text{x} (1-h)^{(N-H)} \tag{3.9}$$

The A' score for the branch must be calculated using the observed hit rate.

$$\begin{gathered}\text{If } H/N=F \text{ then } A'=0.5\\ \text{If } H/N>F \text{ then } A'=0.5+[(H/N-FA)(1+H/N-FA)]/[4H/N(1-FA)]\\ \text{If } H/N<F \text{ then } A'=0.5-[(FA-H/N)(1+FA-H/N)]/[4FA(1-H/N)]\end{gathered} \tag{3.10}$$

Then these scores must be summed for all of the related branches.

---

[22] For the sake of simplicity, and in keeping with the findings of the experiments in the previous chapter, the false alarm rate has been assumed to remain constant across conditions and the product of multiple samples. As such it is treated as 0.1 in all calculations in this paragraph.

Simplified

$$\sum_{H=0-N}^{N} \hat{A}' = \text{Branches x Weighting x A'}$$

Expanded

$$\sum_{H=0-N}^{N} \begin{array}{l} \text{If } H/N=F \text{ then } \hat{A}' = N!/(N-H)!/H! \times h^{H} \times (1-h)^{(N-H)} \times 0.5 \\ \text{If } H/N>F \text{ then } \hat{A}' = N!/(N-H)!/H! \times h^{H} \times (1-h)^{(N-H)} \times (0.5+[(H/N-FA)(1+H/N-FA)]/[4H/N(1-FA)]) \\ \text{If } H/N<F \text{ then } \hat{A}' = N!/(N-H)!/H! \times h^{H} \times (1-h)^{(N-H)} \times (0.5-[(FA-H/N)(1+FA-H/N)]/[4FA(1-H/N)]) \end{array} \quad (3.11)$$

As shown in Figure 3.8 this prediction lines up well with the simulated values produced for Figure 3.2.

By creating a predicted value for each participant in an experiment a curve that assumes that there is no difference between the means of the conditions can be created. While the true hit and false alarm rates are unknown the total hit and false alarm rate can be used in their place. By testing against this curve rather than testing against a global mean a more accurate null hypothesis can be created. While this approach does not correct for discrepancies between sample sizes in the same way as the correction used for d’ this prediction approach is preferable here. This is because, unlike d’, the drift in A’ is not attributable to a pre-existing correction factor.

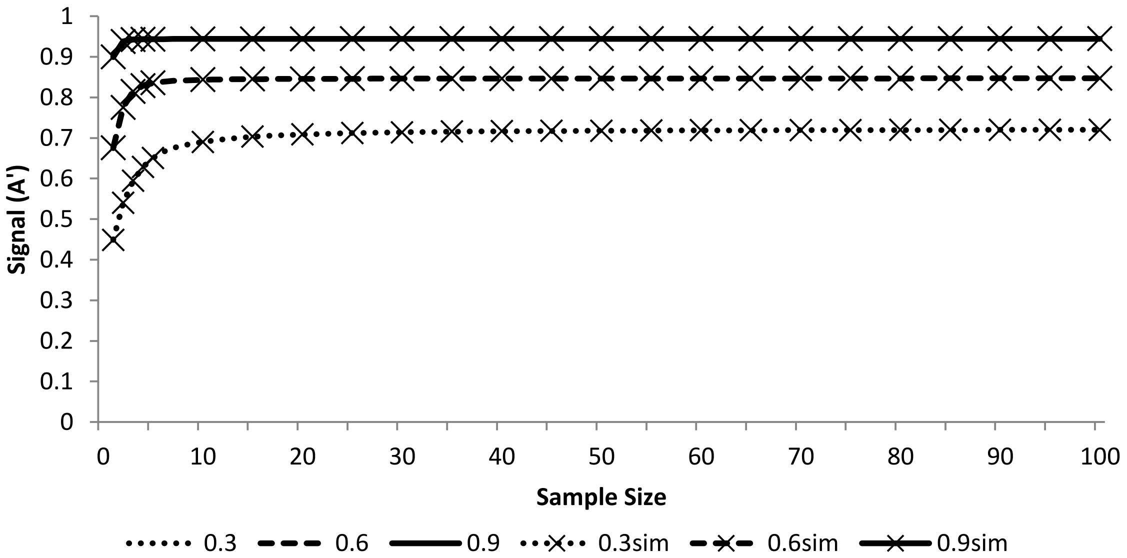

Figure 3.8. Predicted A’ values overlaid on the simulated A’ values.

Using the A' Prediction Method

The first point to consider when using the predicted A' method is that the values will form a new null hypothesis. The predicted values, just like the simulations, assume that there are no real differences between conditions in the observed data. As such the hit and false alarm rates used in the predictions are obtained by collapsing across the conditions in the experiment. Such collapsing formalises the assumption that there are no differences between the conditions.

When using a parametric test the null hypothesis is determined by the predicted mean A' values[23]. The predicted A' values produced can then be used to model the mean signal values across participants observed in a sample. Importantly these correction procedures cannot give an estimate of the actual observed signal if differences between the observed and predicted value are seen to exist. This is because the degree of difference between observed and predicted values changes as the observed hit rate in a condition changes.

Methods.

Participants. For each test 30 simulated participants were used as this was taken to be a reasonable sample size arguably allowing for normal approximation (Agresti & Franklin, 2007).

Materials. The same materials were used as when investigating issues with sample size.

Procedure. The process associated with testing the adjusted d' correction was repeated. The only differences were that a new set of simulated participants were used and A' was calculated in place of d'.

When calculating uncorrected A' values equation 3.3 or 3.4 was used as appropriate. When using the prediction method the predicted A' value was generated by using the observed overall hit rate (total hits/total samples) for each participant as the true hit rate in equation 3.11. This mirrors the approach that would be taken in real world application. The predicted value was then subtracted from the uncorrected A' value.

Results.

Type 1 errors. As shown in Figure 3.6, in most cases the correction reduced the number of significant differences in the pairings. In most cases this lead to a reduction in the total number of

---

[23] Non-parametric tests on the other hand require the creation of new probabilities to test against. For example if no correction were made and no difference existed between the conditions the number of participants with condition X was larger than condition Y would be roughly equal to the number of participants where condition X was smaller than condition Y. This does not hold true when the sample size in each condition is not equal. As an extreme example if the true underlying hit rate is 0.9 and the true underlying false alarm rate is 0.1 the odds of a participant having a higher score in a condition with 1 sample than a condition with 100 samples moves from 0.5 to 0.9. Such predictions can be created by considering the end of the probability trees rather than the mean score.

pairings showing a total that was significantly different to chance. In some specific cases there is a slight increase in type 1 error rate after correction.

Type 2 errors. As shown in Figure 3.7, the uncorrected data showed a power distribution that tended toward the lower left corner of the heatmap and became more asymmetrical at hit rate values close to the false alarm rate. The power distribution of the corrected data was generally symmetrical. In the case of hit rates of 0.6 and 0.9 the corrected data was less symmetrical than the uncorrected data. Visual inspection shows that this is due to a cost in power when the condition with fewer trials has the lower signal value.

Discussion. As expected the corrections reduced the number of false positive results in most cases. While this suggests that the corrections are generally beneficial it is important to note that they are not perfect. For example, the heatmap for an underlying hit rate of 0.9 still contains inflated type 1 error when one of the sample sizes is low. Symmetry was largely restored to the power heatmaps. This suggests that the prediction method can help reduce the bias in the A' measure when a difference actually exists.

The cost to power in the case of 0.6 and 0.9 is quite small and is actually beneficial. In these cases the uncorrected data overemphasises the size of any effect that is present. The correction prevents this from happening.

Notably, for the sake of simplicity, the current example has only considered the effect of different sample sizes in the stimulus present (hit) condition. If sample size also varied in the stimulus absent (false alarm) condition a similar prediction technique could be used by including the false alarm rate and its associated sample size in the calculation of the probability trees.

General Discussion

While d' and A' both produced troublesome artefacts as a result of differences in sample size the reasons underlying these artefacts differed. In the case of d' artefacts were produced through the use of a common correction for hit rates of 1 and 0 regardless of trial number. For A' artefacts were produced as a result of the core equation itself. As the issues with d' were the result of a pre-existing correction the issues were able to be removed, or at least greatly reduced, by factoring sample size into the calculation of the value used for the correction. A' required a change in the statistical tests used when looking for a significant difference.

In both cases consideration of the differences in sample size between conditions reduced the number of type 1 errors and distortions in power. As there are no real costs involved in implementing the d' correction or the A' prediction method implementation of these adjustments is advisable.

Common limitations. While the d’ correction and A’ prediction methods do provide a means of comparing signal values between conditions they do not give true estimates of the signal value present in the conditions themselves. They can only give an indication that there is a difference, not that the difference is of a particular size. This is apparent through simple inspection of the equations for each model. In the case of d’, higher signal values are associated with lower correction values. In the case of A’, higher signal values are associated with higher predicted values. Essentially, if the signal strength in one sample size condition actually does differ from the other, the assumption of equal signal used in the d’ correction and the A’ prediction method is false. While this will not prevent a difference being found it does mean that the estimated value signal produced in each case is not strictly accurate.

Conclusion

Ironically, when the sample size differs between conditions log-linear corrected d’ and standard A’ calculations create bias: increased type 1 error (false alarms) and type 2 error (misses). Proposed corrections to these signal estimates were shown to greatly reduce both issues.

Chapter 4. Re-analysing the Data from Experiment 1

In light of the finding that differences in sample size lead to erroneous statistically significant differences in signal strength, the data from experiment 1A need to be reanalysed. As experiment 1B used a fixed sample size across conditions there should be no need to apply the correction from the signal strength chapter to the data as it was collected. However, data from experiment 1B can be used to make a point about the importance of using equal numbers of trials if the data is re-categorised.

The first investigation will consider the data presented in experiment 1A with the original divisions between temporal distances being maintained. Then, in order to demonstrate the importance of the effect of trial number on the results and the effectiveness of the correction from the previous chapter, data from experiments 1A and 1B will be shown to be able to produce a finding that opposes the original finding of experiment 1A. This will be achieved by dividing trials into same temporal location and different temporal location conditions rather than the specific temporal distance categories originally used. This should lead to the same temporal location condition always having fewer trials than the different temporal location condition. As a result uncorrected signal strength measures should suggest that the signal is stronger in the different temporal location condition than the same temporal location condition. Importantly this would reverse the result for experiment 1A. Such a reversal demonstrates the problems that arise due to the artefact in the signal calculation. If the proposed corrections work, the same conclusion should be drawn from experiment 1A and experiment 1B regardless of how the data is categorised.

Methods

Participants. Data from experiment 1A and 1B were included. Only data from participants who met the selection criteria employed in experiments 1A and 1B were included in the analysis to ensure a fair comparison with the original results.

Procedure.

Re-analysis of 1A. Data from experiment 1A was considered in isolation to determine whether the artefacts described in the previous chapter contributed to the pattern of results originally obtained. Signal measures for the data from 1A were calculated with and without the corrections outlined in the previous chapter.

Re-categorisation. Data from experiment 1A and 1B were combined for the re-categorisation condition. In this condition data from trials with a temporal distance of anything other than 0 were pooled to become the different temporal location condition for each participant individually. Data from the 0 temporal distance condition became the same temporal location condition. This meant that participants from experiment 1A had 8 trials in the same temporal location condition and 24 in the different temporal location condition while those from experiment 1B had 5 trials in the same temporal location condition and 30 in the different temporal location condition.

Plotting corrected A’. When using the A’ prediction method outlined in the previous chapter the null hypothesis is that none of the mean A’ values differ from zero. If these corrected values were plotted alongside the uncorrected values, the difference between corrected and uncorrected A’ would be over-emphasised as the uncorrected scores would be above 0.6 while the corrected scores would be close to zero. In addition to this, the corrected A’ values for the forget-cued and remember-cued words would both be close to zero and would erroneously show no difference between the forget-cued and remember-cued conditions. In order to meaningfully plot corrected A’ along-side uncorrected A’, and in order to show a meaningful difference between the forget-cued and remember-cued conditions, A’ values for the forget-cued and remember-cued conditions will be calculated for each participant. These memory cue condition specific A’ values will be calculated using the hit rate obtained when temporal location is ignored in each memory cue condition. These A’ values will then be added to each forget-cued and remember-cued temporal location condition as appropriate.

Results

Re-analysis of 1A. Corrected A’ data from experiment 1A is presented in Figure 4.1. Unlike the uncorrected analysis presented on page 42 there is no apparent trend in the data for the forget-cued words to show decreased signal strength as a function of increasing temporal distance. This is supported by a repeated measures ANOVA that found a significant main effect of memory condition ($F_{(1,31)}$=82.29, p<.001) but no significant main effect of or interaction with temporal distance. Planned contrasts equivalent to those used in experiment 1 also failed to produce any significant results.

Considering d’ without correction to the log linear transform there is a significant effect of memory condition ($F_{(1,31)}$=145.15, p<.001) time condition ($F_{(1,31)}$=2.58, p=.020) and the interaction ($F_{(6,192)}$=3.40, p=.003). Contrasts also reveal significant linear trends for time distance ($F_{(1,31)}$=5.25, p=.029) and the interaction ($F_{(1,31)}$=4.95, p=.033) as well as significant quadratic trends for time distance ($F_{(1,31)}$=6.60, p=.015) and the interaction ($F_{(1,31)}$=10.00, p=.003). Visual inspection of Figure 4.1 suggests this is due to the presence of a quadratic pattern in the remember data.

Corrected transformation of the d’ data only shows a significant effect of memory instruction ($F_{(1,31)}$=136.44, p<.001). No other significant differences are found.

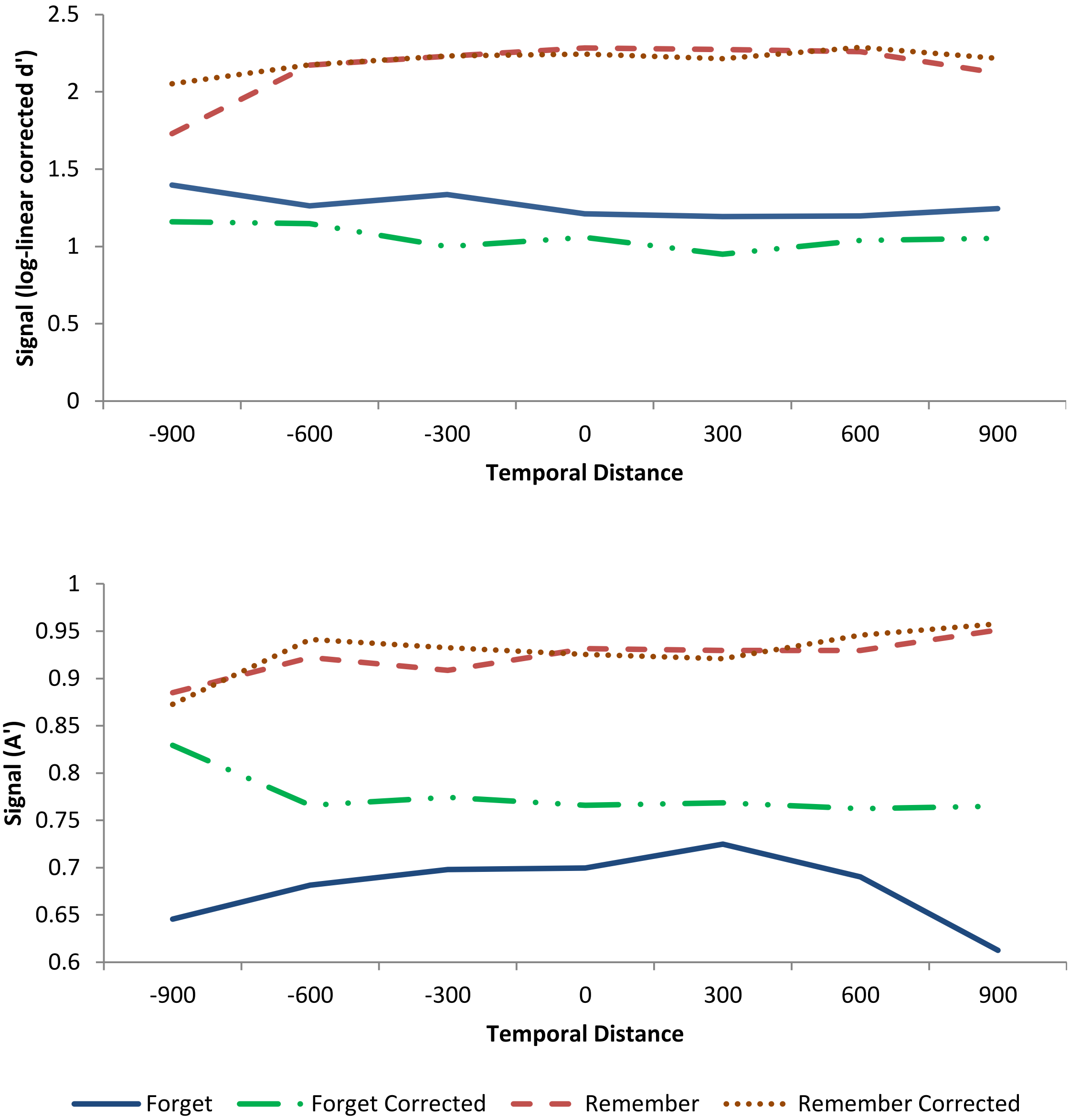

Figure 4.1. Corrected and uncorrected signal strength (d' and A') data for remember-cued and forget-cued words from experiment 1A. The X axis represents the difference in time between the duration of the fixation cross at study and test.

Re-categorisation. Re-categorised data is presented in Figure 4.2. For A', the difference between the uncorrected and corrected signal values is clear and supported by the patterns of significance seen in the repeated measures ANOVAs. In particular, while a significant main effect of memory condition is present for the corrected ($F_{(1,63)}$=70.98, p<.001) and uncorrected data ($F_{(1,63)}$=120.28, p<.001) the main effect of time location is not significant with standard criteria in either case although it is significant if a one-tailed test is used for the uncorrected data ($F_{(1,63)}$=2.82, p=.098 or p=.049$_{\text{(one tailed)}}$). Planned contrasts within memory conditions were not significant. The interaction was not significant in any case.

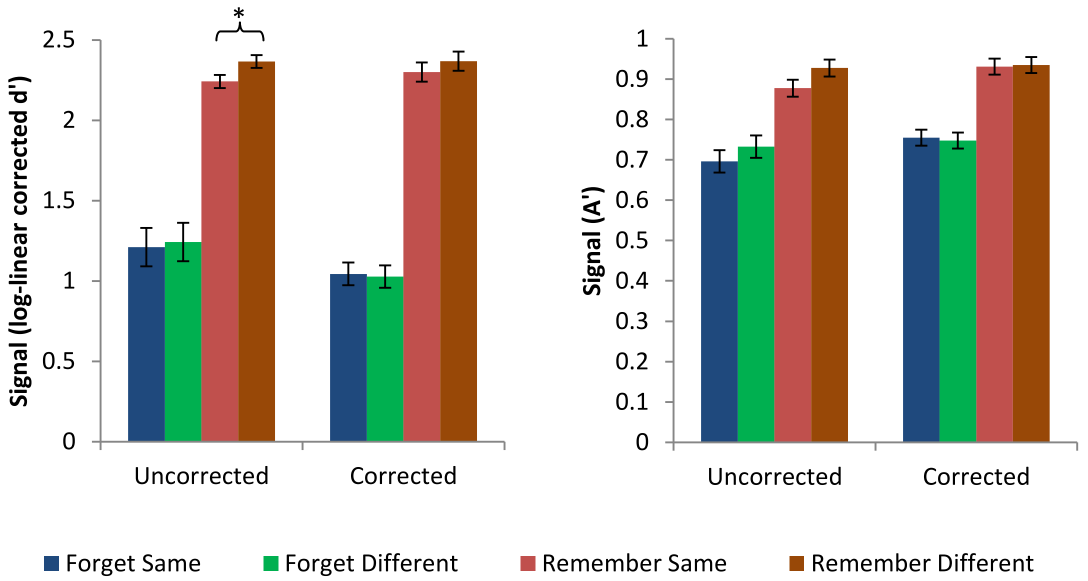

Figure 4.2. Corrected and uncorrected signal scores (d' and A') for remember-cued and forget-cued words when categorised as same and different time locations at study and test. The presented data contain values from the original and balanced versions of the designs. In this case the corrected and uncorrected values map on top of one another despite producing a different outcome in the ANOVA. An * is used to denote where planned contrasts indicate that there is a significant difference (p<.05) between locations within a memory condition. Within subjects error bars (Masson & Loftus, 2003) represent the planned contrasts between temporal locations within memory cue conditions e.g. same vs. different temporal locations for remember cued words.

Log-linear corrected re-categorised d' scores show significant main effects for memory ($F_{(1,63)}$=108.44, p<.001) and time location ($F_{(1,63)}$=5.11, p=.027) as well as a significant interaction ($F_{(1,63)}$=9.36, p=.003). Contrasts reveal a significant difference between same and different temporal location words in the remember-cued words ($F_{(1,63)}$=8.83 p=.004) but not the forget-cued words. Corrected d' scores only show a significant main effect of memory ($F_{(1,63)}$=200.31 p<.001). No other significant differences were found in the corrected data.

For completeness, analyses of re-categorised accuracy and reaction times for correct responses were also considered. The data is presented in Figure 4.3. In the case of reaction time the data of 11 participants was excluded due to accuracy values of 0 in one or more conditions. Main effects of memory were found for accuracy ($F_{(1,63)}$=254.7, p<.001) and reaction time ($F_{(1,52)}$=28.94, p<.001). No other significant differences were found.

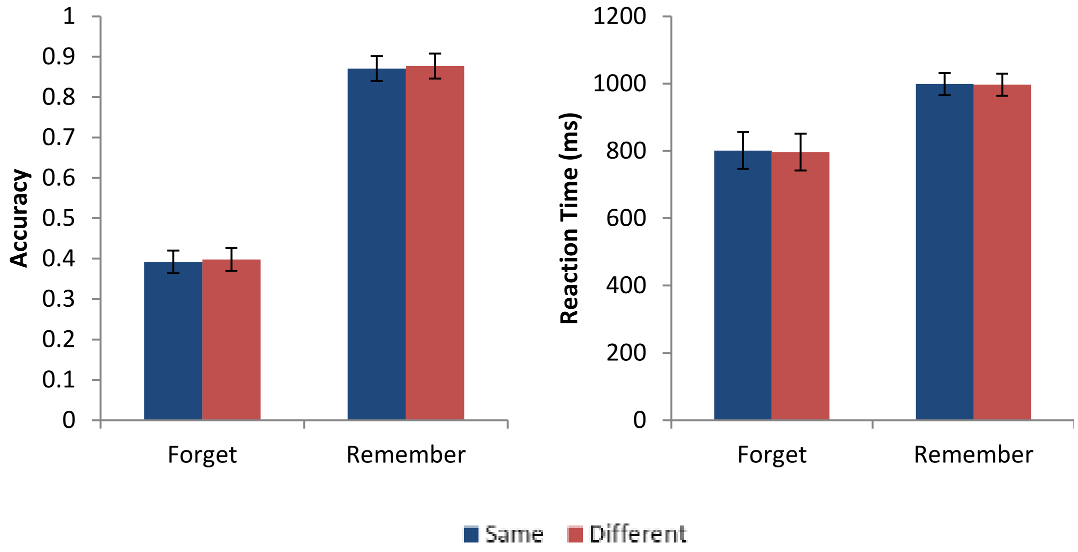

Figure 4.3. Accuracy and reaction time data when combined across experiments 1A and 1B and re-categorised as same and different time location. In the case of Accuracy the false alarm rate in the foil condition is 13%. For reaction time the average correct reaction time to foils is 935ms. Temporal locations are indicated by the legend at the bottom of the figure. Within subjects error bars (Masson & Loftus, 2003) represent the planned contrasts between temporal locations within memory cue conditions e.g. same vs. different temporal locations for remember cued words.

Discussion

The results observed here support the importance and effectiveness of the corrections outlined in the previous chapter. Specifically, when there is an imbalance in the number of trials between conditions this will lead to an erroneously low estimate of signal in the condition with fewer trials unless an appropriate correction is used. While the results of experiment 1A do show this the clearest demonstration comes from the re-categorisation of trials across experiment 1A and 1B. In the re-categorised case failure to use a correction produces a pattern that runs in the opposite direction to that seen in the initial analysis of experiment 1A. In the case of A' a main effect would be significant if a one way test were used, in the case of d' a comparison of remember-cued words in the same location with remember-cued words in different locations reaches significance. Most importantly, in both cases, the corrections remove these erroneous differences leading to the same conclusion being drawn regardless of how the data is categorised.

Interestingly the impact of the different sample sizes appears in the remember-cued words for d' and the forget-cued words for A'. This is likely to be a reflection of the differences between the measures. In the case of d' errors are the result of trials with accuracy of the hits being close to 0 or close to 1. In the case of A' errors are the result of changing between the two versions of the formula. Where the hit rate and false alarm rate are treated independently in the calculation of d' they are interdependent in the case of A'. Re-applying this to the actual data d' errors are seen in the remember-cued condition as this condition contains values close to 1. In the case of A’ errors are seen in the forget-cued condition because the hit rate in the forget condition is closer to the false alarm rate than the hit rate in the remember condition.

In the next set of experiments, examining suprasecond timing versions of the directed forgetting task, the use of such corrections will be important if signal detection measures continue to be employed. However, given that the signal strength measures failed to provide a different message to the accuracy data once corrections were employed, it is worth considering whether or not to proceed with their use.

While signal detection measures should reduce the influence of response bias they can be problematic. Issues beyond the sample size used have been outlined elsewhere (Rotello et al., 2008; Verde et al., 2006). Furthermore as the sample sizes used in the current experiment design tend to be quite low it is worth noting that they tend to fall in the danger zone identified in the previous chapter, particularly in the peaked temporal distribution design. As such, in order to maximise the sample size per condition and remain in line with the design and analysis style used in Hourihan et al. (2007), future use of signal detection measures will be restricted to same time vs. different time comparisons. While this still leads to an imbalance in the distribution of temporal distances the proposed corrections have shown an ability to cope well with imbalances of this magnitude. This approach is also preferable to proceeding with the flat temporal distribution design alone as it increases the power in the different time condition while maintaining a division comparable to Hourihan et al. (2007). As an additional protection no difference in signal level will be considered to be significant unless the same pattern is found using A’ and d’. Given the apparent tendency for d’ to show erroneous differences in the remember data and A’ to show erroneous differences in forget data this criteria seems a good fit for the current line of experiments.

Chapter 5. Directed Forgetting and Suprasecond Times

Experiment 2

While experiments 1A and 1B failed to produce a benefit for the inclusion of when information, the implications of this failure are unclear. While this may suggest that directed forgetting tasks do not use episodic memory or that the when component is not useful in remembering there is another possibility: the when component was not clear enough or occurred in the wrong time class. These possibilities will be addressed in the current experiment.

The issue of clarity or salience of changes in durations has been noted before (Bottini & Casasanto, 2010; James, 1890; Riemer, Trojan, Kleinbohl, & Holz, 2012). In fact, time and space have been argued to be much different to each other in this regard for some time as changes in visual space are readily perceived whereas changes in times are less noticeable. More specifically, perception of the location or size of something only takes a moment whereas perception of the duration of something takes many moments observed sequentially (James, 1890). This idea is illustrated in theoretical models of timing that imply that time is naturally less salient as accurate recording has been argued to require constant attention (Zakay & Block, 1997).

In experiments 1A and 1B avoidance of the apparent requirement for ongoing attention to time was attempted through the use of subsecond times. Subsecond timing is argued to be automatic and not truly dependent on attention (Lewis & Miall, 2003a). The downside being that the durations themselves are very brief meaning the difference between durations must also be small, perhaps too small to notice a difference in the durations being shown.

Perhaps suprasecond, rather than subsecond times should have been used. There are two reasons to believe that this may be the case. The first is that suprasecond times are longer than subsecond times. Use of suprasecond times may make it easier to more accurately observe or experience the time that is shown and allows for a greater difference between durations to be used. Essentially, increasing the size of the difference in durations may make time more salient and increase its availability as a memory cue.

The second is that suprasecond times have been shown to be served by right hemisphere structures while subsecond times have been linked to left lateralised structures (Lewis & Miall, 2003b). In addition, the structures associated with suprasecond time, but not subsecond time, are also used for spatial working memory tasks. Essentially suprasecond timing tasks may be more similar to spatial tasks than subsecond timing tasks would be. As the Hourihan et al. (2007) study found an effect related to space, perhaps this link between spatial processing structures will make an effect of suprasecond timing more likely to occur and/or be detected.

In contrast, there is also reason to expect suprasecond times to be even less likely to produce an effect than subsecond times. Unlike subsecond times suprasecond times are argued to be cognitively controlled (Lewis & Miall, 2003b). Rather than suprasecond times being measured automatically attention is generally thought to be required to accurately track their passing. In a directed forgetting paradigm such as that used in the previous experiment attention is not directed towards the duration at any point in the experiment. Furthermore attention would be expected to be directed towards rehearsal of remember words. As such, without any explicit cue to focus on the change in the delay, participants could be expected to entirely ignore the temporal component of the stimulus.

However, attentional control is only one candidate in the use of this system (Lewis & Miall, 2003b). It is possible that non-attended longer durations may still be encoded with the right lateralised system. In fact it could even be argued that this right lateralised bias may increase the odds of an effect being found. Specifically, if it is assumed that the use of word stimuli in the current memory task makes use of verbal working memory this implies that spatial working memory is not being used. As there is overlap between spatial working memory and the right lateralised system used for suprasecond time it is possible that suprasecond time may be registered more easily than subsecond timing in the current task.

When these arguments are combined they suggest important implications if the directed forgetting benefit observed in the spatial task is reproduced with suprasecond but not subsecond times. First it helps to reinforce the idea that subsecond and suprasecond times may use different systems and representations. Specifically, it would suggest that suprasecond times produce a similar result to those of spatial cues while subsecond times do not. This would suggest that time and space may be processed similarly as long as the correct time class is used. This would fit with findings that suggest similar structures are used in spatial working memory and suprasecond timing tasks. Furthermore it would suggest that the time divisions commonly used to describe prospective timing also apply to retrospective timing. This would help to fill the gaps in this particular area. Finally it would suggest that suprasecond times can be encoded automatically or, at the very least, non-deliberately.

It is important to first select times that are comfortably in the suprasecond range. While the previous experiment suggested that the subsecond range might be expected to end at 1500ms it was also pointed out that this was a conservative estimate. Some theorists argue that subsecond timing could extend to times of 3 seconds or longer (Gibbon et al., 1997; Ulbrich et al., 2007). In fact some have suggested that counting can extend subsecond timing indefinitely (Hinton & Rao, 2004). In order to avoid these issues while remaining within the bounds of the experiment, times ranging from 4 seconds (4000ms) to 10 seconds (10000ms) will be used. Times will change in 2000ms steps to maintain a linear representation of time (Wearden & Ferrara, 1995) in line with that produced in experiments 1A and 1B. Finally, at the end of the experiment, participants will be asked if they noticed that the duration between the fixation cross and the word coming on screen varied across trials. If they say they did notice a difference they will also be asked whether or not they timed/counted these delays. This will be done in an effort to check whether participants were aware of the durations varying during the experiment and also identify any participants who may have engaged in chronometric counting[24].

Methods.

Participants. Enough participants to balance gender, response key, word list assignment, study balance, and tone were collected. After participation the remember-cued and forget-cued accuracy score of each participant were tested with chi-square analysis. If no significant difference was found between the remember-cued and forget-cued accuracy scores that participant was excluded from analysis and replaced. In total 103 volunteers were recruited from the University of Auckland. All participants reported normal or corrected to normal vision and were able to hear and discriminate between a 260Hz and an 1170Hz tone. Participant handedness was measured using the Edinburgh handedness inventory (Oldfield, 1971). 64 passed the selection criteria (N_{males}=32, M_{age}=22.29, $\text{Range}_{\text{age}}$=18-27, 28 Right Handed, 4 Left handed; N_{females}=32, M_{age}=22.29, $\text{Range}_{\text{age}}$=17-44, 30 Right Handed, 2 Left Handed)[25]. Participants were evenly split between designs using the original peaked distribution and the flat distribution. None of these participants reported counting the durations and those who did notice the variability in the delay did not report believing that it was important to the experiment[26].

Materials. The same materials were used as in experiments 1A and 1B. The same word lists were used in each balance to ensure a fair comparison between the current version and the previous version of the experiment. The key differences were the change in times from 500, 800, 1100, and 1400ms to 4000, 6000, 8000, and 10000ms respectively as well as the inclusion of a marble draw to randomly allocate participants to a participant number.

[24] Instructions to avoid chronometric counting would invalidate the experiment as it would suggest that time is important in the task. This could cause the task to move from being retrospective to being prospective in terms of timing.

[25] Participants were recruited purposively to ensure equal numbers of males and females in the experiments and in each condition. As no gender effects were found, the practice proved impractical, and Hourihan et al. (2007) did not balance males and females, this is the only experiment in which this occurs.

[26] Often participants attributed variation to fact that the computers were “pretty old”.

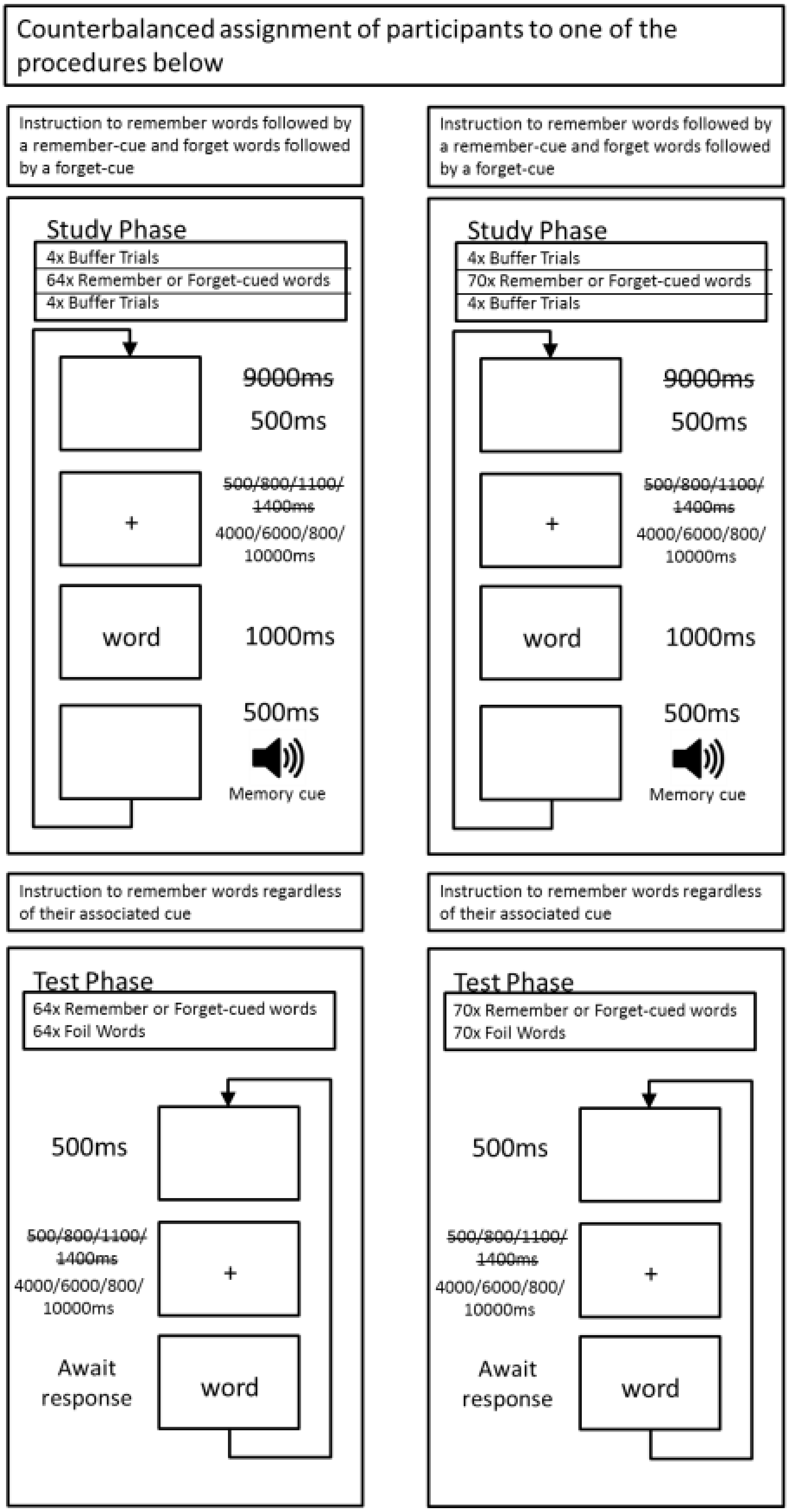

Figure 5.1. The experimental procedure for experiment 2. Stimuli are for illustrative purposes only and are not drawn to scale.

Procedure. The procedure was the same as that used in experiments 1A and 1B with the following exceptions. Participants were randomly assigned a subject number that allocated them to the peaked distribution or flat distribution of temporal distances. These distributions were counterbalanced along with the variables that had been counterbalanced in experiments 1A and 1B.

The delay between word offset and fixation cross onset was returned to 500ms. This is in accordance with Hourihan et al. (2007) and helps keep the average duration between the presentations of words more comparable between the time experiments (6950ms vs. 7500ms instead of 6950ms vs. 1300ms.)[27] The procedure, and its differences to that used in experiment 1, are shown in Figure 5.1.

Results.

Selection and screening*.* Chi-square tests were conducted on individual participants comparing the number of remember words that were recognised to the number of forget words that were recognised. Participants showing a significant difference between the number of remember and forget words recognised were separated from those who did not. Data for the included participants is presented in Figure 5.2.

Separate 2(memory cue) x 2 (temporal location) ANOVAs were conducted on the correct reaction time and accuracy data. Significant main effects of memory condition were found for accuracy ($F_{(1,63)}$=509.43, p<.001) and reaction time ($F_{(1,63)}$=38.64, p<.001). This indicates that remember-cued words are recognised faster and more frequently than forget-cued words. No main effect of temporal location (same vs. different) was found nor was an interaction found in either case.[28]

Simple contrasts. As the effect is expected to be very small, powerful and simple contrasts were initially proposed. It was expected that remember-cued words and forget-cued words would differ in terms of accuracy, reaction time, or both and that there would be a difference in these measures between words in the same and different temporal locations in the forget-cue condition. This comparison would also be replicated in the remember-cue condition for symmetry and in case the effect existed in both memory categories. While this implies the possibility of an interaction no interaction was initially tested for.

[27] While the failure to produce an exact match is a mild concern a larger difference in the pacing of trials in the study phase is later shown to have no significant effect on performance in experiment 3.

[28] If 2x7 Repeated Measures ANOVAs are conducted the same pattern of results is produced. The 2x2 ANOVAs are reported here as they should have had more statistical power and allowed for the retention of more participants. This is also in line with the recommendations of the section on correcting A’ and d’.

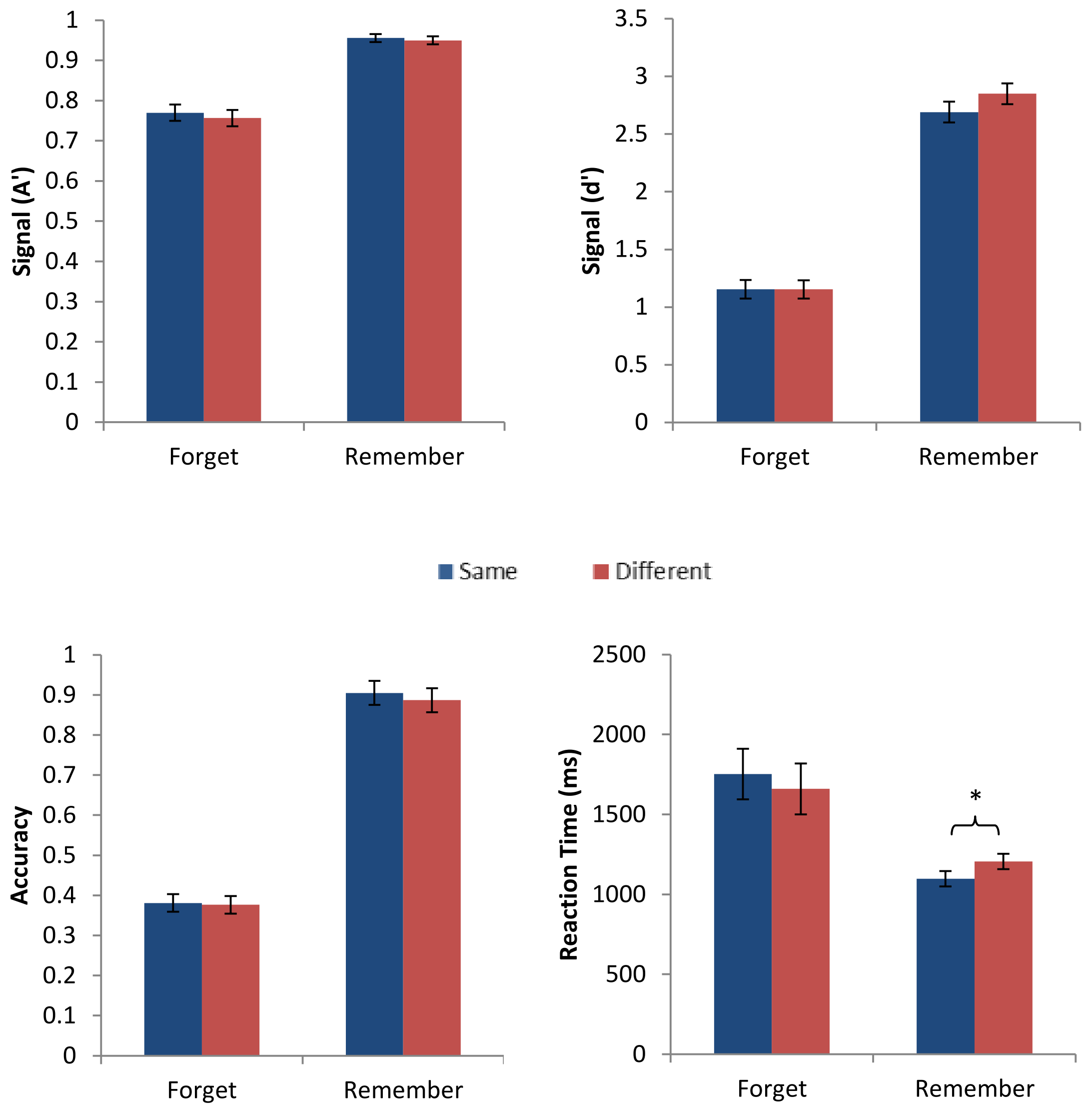

Figure 5.2. Accuracy, reaction time, and corrected signal value data for the remember-cued and forget-cued words in the long time version of directed forgetting task as a function of whether the fixation time at study and test was the same or different. The false alarm rate in the foil condition was .08. The reaction time to foils was 1367ms. An * is used to denote where planned contrasts indicate that there is a significant difference (p<.05) between locations within a memory condition. Within subjects error bars (Masson & Loftus, 2003) represent the planned contrasts between temporal locations within memory cue conditions e.g. same vs. different temporal locations for remember cued words.

Accuracy. A contrast comparing remember-cued and forget-cued words was conducted. A significant difference was found indicating that remember-cued words were recognised more frequently than forget-cued words ($F_{(1,63)}$=509.43, p<.001). Contrasts comparing the same time and different time words were conducted separately for the remember-cued and forget-cued words. There was no evidence to suggest that same and different time words differed in either memory cue condition.

Reaction time. When analysing reaction times for correct responses the data of 8 participants was excluded due to missing values (accuracy score of 0) in one or more conditions. A contrast comparing remember-cued and forget-cued words was conducted. A significant difference was found indicating that remember-cued words were recognised more quickly than forget-cued words ($F_{(1,55)}$=38.64, p<.001). Contrasts comparing the same time and different temporal location words were conducted separately for the remember-cued and forget-cued words. There was no evidence to suggest that same and different time words differed in the forget-cued condition ($F_{(1,55)}$=0.30, p>.05) but the same vs. different contrast in the remember-cued words was significant ($F_{(1,55)}$=6.61, p=.013). This suggests that remember-cued words presented at the same temporal location at study and test were recognised faster than those presented at different temporal locations at study and test.

This finding was tested for robustness. The first test was to see if the pattern would still emerge if all participants were included rather than only those who met the chi-square criterion that tests for a significant difference in accuracy between the remember-cued and forget-cued conditions. In this case the pattern did not re-emerge ($F_{(1,93)}$=1.87, p>.05). The pattern was then tested when only remember-cued data was considered. Where the standard analysis drops participants with missing reaction time data for remember-cued or forget-cued values this analysis only drops participants with missing reaction time data for remember-cued words. This change in criterion leads to the retention of more participants in the analysis. When the chi-square criterion was employed the effect was significant ($F_{(1,63)}$=8.37, p=.005). When the criterion was not used the effect bordered on significance ($F_{(1,101)}$=2.94, p=.09). Figure 5.3 compares the data between the

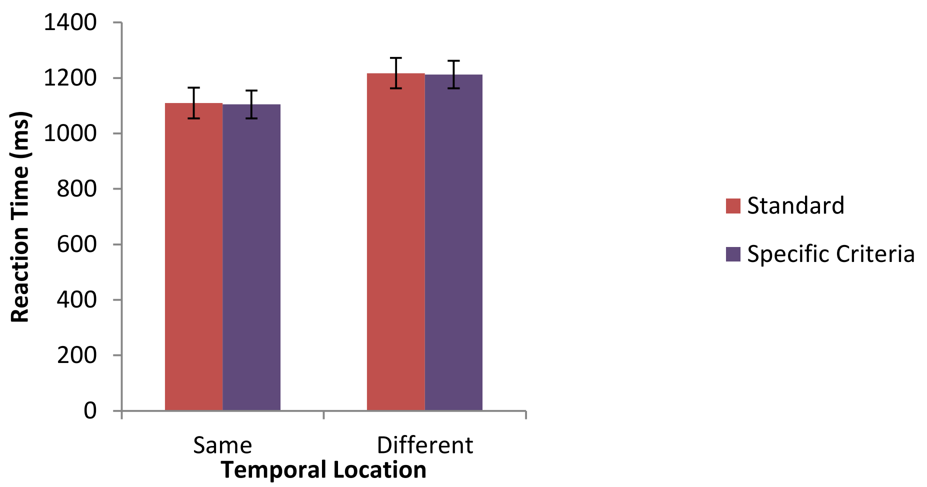

Figure 5.3. The patterns obtained when selection considers remember-cued and forget-cued words (standard) compared to when it only considers remember words (specific criteria). Within subjects error bars (Masson & Loftus, 2003) represent the planned contrasts between temporal locations.

remember-cued conditions where the selection criterion is employed and significant differences are found.

Signal. In accordance with the re-categorised results from experiments 1A and 1B presented in chapter 4, the data was also analysed using the corrected A' and d' signal measures outlined in chapters 3 and 4. Separate 2 (memory cue) x2 (temporal location) ANOVAs found a significant effect of memory condition for A' ($F_{(1,63)}$=202.68, p<.001) and d' ($F_{(1,63)}$=381.05, p<.001). The interactions with and main effects of temporal location were not significant. Contrasts found no significant effects between the same and different conditions in either the remember-cued or forget-cued words.

Discussion. As expected a significant difference was found between memory cue conditions. This difference between accuracy, signal, and reaction times for remember-cued and forget-cued words is common in the literature and strengthens the argument that participants are completing the task in the manner that would be expected. In the case of accuracy, this difference is also ensured by the selection process used in the current design.

The effect of temporal location is less conclusive. In standard ANOVAs no main effect of temporal location is apparent for accuracy, signal, or reaction time nor is there a significant interaction between memory condition and temporal location. Only when very targeted contrasts performed on a subset of participants while examining remember and forget words separately are used does any effect emerge. Furthermore, unlike the Hourihan et al. (2007) study, this effect is in the reaction time data rather than the accuracy data and the remember-cued words rather than the forget-cued words. While separate analysis of remember-cued and forget-cued data is justified, given the contrasts used in the Hourihan et al. (2007) study and the expected lack of salience of time in the current design, the fact that effects are only visible if examined in this way does raise the possibility that this finding is the result of a type 1 error. Given recent findings around issues with the .05 criterion (Johnson, 2013), the failure to observe the finding in anything other than the reaction time data, and the fact that the reaction time data does not completely counterbalance other variables[29] caution is warranted.

In exercising such caution it is necessary to consider whether there is a justifiable reason to expect such a difference to occur. The finding of a difference in reaction time in the remember-cued

---

[29] Reaction time data comes from participants with at least one correct response in each condition. As some participants fail to meet this criteria their data is excluded from the analysis leading to a failure to counterbalance. While none of the counterbalanced variables show a significant effect in the analysis this shortcoming is worth considering. This problem has been reduced by categorising the times as same and different rather than the categories originally used in experiments 1A and 1B but it has not been eliminated.

but not the forget-cued words was not expected and opposes the findings of the Hourihan et al. (2007) study. To recap, the Hourihan et al. (2007) study found that matched spatial location aided memory for the forget-cued but not the remember-cued words and this was argued to be due to remember-cued words having more memory cues in place than the forget-cued words. Essentially the provision of an extra memory cue was more beneficial for forget-cued words than remember-cued words. The current results seem to suggest the opposite. Somehow the provision of an extra memory cue is more beneficial for remember-cued words, where memory cues are already in place, than forget-cued words, where fewer memory cues should exist. Additionally it leads to faster recognition rather than more frequent recognition.

If the finding is assumed to be true, the most immediate explanation is that temporal and spatial information are treated differently in the directed forgetting paradigm. While this is not incompatible with an episodic memory account, under which where and when data are implicitly presented as different, it is less compatible with accounts that propose a shared representation for time and space. Why two concepts with a shared representation should impact on different memory classes (remember-cued and forget-cued) is not clear under symmetrical or asymmetrical accounts.

A better explanation might come from ignoring the distinction between the remember-cued and forget-cued conditions. Under such an account space and time both aid recognition of words rather than the recognition of a particular class of words[30]. Under such an interpretation a common representation for time and space makes sense.

In explaining the shift from a difference in accuracy data to a difference in reaction time the suggestion would be that a speed accuracy trade-off is occurring. This suggests that participants in the current study are behaving differently to those in the Hourihan et al. (2007) study. As reaction times in the forget condition are slower in the current study than those in the original study this may suggest that participants are waiting to be sure in the forget-cued condition. As such the effect seen here occurs where certainty is highest, the remember-cued condition, whereas the original study showed the effect in the condition where certainty was lowest, the forget-cued condition.

Finally, the fact that the difference only reaches significance when participants reaching the selection criteria are met could be argued to be in line with the idea of salience. As the effect size is expected to be quite small and the number of trials is relatively low, the presence of an effect may only be detectable when participants adhere strongly to the requested division between remember and forget words. Participants who segregate the words into two distinct categories may be

[30] While this would imply that a main effect should have been observed the low number of trials may make the division necessary in order to reveal the effect.

reducing the variability in the results allowing for better detection of any underlying effect. Unfortunately as the criteria is new to this series of experiments there is no way to know if this would also be the case in the Hourihan et al. (2007) experiment.

Conclusion. It is unclear whether or not the current results show an effect of time location on memory. As the only significant effect requires specific contrasts to be used and does not entirely match the results of Hourihan et al. (2007) the safest conclusion is that an effect of time on memory does not exist. However, as it is possible to explain the results attained through reinterpreting the Hourihan et al. (2007) study, it would be best to re-conduct a spatial version of the experiment using the current study population and its selection criteria before any further discussion is considered.

Chapter 6. Directed Forgetting and Space

After the failure to clearly reproduce the pattern of results observed by Hourihan et al. (2007) using subsecond and suprasecond times it becomes important to determine if the pattern observed in the original study can itself be replicated. The apparent incompatibility of the results could suggest that time and space do not share a common representation in memory, changes in experimental protocol lead to changes in the results, or that no spatial or temporal effect exist.

Adding to the complications are differences in the inclusion criteria used between the Hourihan et al. (2007) study and the time experiments. Unlike experiments 1 and 2, participants in Hourihan et al. (2007) were only excluded from analysis if their data suggested that they had not understood the task correctly. This criterion led to the exclusion of 1 out of 32 participants with no replacement participant being used. In contrast, experiments 1 and 2 excluded any participant who did not show a significant difference between the number of remember-cued and forget-cued words that were recognised. This typically led to the exclusion of about 1 in 3 participants, each of which was replaced to complete counter-balancing. Importantly this difference in selection criteria can already be shown to make a difference: when the Hourihan et al. (2007) criteria are used in the time experiments no effect of temporal location is found. This difference points to a potential issue with the use of the significant memory effect criterion. Specifically, the differences observed in the time studies may be the result of the change in criterion rather than the change in stimulus itself. While such a change of inclusion criterion has been justified, on the basis of increasing the number of participants adhering to the task requirements to aid in finding what was likely to be a small effect, there is presently no means of investigating what impact such a criterion would have had if it had been used in the Hourihan et al. (2007) study. As such there is no real way of knowing if the results are truly comparable.

This also draws attention to another potential issue. Given the small size of the Hourihan et al. (2007) space effect it becomes important to determine if the space effect can be replicated. Suspicion that the effect seen in Hourihan et al. (2007) results may be the product of type 1 error is amplified by the fact that the results contradicted original predictions made by the experimenters. Where Hourihan et al. (2007) had expected to see that forget-cued words appearing at the same location at study and test would be recognised less frequently than those appearing at different locations at study and test, they actually observed the opposite pattern. Had a one way test for the expected effect been invoked the results would not have been significant.

Before making a call on the presence or absence of a temporal location effect in the recognition item method directed forgetting paradigm, it is important to determine whether or not a spatial location effect exists. Confirmation or disconfirmation of this result is necessary for proper interpretation of the time experiments and in terms of recommendations for future use or non-use of this paradigm.

Experiment 3

In this experiment the intention was to replicate Hourihan et al. (2007). Unfortunately an error in the program caused the time between words to differ from that of Hourihan et al. (2007) for the first 34 participants who completed the task. Conveniently this error leads to the delay between words producing a near match to the average delay in the long time version of the study. These results were retained and a second set of participants ran under conditions properly replicating those of Hourihan et al. (2007). Data from both runs are presented below. For ease of reference the first run is referred to as experiment 3A, the second 3B, and the combined data is referred to as experiment 3.

Methods.

Participants. Volunteers were recruited from the University of Auckland student population. All participants reported normal or corrected to normal vision and stated that they were able to discriminate between the low (260Hz) and high (1170Hz) tones. Handedness was measured using the Edinburgh handedness inventory. The chi-square criterion was not employed at this point.

In experiment 3A 34 participants were recruited. Of these, 2 failed to complete the experiment. This left 32 participants (N_{males}=5, M_{age} = 20.80, $Range_{age}$ = 17-26, all right handed; $N_{females}$=27, M_{age} =21.74, $Range_{age}$ = 17-32, 24 Right Handed, 1 Ambidextrous, 2 Left Handed).

In experiment 3B an additional 34 participants were recruited. Of these, 2 failed to complete the experiment. This left 32 participants (N_{males}=10, M_{age}=21.23, $Range_{age}$ = 18-29, 9 Right Handed, 1 Ambidextrous; $N_{females}$=22, M_{age}=25.09, $Range_{age}$ =18-54, 17 Right Handed, 2 Ambidextrous, 2 Left Handed).

Materials. The same materials as experiments 1 and 2 were used.

Procedure. The procedure outlined here applies to experiment 3A. Unless otherwise specified the procedure is the same as that used for experiment 1A.

As in Hourihan et al. (2007) words appeared on screen, one at a time, in one of four spatial locations. Words appeared at the top, bottom, left, or right of the screen. In all cases they were 20% from the nearest edge and centred on the midline. This variation in spatial location replaces the variation in temporal location used in previous experiments. Clarification of the locations used is provided in Figure 6.1.

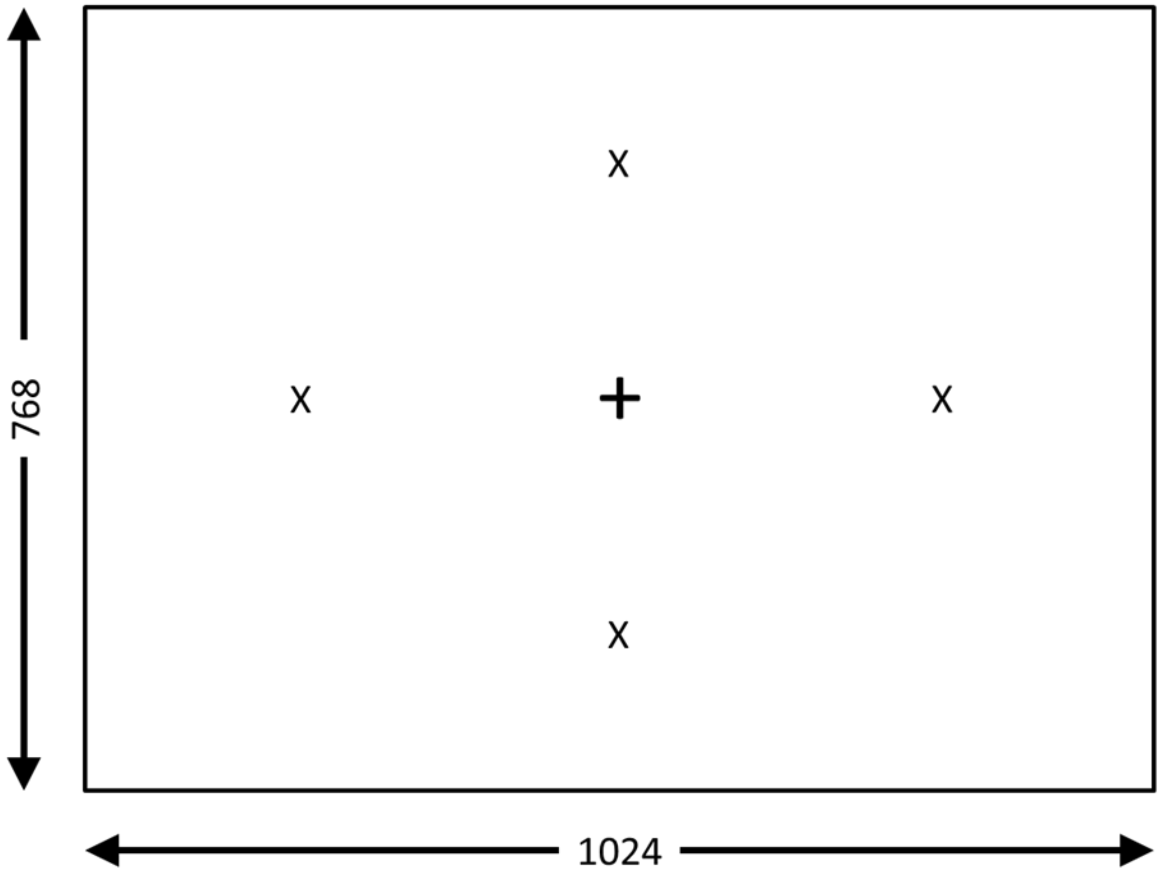

Figure 6.1. Screen positions of words for experiment 3, marked by xs. The figure is for illustrative purposes only and is not drawn to scale.

As in experiment 1A, words were randomly assigned to a study-test location pairing at the start of the experiment such that every location used at study was paired with each location at test. This is a more restricted version of the random pairings used by Hourihan et al. (2007). Unlike the Hourihan et al. (2007) technique, this technique ensures all locations are paired with all others in all conditions.

During the study phase trials were preceded by a 6000ms fixation cross. After the cross a word was shown on screen for 1000ms and was immediately followed by a 500ms tone indicating whether participants should try to remember that word. The screen remained blank for 3000ms then the process repeated itself until all of the word and tone pairings had been displayed.

During the test phase words were preceded by a 6000ms fixation cross and stayed on screen until participants made their response.

In experiment 3B, the only difference to experiment 3B was the change in the duration of the fixation cross in the test section from 6000ms to 500ms. This was done to match the delays used in the Hourihan et al. (2007) study. Both procedures are outlined in Figure 6.2.

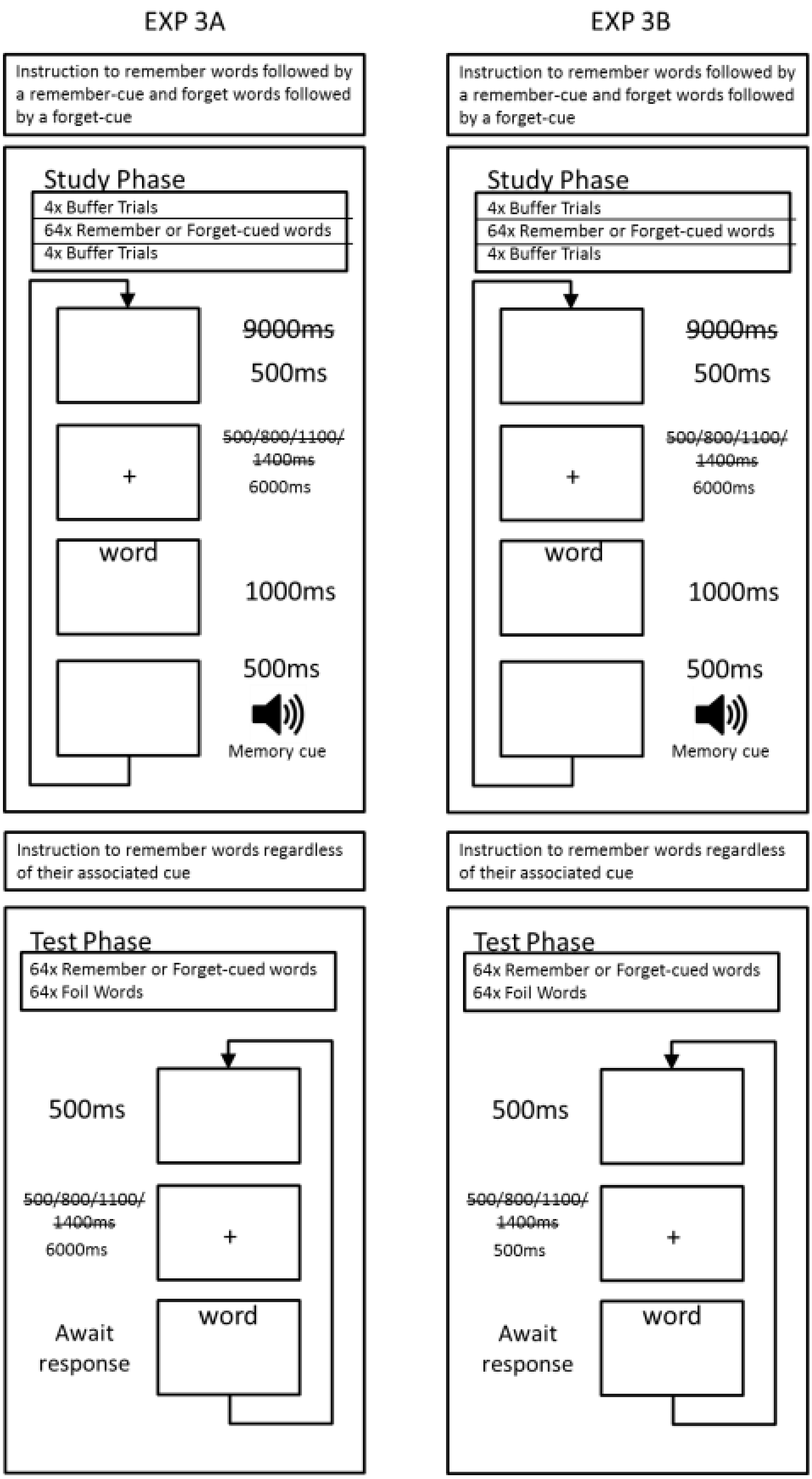

Figure 6.2. Experimental procedures for experiments 3A and 3B. Stimuli are for illustrative purposes only and are not drawn to scale. In each case, the depictions show a same location condition.

Results. Separate mixed design $2_{(experiment)}x2_{(memory\ cue)}x2_{(temporal\ location)}$ ANOVAs for accuracy and reactions time were conducted with experiment as a between participants factor and memory condition and spatial location (same vs. different) as within subjects factors. As there was no significant main effect of experiment and no significant interactions involving experiment in either case data was collapsed across experiments 3A and 3B. Only the combined results are presented here.[31] A summary of this data is presented in Figure 6.3.

Accuracy. A significant main effect of memory cue condition was found ($F_{(1,62)}$=178.05, p<.001) suggesting that remember-cued words are recognised as old words more frequently than forget-cued words. While there was no significant main effect of spatial location ($F_{(1,62)}$=2.01, p>.05) a significant interaction between memory condition and spatial location was found ($F_{(1,62)}$=4.76, p=.033). Planned contrasts found a significant difference between words in the same location and different location in the remember-cue condition ($F_{(1,63)}$=7.33, p=.009). This suggests that remember-cued words in the same position at study and test were recognised more frequently than remember-cued words that appeared at different locations at study and test. There was no significant difference between forget words presented at the same location and those presented at different locations at study and test ($F_{(1,62)}$=0.15, p>.05).

When 21 participants who did not show a significant difference between the number of remember-cued and forget-cued words they recognised were excluded the pattern of results remained the same. There was a significant main effect of memory condition ($F_{(1,41)}$=259.86, p<.001), a significant interaction between memory condition and location ($F_{(1,41)}$=7.94, p=.007), and a planned contrast found a significant difference between remember words in the same location and those in different locations at study and test ($F_{(1,41)}$=5.85, p=.020).[32]

[31] The pattern of results is the same if data from experiment 3B is considered without including the data from experiment 3A. Combined data is presented here as it provides a better representation of what is likely to be happening at the population level and also allows for a mock replication of the selection criteria used in the time experiments.

[32] When participants (N=21) who do not show a difference between the number of remember-cued and forget-cued words recognised are analysed on their own no significant differences between the same and different temporal location results are found ($F_{(1,20)}$=1.78, p=.20). A follow up analysis found no significant interaction relating to time location between participants that did and did not meet the inclusion criterion ($F_{(1,62)}$=1.27, p=.27) nor was a difference found when looking at the relative size of the difference between the remember and forget cued words between groups ($F_{(1,62)}$=0.89, p=.35), visual inspection of the data also suggests that participants who do and do not meet the inclusion criterion are trending in the same direction as those who do. As a result, the failure to find a significant difference in the participants is likely to be due to insufficient sample size (N=32 or more in previous work). An interaction between memory condition was found ($F_{(1,62)}$=51.26, p<.001) suggesting that the difference between performance on remember and forget cued words was greater for those meeting the selection criterion than those not meeting the selection criterion. This suggests that the criterion is working as intended.

Reaction time. When analysing reaction times for correct responses, the data of seven participants with an accuracy score of 0 in one or more word conditions was excluded. The ANOVA revealed a significant main effect of memory condition with remember-cued words being responded to faster than forget-cued words ($F_{(1,56)}$=24.74, p<.001). The main effect of location, the interaction, and the planned contrasts all failed to reach significance although the contrast testing for a

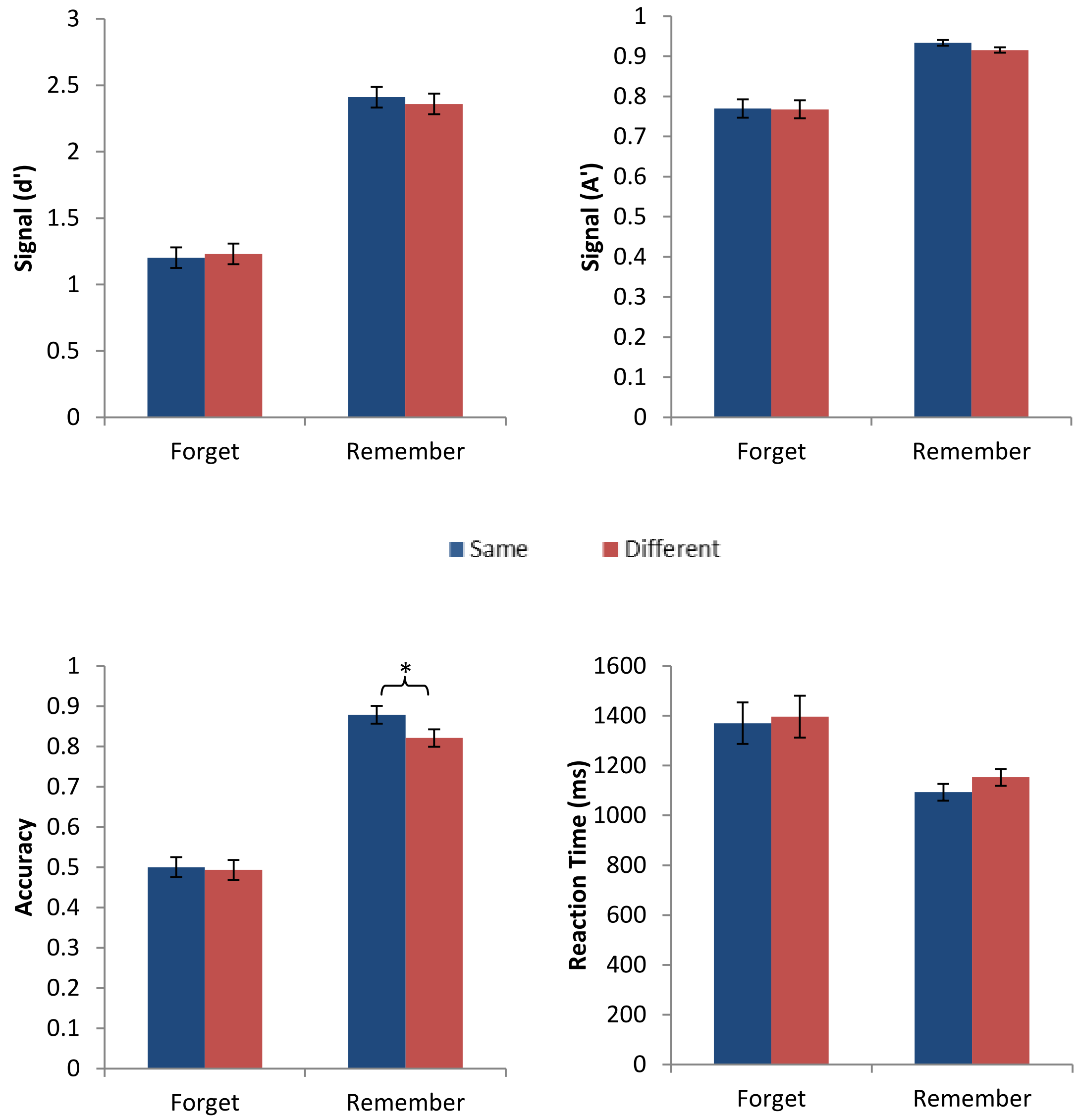

Figure 6.3. Accuracy, reaction time, and corrected signal data for the combined space data when all participants are considered. Data is plotted as a function of the location of the words at study and test (same or different) memory condition (remember-cued words or forget-cued words). The false alarm rate in the foil condition was .10. The reaction time to foils was 1272ms. An * is used to denote where planned contrasts indicate that there is a significant difference (p<.05) between locations within a memory condition. Within subjects error bars (Masson & Loftus, 2003) represent the planned contrasts between temporal locations within memory cue conditions e.g. same vs. different temporal locations for remember cued words.

difference between spatial locations in the remember-cued words did approach significance ($F_{(1,56)}$=3.83, p=.055).

When participants who did not show a significant difference between the number of remember-cued and forget-cued words they recognised were excluded this pattern, shown in Figure 6.4, remained the same. The only significant difference was for the main effect of memory condition ($F_{(1,35)}$=22.04, p<.001) although the remember contrast did approach significance ($F_{(1,35)}$=3.31, p=.077).

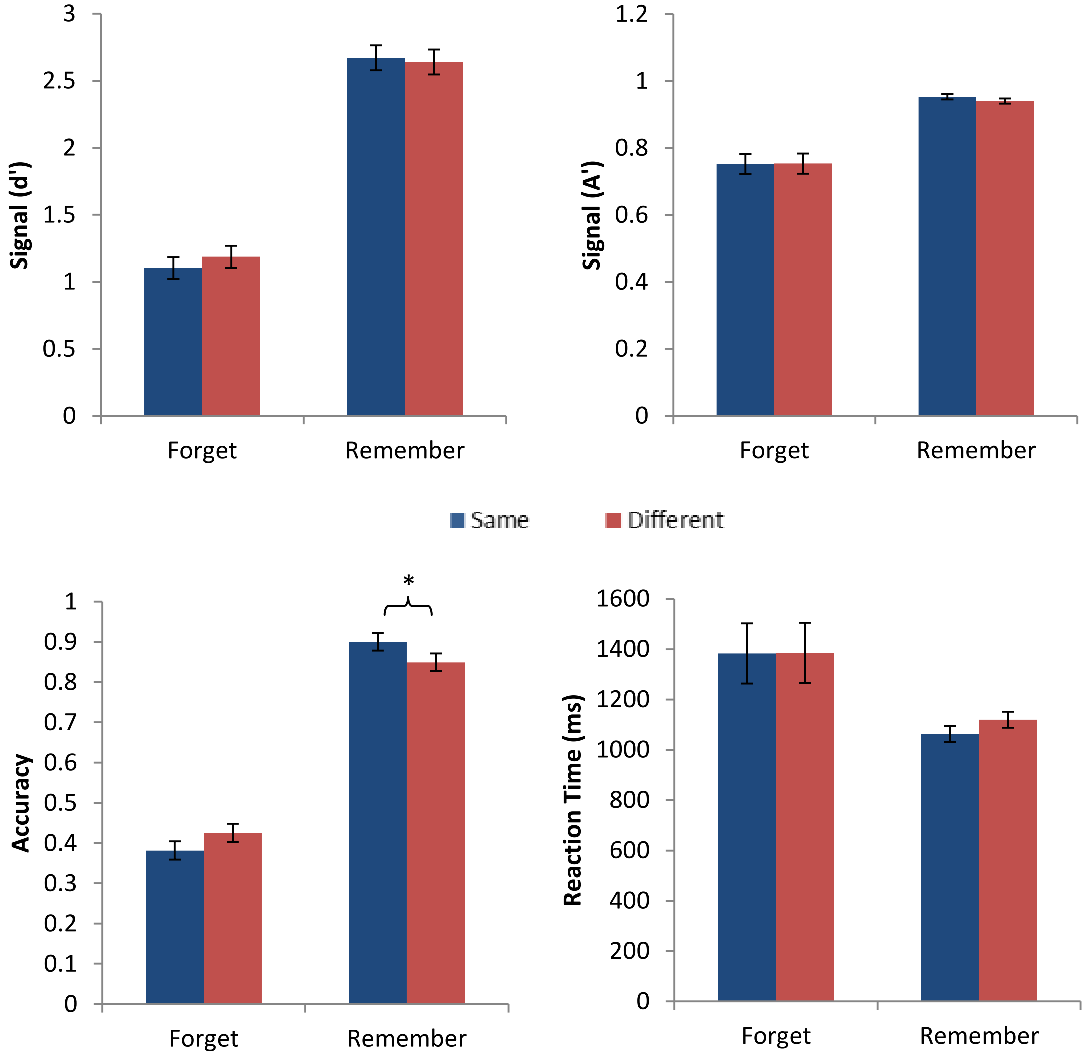

Figure 6.4. Accuracy, reaction time, and corrected signal data for the combined space data when only participants meeting the selection criterion used in the temporal experiments are considered. Data is plotted as a function of the location of the words at study and test (same or different) memory condition (remember-cued words or forget-cued words). The false alarm rate in the foil condition was .08. The reaction time to foils was 1223ms. An * is used to denote where planned contrasts indicate that there is a significant difference (p<.05) between locations within a memory condition. Within subjects error bars (Masson & Loftus, 2003) represent the planned contrasts between temporal locations within memory cue conditions e.g. same vs. different temporal locations for remember cued words.

Signal. In accordance with the re-categorised results from experiments 1A and 1B presented in chapter 4, the data was also analysed using the corrected A' and d' signal measures outlined in chapters 3 and 4. ANOVAs revealed a significant main effect of memory for A' ($F_{(1,62)}$=123.24, p<.001) and d' ($F_{(1,62)}$=160.68, p<.001). No other comparisons reached significance. This pattern was unchanged when the inclusion criterion from experiments 1 and 2 was employed.

Discussion. The results did not differ between the slow paced and standard paced experimental design, therefore both experiments were combined. This increased power and increases confidence in the findings. Importantly the results broadly replicated those of Hourihan et al. (2007). Furthermore when selection criteria in line with those used in the time experiments were used the pattern of results did not change and effect sizes actually increased.

Visual inspection of the plots suggests that the significant interaction between memory condition and spatial location is the result of a same location benefit for the remember-cued words and a same location cost for the forget-cued words. Tested on their own, only the difference in accuracy between remember-cued words at the same and different location reaches significance. While the visual pattern seems to show some support for Hourihan et al. (2007) original hypothesis, wherein forget-cued words presented in the same spatial location should be harder to recognise, statistical analysis does not support this.

While the change from the benefit appearing in the forget-cued words to the remember-cued words does not replicate the findings from the Hourihan et al. (2007) study, it is not incompatible with their interpretation. In particular, in a follow up experiment in Hourihan et al. (2007) no significant difference was found in the memory for location information between memory conditions. Essentially, there was no evidence to suggest that spatial information was encoded better for forget-cued words than remember-cued words.

While it may seem strange to need to make such a statement, as it implies that spatial information linked to remember-cued words might somehow be worse than that associated with forget-cued words, there are reasons to expect this to be the case. In fact these are the same reasons that Hourihan et al. (2007) originally outlined in support of their finding that spatial information may create a larger benefit for forget-cued words compared to remember-cued words. Remember-cued words are argued to have undergone more elaborative rehearsal. Through this process remember-cued words come to be associated with more memory cues, thus diluting the impact of the spatial information.

Regardless, in the current experiment this dilution does not appear to have happened to the same extent or has not had the same impact as it did in Hourihan et al. (2007). An alternative explanation could be that participants in the current experiment are somehow not as close to producing a ceiling effect in the remember-cued words. Conversely it may be possible that participants are producing some kind of floor effect in the forget-cued words. Such accounts are compatible with the idea of participants in the current experiments behaving in a more conservative manner than those in the original Hourihan et al. (2007) experiment.

This may be related to the fact that participants in the Hourihan et al. (2007) study participated in exchange for course credit whereas participants in the current experiments participated without being offered any incentives. Some work has suggested that participants who participate in exchange for some incentive may be less motivated to perform well than those who participate on a purely voluntary basis, particularly if the incentive is of a low value (Gneezy & Rustichini, 2000; Heyman & Ariely, 2004).

If the volunteer participants were more motivated than the incentivised participants their remember-cued word performance could be expected to be higher than that of the incentivised participants. Furthermore less contamination of forget-cued words with remember-cued word behaviour may occur if volunteer participants are more motivated to follow the instructions of the task. It may be possible that the reason for the switch from an effect in the forget-cued words to the remember-cued words could be tied to such a pattern of responding. If such an account is correct it could be expected that participants who show a more pronounced difference between the remember-cued and forget-cued conditions should show a more pronounced pattern of results. If the criteria used in the temporal studies is used here this is exactly what is found with the p-value for the interaction between memory cue condition and spatial location dropping to .007.

Conclusion. Spatial location information does appear to be beneficial in recognising words in a recognition memory paradigm. Unlike Hourihan et al. (2007) a benefit for matched spatial locations at study and test was found for the remember words instead of the forget words. While the current experiments did not directly replicate the results of Hourihan et al. (2007) the differences may have been due to differences in participant motivation. This is supported by the finding that the current pattern is strengthened when the selection criterion from the time experiments is used.

Chapter 7. General Discussion: Directed Forgetting

Whether the current results are in line with ideas that assume that space and time share a common representation depends on how liberally they are interpreted. If concessions are granted to the time results, on the basis of time having less salience and the work being more exploratory, some support for a common representation can be found. However, at the most conservative level the results from the current line of experiments do not support this notion. The spatial benefit seen by Hourihan et al. (2007) was reasonably well replicated while the time results only met significance when investigated using specific criteria and measures.

Time Works

If matched time information does produce a benefit it only does so under specific circumstances. In order to observe an effect of time on recognition it was necessary to consider data from a subset of the sample and employ a specific contrast: one that considered the difference between reaction time in same and different temporal location words in the remember-cued words only.

While the use of the performance based selection criterion did not match the approach originally taken by Hourihan et al. (2007) the pattern of results found in the suprasecond study was backed up in the replication of the spatial study. Despite the spatial effect appearing in the accuracy data and the time effect appearing in the reaction time data, both datasets show an effect in the remember-cued words that lead to a similar conclusion. In each case reinstatement of the relevant cue lowered a barrier to recognition. If this discrepancy were to hold across future replications it would suggest that time and space were differentially influencing some form of speed accuracy trade-off.

In terms of the issue related to addressing a subset of the participants it is important to note that the use of the selection criterion in the spatial version of the design increased the significance of the interaction while decreasing the size of the sample. Essentially a manipulation that should have reduced statistical power increased the significance of an effect. This suggests that the criterion is beneficial and probably fulfils its intended role of excluding participants who are performing the task incorrectly. Put another way, it removes variability from the sample without sacrificing the detectability of the signal.

Another criticism that could be made about the temporal finding was that the reaction time data was not effectively counterbalanced as some participants were dropped due to an accuracy rate of 0 in one or more conditions. While this is true the same criticism applies to the accuracy data of the Hourihan et al. (2007) study. In the Hourihan et al. (2007) study participants who did not reach inclusion criteria were excluded but not replaced thereby preventing effective counterbalancing. This did not appear to impact the pattern of results as it was replicated by the spatial version used here.

Finally, the current data suggest that if time acts as a memory cue the times used must be in the suprasecond range rather than the subsecond range. This is consistent with the idea of time and space sharing a common representation when a right lateralised network is used. It also suggests that prospective time divisions may apply to retrospective timing. However this final point does come with additional important caveats. The first is that the suprasecond durations used longer durations and longer differences between durations. As such it is unclear which of these differences is important for retrospective time memory cues.

Different Representations - Assume Time Fails

In terms of parsimony - given the number of conditions that must be met, the small effect size, and the lack of clarity in interpretation - for the time being it appears safest to suggest that there is no effect of time on the likelihood of participants to recognise words. On the other hand the fact that a benefit for matched spatial location was found in experiment 3 as well as Hourihan et al. (2007) suggests there is an effect of space. When considering the surrounding literature there are a number of reasons why this might be the case.

Perhaps the most reasonable claim is to suggest that time and space do not share a common representation in memory. While prospective time work has shown a number of relationships between spatial and temporal processing (Walsh, 2003) it is important to remember that this link has not been established with retrospective timing paradigms. In fact it is not even clear whether the timing systems used in prospective timing are used in retrospective timing. Whereas prospective time is talked about in terms of suprasecond and subsecond time, in memory time typically is not divided at all or is spoken about as microtime, the time between events, and macrotime, the time at which an event happens (Hassabis & Maguire, 2007). With mounting evidence suggesting that memory for events is both unreliable and more of a reconstruction than a replay (Hassabis & Maguire, 2007; Loftus, 2005) these unattended elements may not be faithfully recorded and/or replayed at all, preventing them from being used in subsequent recognition.

While the prospective timing literature did provide reasons to expect time to show an effect at both the subsecond and suprasecond levels, respectively its apparent automaticity (Lewis & Miall, 2003b) and relationship to spatial processing (Walsh, 2003), it was also possible that there would be no effect. First, assuming that there is something special about space in its ability to provide information that will aid in the recognition of remember or forget words, there is no real reason to expect subsecond times to show a space like effect on memory. Unlike suprasecond time, subsecond timing is argued to activate a left hemisphere network with no real overlap with spatial processing (Lewis & Miall, 2003b).

Second, while the association of suprasecond times with a network implicated in spatial working memory suggests an overlap with space may occur, other aspects of the system suggest it cannot occur. Specifically, cognitive (Lewis & Miall, 2003b) and explicit (Coull & Nobre, 2008) timing both suggest that participants would need to direct their attention toward temporal information in order to encode time information. In essence, this suggests that unattended time might as well have not happened under such systems. At no point in any of the experiments was attention directed towards the passing of time. In fact attention was actively directed away from time. This may mean that participants were essentially unaware of any differences in timing occurring at all. Put more simply, as time was not noticed it could not be encoded. As it was not encoded it could not act as a memory cue. This idea is supported by a number of participants expressing surprise when being told that the time of the fixation cross was varied throughout the experiment.

Future Directions: Assessing the Directed Forgetting Paradigm's Use in Retrospective Tasks

Despite the failure to find a clear effect of temporal location on recognition the current study design has made important advances in the field of retrospective timing. The clearest advance is the shift from basing retrospective timing tasks on a single trial. While the degree of success associated with this approach is unclear in this case, it does provide a starting point for future work. In fact the current results suggest that slight variations to the current paradigm may be beneficial for future retrospective work.

If the reaction time benefit observed in suprasecond time in experiment 2 is accepted as real, future work may benefit from removing the division between remember-cued and forget-cued words. Removing the forget-cued condition is supported by four key findings. First, effects were found in the remember-cued words for suprasecond time locations and spatial locations. As suprasecond times have been argued to rely on spatial representations this makes sense. Second, these effects are strongest when the selection criterion is employed. In these cases participants were producing a clear difference in behaviour between the remember-cued and forget-cued conditions. Third, the effects were most easily detected when the remember-cued words were considered in isolation. Finally, a shift to only using remember-cued words has a key benefit: it would increase the number of trials where a temporal location effect would be likely to be detected. This increase in trial number should increase the power of the resulting analyses. Given the small size of the effects such an increase in power is highly desirable.

The way in which the stimuli were displayed may also be important. Critically the spatial information that helped participants recognise words in the test phase of experiment 3 and

Hourihan et al. (2007) was locked to the word itself. In the temporal version of the task the temporal information preceded the word and was actually attached to the fixation cross instead. This may have prevented participants from associating the temporal information with the word. Instead they may have associated it with the fixation cross only. Future research may benefit from considering pairing the duration with the word rather than the fixation cross.

While this is simple enough in the study phase it becomes more difficult at the test phase. In the study phase the time each word is on the screen is determined by the experimenter. During the test phase participants are encouraged to respond as quickly and as accurately as they can. Words remain on screen until a response is made. Essentially the time the word is on the screen falls under the participant's control. In order to preserve the temporal information associated with the word, responses would have to be prevented until the interval had passed. As participants would be ready to respond prior to the time location occurring it is likely that reaction time information would have to be discarded.

Future work should also consider the implications of the way in which retrospective time is represented. While the current division into subsecond and suprasecond time experiments assume that retrospective time may obey the same rules that prospective time does, including that time information in memory will be preserved reasonably accurately, this may not be the case. When looking at the way that time information appears to be represented in episodic memory researchers typically refer to time as more like a label than an actual duration (Hassabis & Maguire, 2007). This may mean that, as with episodic memory itself (Hassabis & Maguire, 2007), it is likely that time information is fluid: time information may be open to reinterpretation and outright replacement.

This may raise a question of whether future designs should explicitly draw attention to time locations in order to increase the reliability of time estimates and increase the chances of unearthing an effect. While this would probably make time more equivalent to the spatial information presented, as shifts in space rapidly draw exogenous attention, such a change would invalidate the paradigm. Retrospective timing refers to time that has passed without participants explicitly being asked to take note of it: it is the judgement of time information from memory rather than through some online process. Drawing attention to time is likely to shift timing to an online process, thereby leading to the investigation of prospective rather than retrospective time.

That being said it may be fair to inform participants that time information may be a useful cue for recognition after the study phase. While this may create a shift towards prospective time in the test phase, successful use of temporal location cues would rely on the existence and successful use of retrospective timing for the study phase.

Combining an instruction to attend to time at test with linking the time to word stimuli suggested earlier could also allow for the analysis of reaction time data, particularly if a design similar to that used in the second Hourihan et al. (2007) experiment was used. In this situation participants would be asked to indicate whether the time the word was displayed at test matched the time the word was displayed for at study. This would also shift the focus of responses away from memory and towards time. Such a change is highly desirable.

Conclusion

Retrospective time is a difficult area to investigate. The current line of experiments has produced some data tentatively suggesting that suprasecond time may work as a memory cue in a similar way to space. However this effect can only be found using contrasts conducted under very specific conditions. While the effect is sufficiently fragile that concerns exist this it may be spurious, the experimental protocol has shown some promise. Future research should work on increasing time salience and/or statistical power through optimising the design. Suggestions include removing the forget-cued condition, pairing time with the word rather than the fixation cross, drawing attention to time after the study phase, and changing the task to indicating whether the duration of a word at test matches its duration at test. Such changes should maintain the retrospective nature of the task.

Chapter 8. Moving to Prospective Time

Why Look into Kappa and Tau Effects?

While retrospective timing work suffers from a lack of established paradigms prospective timing does not. Where retrospective timing tasks require participants to be unaware of the importance of time information until testing there is no such requirement for participants in prospective tasks. This means individual participants can run and attend to times in multiple trials per experiment. This difference in the paradigms is likely to be a key contributor to the wider body of work available in prospective time.

Indeed, attention to time plays a strong role in the theory surrounding prospective timing tasks. The most influential theory of prospective time perception, STT, has been referred to as one of a gated accumulator (Zakay & Block, 1997). Under such a model, ticks are collected by an accumulator as they occur. However ticks can only be collected if they are attended to. This explains why time can seem to fly when having fun and drag when there is nothing else to capture attention. At an experimental level it also explains how including attention demanding tasks leads to the underestimation of prospective time (Predebon, 1996).

While the STT model does not hold a spatial component within it, models that advocate an overlap of temporal and spatial information can be considered to be a modification to STT. In such cases the accumulator can be thought to gather a measure of magnitude rather than the number of ticks of an internal clock. As overlaps in spatial and temporal information are often relatively small (e.g. kappa and tau effects) this implies that the accumulated space, time, and number estimates are separated by porous barriers as depicted in Figure 8.1. In this way the medium used to represent time, space, and number is the same, hence they can be said to share a common representation, but the storage space of each is different. As representational units leak from one storage space to another, representations of time, space, and number become confused.

While a case for the overlap of time and space, potentially related to spatial working memory, was outlined in the general introduction, one key question was left unanswered. Which of the paradigms available is likely to best serve the purpose of investigating relationships between space and time?

While more complex means of assessing overlap between time and spatial information have been used, the most direct measure appears to be kappa and tau effects. Unlike more complex paradigms that look for time effects that are equivalent to space effects, such as temporal neglect (Danckert et al., 2007), kappa and tau effects are found in paradigms that should allow for space and time to be represented completely independently. Yet, for the most part, they are not.

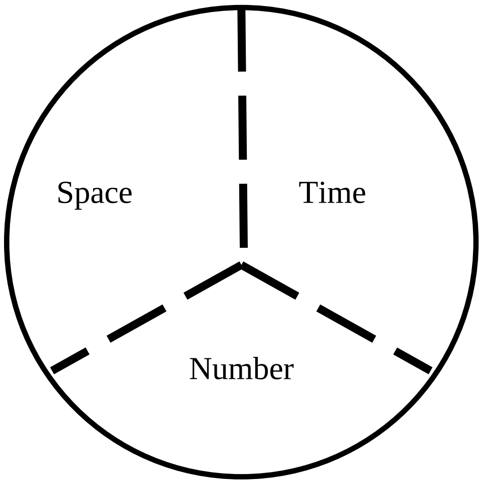

Figure 8.1. A depiction of the relationship between the representation of space, time, and number compatible with that proposed under ATOM (Walsh, 2003). All three occupy separate areas of a common representation of magnitude. They are separated by porous barriers meaning that information representing any one concept can leak into the representation of any other concept.

In a standard kappa-tau experiment participants are presented with both spatial and temporal information on each trial. In each case they are asked to attend only to the spatial or temporal information. In each case the irrelevant dimension impacts on the perception of the relevant dimension. In this respect kappa and tau effects can be thought of as illusions as they lead to predictable differences between presented stimuli and perceived stimuli.

Kappa effects were first reported by Cohen, Hansel, and Sylvester (1953) and relate to the impact of spatial information on perceived time. Typically, larger spatial stimuli are associated with the perception of longer intervals. Tau effects owe their name to Helson and King (1931) and describe the opposite scenario: longer intervals leading to the production of longer lines. Notably, while the spatial and temporal information is generally conveyed visually through the presentation of two points of light, kappa and tau effects have also been found in other modalities (Kawabe, Miura, & Yamada, 2008).

The most popular account of kappa and tau effects contains a role for speed. In a core study and review of the area, Jones and Huang (1982) considered 5 key explanations (relativity, physiological excitation, temporal duration discrimination, temporal order effects, and imputed velocity) and concluded that imputed velocity provided the best account of kappa and tau effects. Under imputed velocity, kappa and tau effects are said to be the product of participants combining presented spatial and temporal information into a representation of velocity. This association with speed can then lead to distortions in perception of space and/or time if the subsequent stimuli do not produce the expected speed, a suggestion that is in line with more recent work involving a Bayesian model of responses (Goldreich, 2007). Both Jones and Huang (1982) and Goldreich (2007) argue that the system is intended to resolve rather than create uncertainty. Key to Jones and Huang's (1982) preference for the imputed velocity account was the fact that it could explain findings accounted for by the other explanations in addition to findings showing modifying effects of acceleration (Cohen, Hansel, & Sylvester, 1955), and the existence of kappa and tau effects across a number of paradigms and modalities (recent examples include Cai, Connell, & Holler, 2013; Firmino, Bueno, & Bigand, 2009; Henry & McAuley, 2009; Henry, McAuley, & Zaleha, 2009; Jones & Huang, 1982; Kawabe et al., 2008).

The relationship with imputed velocity was expressed more formally by Jones and Huang (1982).

$$V= s_1/t_1 = s_2/t_2 \tag{8.1}$$

In this equation imputed velocity is represented by V, space is represented by s, and time is represented by t. Space and time each have numbers beside them to refer to two sets of stimuli that are presented one after the other, 1 representing the first pairing and 2 representing the second. However, in line with the suggestion put forward in Goldreich (2007), these numbers can also refer to the expected relationship (1) and the observed relationship (2) (Jones & Huang, 1982). In either case V is formed by the combination of space and time in s_1/t_1. If the relationship is the same in s_2/t_2 the requested spatial or temporal response can be expected to be accurate. Using the example where a spatial response is expected, s_2 and t_2V will both produce the same estimate. If the imputed velocity is not the same in each pairing, inaccuracy can be expected. Using the same example where a spatial response is expected, s_2 and t_2V will not produce the same estimates. Assuming that space is fixed ($s_1= s_2$) and the first time is shorter than the second time ($s_1/t_1 =V <s_2/t_2$) the estimate of s_2 can be expected to be longer than that of s_1. Put more simply, if the same space is paired with a longer time a larger estimate of space should be expected.

While imputed velocity is still the dominant account of kappa and tau effects there have been some challengers. The first challenge comes from Sarrazin et al. (2004). In a series of experiments participants viewed eight dots presented sequentially on screen from left to right. Temporal delays and spatial gaps between stimuli could vary (V) or remain constant (C). This gave rise to four conditions; CC where temporal delays and spatial gaps both had a fixed value; CV where all the temporal delays were the same but the spatial gaps varied; VC where the temporal delays varied but the spatial gaps were all the same; VV where temporal delays and spatial gaps varied with one another. For clarity, all of the Sarrazin et al. (2004) conditions are represented in Figure 8.2. The majority of their results supported the predictions of an imputed velocity account. Specifically,

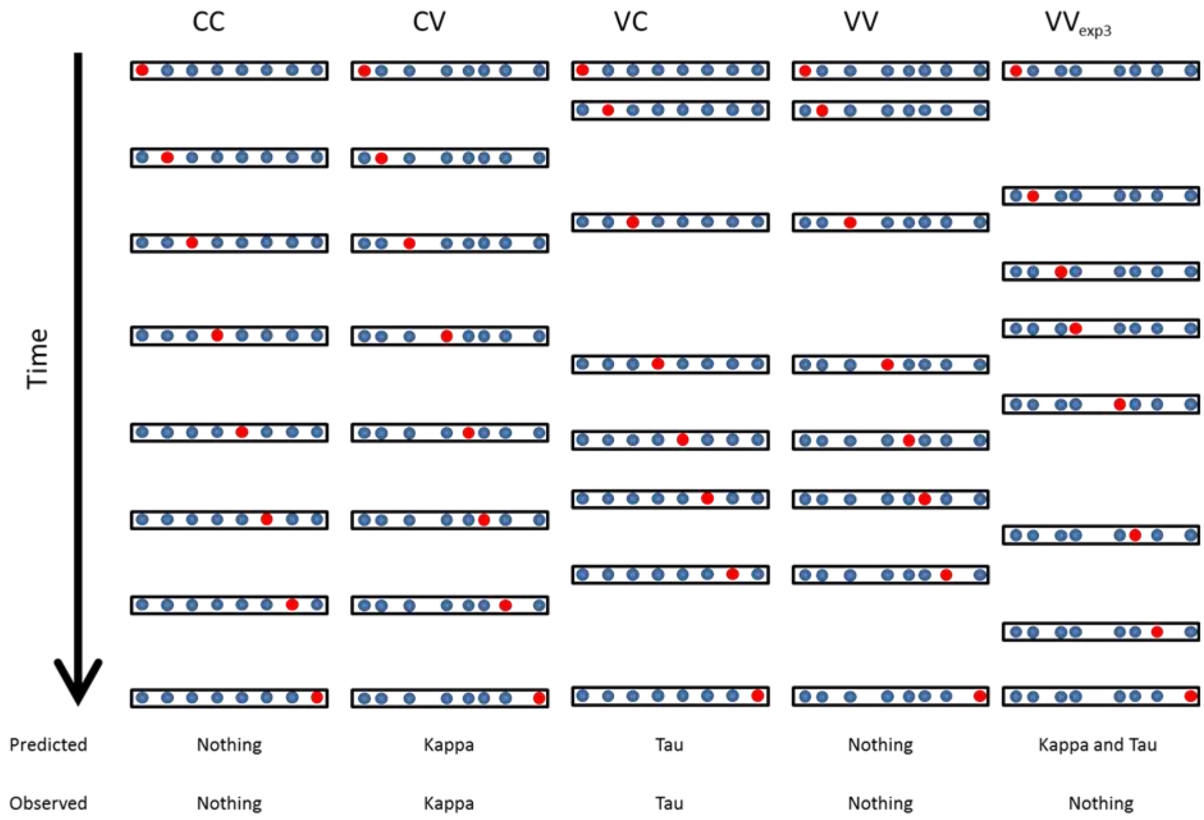

Figure 8.2. A pictorial representation of the conditions used in the Sarrazin et al. (2004) experiments. Condition names are presented at the top. Predictions under an imputed velocity account and the actual findings are presented at the bottom for each condition. Blue dots represent the spatial position of inactive stimuli while red dots represent the stimulus that would be showing on screen at that time point. The figure is representative only and does not provide a scaled replication of the actual spatial lengths and durations used in Sarrazin et al. (2004).

condition CV should and did produce kappa effects, condition VC should and did produce tau effects, and conditions CC and VV should not and did not produce kappa or tau effects.

The challenge to the imputed velocity account comes from one crucial change in their third experiment. Sarrazin et al. (2004) reasoned that if using temporal delays and spatial gaps that maintained imputed velocity by varying *with each other* (condition VV) prevented kappa and tau effects, using temporal delays and spatial gaps that violated a constant imputed velocity *by not varying with each other* (condition VV_{exp3}) should produce both kappa *and* tau effects. However VV_{exp3} failed to produce either a kappa or a tau effect.

While the differences observed in Sarrazin et al. (2004) are potentially fatal to the imputed velocity explanation there are reasons to discount them. The first is that no alternative explanation was offered. In fact, the experimenters state that they are unsure of what might have led to such results. Second, kappa and tau effects in time ranges similar to those used by Sarrazin et al. (2004) have proven unreliable in the past (Jones & Huang, 1982) suggesting that there may have been a design issue. Finally, and most damningly, the results in this failed condition had substantially more variability than any other condition. This once again suggests that there may have been general issues with the condition itself that may have masked the effects or prevented them from emerging at all. As such, while interesting, these results are not sufficient to put an end to the imputed velocity account.

The other key challenge to an imputed velocity account comes from Casasanto and Boroditsky (2008). In a series of studies Casasanto and Boroditsky (2008) argue that imputed velocity can be suppressed and is relevant to the tau effect but not the kappa effect. Like a standard kappa-tau paradigm, varied pairings of space and time were presented to participants. Unlike a standard kappa-tau paradigm, stimuli were made of static lines, growing lines, or travelling specks instead of two lights flashed in different locations. Across six versions of their kappa-tau-like paradigm the same finding was produced: space impacted on perceived time (a kappa effect) but time did not impact on perceived space (no tau effect). This was argued to support the metaphorical account of time representation. Specifically, language enables us to use space to think about time, meaning that time perception should be more dependent on spatial perception than the reverse. Their findings are controversial for two key reasons. The first is that these findings are out of step with previous kappa-tau literature. The second is that this inconsistency is explained by arguing that the paradigm suppresses imputed velocity.

While Casasanto and Boroditsky's (2008) results were out of step with the common finding of kappa *and* tau effects, their paradigm was different to common practice. In fact, judging by the approach taken in the discussion, it appears that their paradigm was constructed without knowledge of, or an attempt to tap into, kappa and/or tau effects. Unlike traditional kappa-tau experiments these experiments used growing lines, dots travelling at a fixed speed, or complete lines instead of brief presentations of dots on screen. This methodological difference was later argued to prevent the perception of imputed velocity as an actual speed, rather than imagined, speed was supplied in the growing and moving conditions and speed was unlikely to be perceived in the static line condition.

While this may be true, it is a problematic explanation for the Casasanto and Boroditsky (2008) results as the argument is circular: the absence of tau effects is attributed to the lack of imputed velocity, and since imputed velocity itself has not been measured, the lack of a tau effect is argued to show that imputed velocity is lacking. Furthermore kappa and tau effects have been shown to be robust across experimental protocols that have included variation in stimulus modality and judgement procedures such as reproduction and forced choice (Jones & Huang, 1982). Beyond this, unlike the phi phenomenon or illusory line motion tasks, imputed velocity does not necessarily involve the perception of illusory motion (Jones & Huang, 1982). Instead it is the result of participants making use of *knowledge* about the relationship between space and time. As such it is conceivable for imputed velocity to impact on the stimuli that Casasanto and Boroditsky (2008) used as long as participants are still able to perceive differences in space and time. Finally, if imputed velocity has been successfully suppressed, there is no clear reason to predict a kappa effect but not a tau effect. As kappa *and* tau effects are argued to be the result of imputed velocity (Jones & Huang, 1982) it follows that suppression of imputed velocity should suppress kappa and tau effects.

An alternative possibility as to why these findings were out of step with previous literature is that perhaps the changes in the times used simply were not equivalent to the changes in the spaces used. If the changes in time were perceived as smaller, and therefore potentially harder to detect, than the changes in space, it might be reasonable to expect the changes in time to have less of an effect than the changes in space. Indeed such a relationship between the length of time intervals and their ability to produce tau effects has been shown before (Jones & Huang, 1982) and the potential for such a difference was later recognised by at least one of the authors (Bottini & Casasanto, 2010).

The need to equate the dimensions of time and space is highlighted further when considering one of the ways in which an asymmetrical relationship was implied. Casasanto and Boroditsky (2008) note that the slope for the tau effect, with time measured in ms, was smaller than the slope for the kappa effect, with space measured in pixels. This was used to imply that space impacts on time more than time impacts on space. However there is no real reason to assume that ms, rather than seconds for example, is the correct time unit to use. In fact if time were to be measured using a coarser scale, such as seconds or minutes, a supposed asymmetry in the trends could be reversed. This highlights the need to use a standardised metric when suggesting whether a given relationship is symmetrical or asymmetrical[33].

Furthermore, no consideration was given to whether sub or suprasecond times were being used in the experiment. While kappa and tau effects have been found separately in both subsecond (Jones & Huang, 1982; Lebensfield & Wapner, 1968) and suprasecond intervals (Bill & Teft, 1972; Jones & Huang, 1982; Pricewilliams, 1954) kappa and tau effects have been found to be less consistent at the subsecond level (Jones & Huang, 1982). The times used in the Casasanto and Boroditsky (2008) study ranged from 1000ms to 5000ms in 500ms steps. This particular range of

[33] In fairness the original study also reported a much larger effect size for the kappa effect than the tau effect. It would have been better for them to focus on effect sizes and ignore slopes as effect sizes are already standardised.

times may be problematic as, in the worst case scenario, this could mean half of their intervals would be using one system while the other half would be served by another. Worse still the boundary point between subsecond and suprasecond timing has been argued to differ from participant to participant (Coull & Nobre, 2008; Gilden & Marusich, 2009; Lewis & Miall, 2003b) and even task to task (Coull & Nobre, 2008; Lewis & Miall, 2003b). Each of these factors has the potential to increase variability in the data and that could obscure any effect that exists. As kappa and tau effects are very small any increase in variability could mask their presence. Furthermore, results from one study that definitely crossed these boundaries suggest that kappa and tau effects were strong when the two classes of time were separate but diminished when the intervals were closer together (Bill & Teft, 1972). While this crossing was not considered at the time the work was published it does seem a reasonable explanation for the findings produced and would help explain why a discrepancy exists between the Casasanto and Boroditsky (2008) study and the wider kappa-tau literature.

Measuring Imputed Velocity

While Jones and Huang (1982) proposed that space and time could be related through imputed velocity, and experiments have demonstrated that actual and perceived velocity influences kappa (Brown, 1995) and tau (Jones & Huang, 1982) effects, there is no clear measure for the level of imputed velocity that individuals are likely to associate with a stimulus. If such a measure could be found it could be expected that participants that are inclined to expect items to move faster or slower might have varying tendencies to experience imputed velocity from on-screen stimuli. In theory, participants whose expectations fall in-line with the imputed velocity created in the kappa/tau paradigm could be expected to show kappa and/or tau effects of a different strength to participants whose expectations are not in line with what is presented on screen.

A candidate measure of imputed velocity comes from the scanning paradigm used in experiments that investigated visual imagery (Kosslyn, Ball, & Reiser, 1978). Scanning describes a process where participants imagine a small speck travelling as quickly as it can, while still remaining visible, from one point on an image to another. Participants press a key when the speck begins moving and press the key again when the speck arrives at the final destination. Scanning has been used as a means to assess the accuracy of mental representations of stimuli that participants had been asked to remember (Jolicoeur & Kosslyn, 1985; Kosslyn et al., 1978; Kosslyn, Margolis, Barrett, Goldknopf, & Daly, 1990; Kosslyn, Reiser, Farah, & Fliegel, 1983). When stimuli were free of manipulations scan times were in line with the true physical distance between objects, increasing linearly with actual distances (Richman, Mitchell, & Reznick, 1979). While debate continues over whether or not such studies can be used to demonstrate the nature of visual imagery (Pylyshyn, 1973, 1981) it is generally accepted that scanning is a measure of some form of mental representation. Additionally, as kappa effects have been shown to rely on perceived spaces (Lebensfield & Wapner, 1968) rather than actual spaces, working with a procedure that taps into the mental representations of spatial distances makes sense.

Furthermore Jones and Huang (1982) specify that imputed velocity is not something that is an emergent property of the stimuli e.g. the phi phenomenon. Instead imputed velocity reflects the *knowledge*[34] that that times and spaces are related to each other through speed rather than the perception of this relationship. One example of this is the finding that the kappa effect is modulated by gravity (Masuda, Kimura, Dan, & Wada, 2011). It is stronger when participants move from top to bottom (in line with gravity thereby being associated with greater speed) rather than bottom to top (moving against gravity with the expectation of less speed) (Masuda et al., 2011). In this sense using scan-speed to assess this subjective understanding of the relationship between space and time holds merit as scan-speeds are also modulated by beliefs about movement difficulty (Jones & Huang, 1982; Masuda et al., 2011) whereas emergent phenomena, such as the phi phenomenon, are not affected by such things (Jones & Huang, 1982). Finally if imputed velocity is the result of knowledge about how space and time might be related rather than the perception of some kind of illusory motion there is reason to believe that it may still be present in the static line task.

For the current experiments the speed at which the speck travels is the key component of interest. The fact that the scan estimates typically have a strong linear relationship with actual distances (Kosslyn et al., 1978) suggests that participants imagine scanning at a constant speed in these tasks. This imagined speed element might be useful in assessing the speed at which participants expect things to travel. In other words by obtaining an estimate of a participant's scan-speed it is hoped that an estimate of their maximum visible imputed velocity may be obtained. After all, what is scan-speed if not the maximum visible imputed velocity?

[34] Conveniently this is not problematic under Pylyshyn's (1973, 1981) account of mental imagery results as Pylyshyn suggests that 'mental imagery' is really just knowledge based rather than image based.

Chapter 9. Experiment 4

In an attempt to test the assertions made by Casasanto and Boroditsky (2008), namely that time does not impact on space in their paradigm and the use of static lines rather than separate locations prevents imputed velocity from influencing results, their experimental procedure will be replicated with some key changes. First longer time intervals will be used to increase confidence that a single time system, the suprasecond system, is being used. Second, a scanning task will be introduced to provide an index of imputed velocity for each participant. Assuming that scanning provides an index of imputed velocity, if time is based on space and the presentation of lines prevents imputed velocity there should be no tau effect and no relationship between scan time and either kappa or tau effects. Third, conditions that should be the least likely to produce kappa and tau effects will be used. Specifically, static lines will be presented on screen and verbal cues will be used to identify the required tasks to the participants. Static lines were argued to prevent imputed velocity while verbal cues were originally largely avoided in an effort to avoid imposing language based metaphors on participant responses (Casasanto & Boroditsky, 2008).

While the use of verbal cues does differ from the initial Casasanto and Boroditsky (2008) experiment there is little reason to believe that they will contribute to the finding of a tau effect. Firstly, they were avoided in the original study because the investigators felt that the use of verbal cues would make it less likely to find a tau-like effect. This was argued to be the case as the assumption was the bias in language to talk about time as space rather than space as time reflects the underlying relationship between space and time processing. As such if language were used this verbal asymmetry may be highlighted, leading participants to favour space over time. Assuming their suspicions were correct, the use of words instead of symbols in the probe phase should make it less likely to find a tau effect in the current experiment not more likely.

Secondly, verbal cues were not entirely avoided in the Casasanto and Boroditsky (2008) study. In fact from experiment 2 onwards words were shown on screen before each trial indicating whether a space trial or a time trial would follow. As no difference was found between any of their experiments there is little reason to believe that the inclusion of word stimuli impacts on the results in a meaningful way.

Methods

Participants. In total 44 volunteer undergraduate and postgraduate students from the University of Auckland with normal or corrected to normal vision participated in this study. The final dataset consisted of 34 participants (N_{males}=5, M_{age}=20.60, $Range_{age}$=18-24, 4 Right Handed, 1 Ambidextrous; $N_{females}$=29, M_{age}= 23.55, $Range_{age}$=18-49, 24 Right Handed, 4 Ambidextrous, 1 Left

Handed) after 10 participants who did not complete the task or performed one or more of the tasks incorrectly were removed[35].

Materials. The experiment was coded with E-prime and was run using E-run. The experiment was presented in a dimly lit room on a 450MHz Windows 98 Pentium III computer with 128MB of RAM and a 17 inch VGA monitor operating at 60Hz and a resolution of 1024x768 pixels. A second computer was also used. The second computer had a 15 inch VGA monitor. Otherwise the second computer was identical to the first. Participants were seated 57cm from the screen[36]. Responses, in the form of click location and the time between clicks, were collected from the input of a standard computer mouse.

Nine lengths of horizontal line and 9 durations were used in the experiment. Line lengths were measured in pixels and ranged from 200-800 pixels in 75 pixel steps. Durations ranged from 3000-7000ms in 500ms steps[37]. All line length duration combinations were presented once in each of 3 conditions (space, time, and scan) to each participant in a random order during the experimental block. This made for a total of 243 trials per participant excluding practice trials. In line with Casasanto and Boroditsky (2008), lines were presented on the vertical midline of the screen. Lines were randomly allocated to align with the right or left of the screen. In either case the distance between the end of the aligned side and the closest end of the line was randomised between 62 and 162 pixels horizontally. Responses were collected from crosses that were randomly allocated to one of the locations shown in Figure 9.1. Participants were instructed to make their first response by clicking on the cross in all cases. If their second response required them to click another location they were instructed to click to the right of the cross if the cross appeared on the left. If the cross appeared on the right their second response was to be to the left. Absolute horizontal distances and the time between the first and second click were recorded and used for calculating the results. Vertical distance was ignored.

---

[35] If these participants are included the pattern of results remains the same but confidence in the overall data is diminished. Two participants did not complete the experiment while the other participants' data suggested they were performing the wrong task at the wrong time and/or showed a negative correlation on the time judgement, space judgement, or scanning task.

[36] Ideally the screen size would have been matched between computers or the seating distance would have been varied in order to maintain the visual angle occupied by stimuli. Unfortunately the difference in screen size was only noticed once a substantial number of participants had been run in experiments 4 and 5.

[37] These step sizes are in line with those used by Casasanto and Boroditsky (2008). The times used here are longer. Line lengths are exactly the same in pixels, although Casasanto and Boroditsky (2008) have not provided enough information to replicate the visual angle occupied by their line lengths.

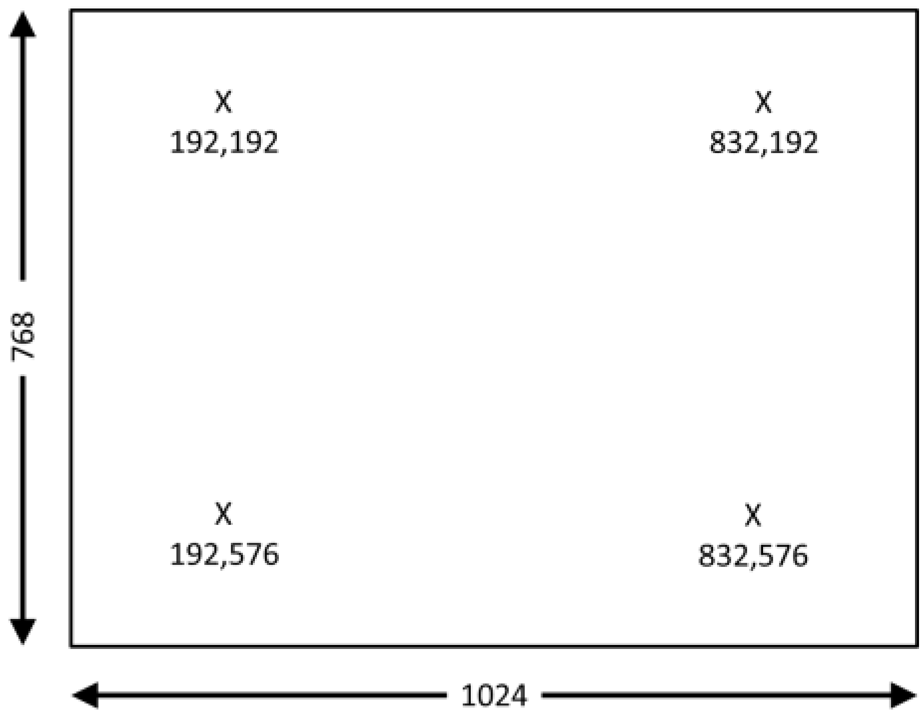

Figure 9.1. Anchor points on the screen. Values are expressed in pixels. When paired they are presented in the format x,y. The figure is for illustrative purposes only and is not drawn to scale.

Procedure. Trials began with a word in the centre of the screen indicating what type of trial participants would be taking part in. The word was SPACE, TIME, or SCAN and it was visible for 1000ms. After a 1000ms blank screen a randomly determined pairing of length and duration was then presented in a randomised location as outlined previously. When the line disappeared the word, SPACE, TIME, or SCAN, was shown in the centre of the screen again along with an X indicating where participants should make their response. The position of the X was randomly assigned to one of the locations in Figure 9.1.

In space trials participants were asked to reproduce the length of the line by clicking the first X then clicking a second location of their choosing. The nature of the starting positions meant that in some cases participants would make their second response to the right of their first response, in others their second response would occur to the left.

In time trials participants were asked to click the first X, wait until they felt that the same amount of time had passed, then click the X again. Time production using this method has been shown to be more accurate than alternatives (Mioni, Stablum, McClintock, & Grondin, 2014). The space and time reproduction tasks are also in line with those used in Casasanto and Boroditsky (2008).

In scan trials participants were asked to imagine a small speck travelling as quickly as it can while still remaining visible from one end of the line to the other. Participants were asked to click the

X when the speck left one end of the line and click the X again when it arrived at the other end of the line. This form of instruction is in line with that used by Kosslyn et al. (1978).

In line with Casasanto and Boroditsky (2008) the cue indicating task type disappeared when participants first clicked the X. The X remained on screen until the second click was made. At this point the next trial began. The general trial procedure is depicted in Figure 9.2.

The experiment began with instructions and practice blocks for each of the tasks in the experiment. Each practice block consisted of 10 trials of a single task type (time, space, or scan). In time blocks line lengths were fixed at the average length (500 pixels) and durations varied. In scanning and space trials the duration was fixed at the average duration (5000ms) and line lengths varied. Participants were free to run through the practice blocks as many times as they felt necessary. Participants were informed that vertical information from their spatial responses would be ignored during analysis. Participants were also instructed not to count when timing the intervals. Instead they were encouraged to time by feel as such instruction has been shown to prevent chronometric counting (Hinton et al., 2004; Hinton & Rao, 2004). All practice and experimental trials were self-paced.

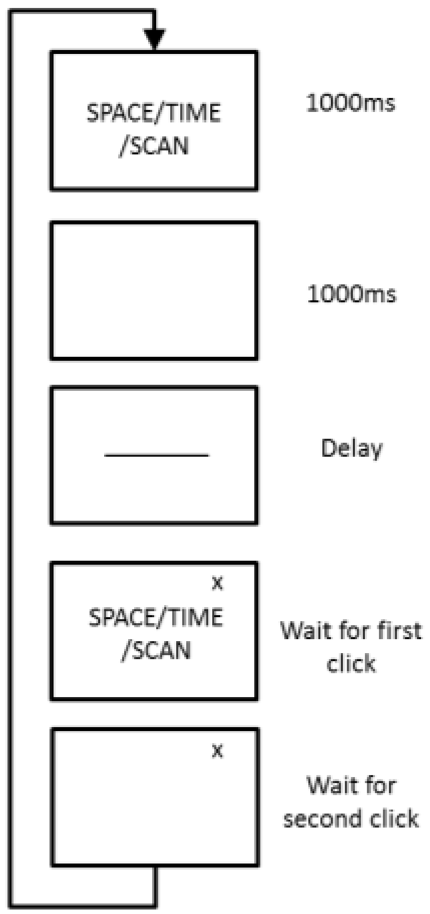

Figure 9.2. General trial procedure for experiments 4-7. Stimuli are for illustrative purposes only and are not drawn to scale.

During the experiment itself all combinations of line length, duration, and task were presented in a randomised order. Response times and mouse click locations during the probe section were recorded. Response times were the variable of interest for the scan and time trials. The horizontal distance between mouse click locations was the variable of interest for the space trials. The task relevant and task irrelevant response dimension were both considered during analysis to ensure that participants were performing the tasks correctly.

Results

Results were considered as a function of presented time or presented line length in each condition. Space, time, and scan are the target conditions. Target conditions are those where the explanatory and stimuli dimension of interest are producing a measure of the same dimension. For simplicity, effects arising from these conditions will be referred to as space, time, and scan effects. For example, the space effect refers to an analysis where the independent variable is the number of pixels in the stimulus and the dependent variable is the number of pixels a participant produces. The kappa, tau, and tau_{scan} conditions are distracter conditions. Here the explanatory dimension opposes the stimuli dimension of interest. For simplicity, effects arising from these conditions will be referred to as kappa, tau, and tau_{scan} effects. For example, the kappa effect refers to an analysis where the independent variable is the number of pixels in the stimulus and the dependent variable is the duration, measured in ms, which a participant produces. Additional clarification is provided in Figure 9.3.

All participants who successfully and accurately completed the task were considered. Accuracy of completion was defined as having a strong (Cohen, 1992) positive effect size (0.5 or greater) for the time, space, and scanning tasks[38]. Similarly if participants were found to be performing the wrong task in any of the conditions they were excluded from analysis.

Standard effects. Standard effects are a shown visually in Figure 9.3. Planned linear contrasts were conducted as part of separate $9_{(times)}x9_{(lines)}$ repeated measures ANOVAs for the time, space, and scanning responses. These contrasts revealed a significant linear main effect in the time ($F_{(1,32)}$=271.68, p<.001), kappa ($F_{(1,32)}$=7.14, p=.012), space ($F_{(1,32)}$=2079.66, p<.001), tau ($F_{(1,32)}$=12.78,

[38] These criteria are used in place of those use in the Casasanto and Boroditsky (2008) study as both sets of criteria lead to the same pattern of results and these criteria lead to the retention of more participants (N=19 vs. N=34). The inclusion of more participants made no difference to whether any of the comparisons except those between scan-speed and the tau effect size came out as statistically significant and gives greater confidence that the data reflect the population.

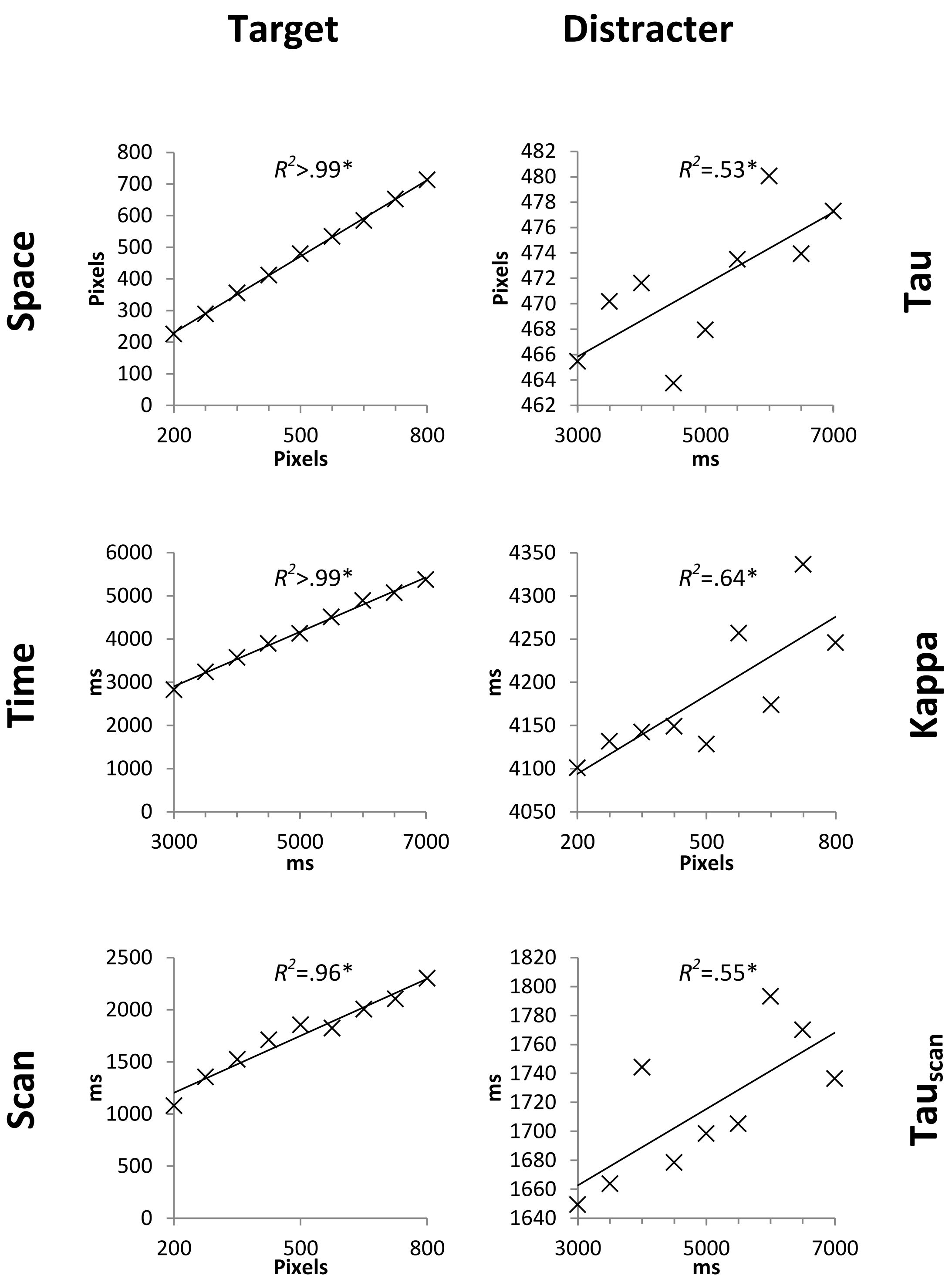

Figure 9.3. The relationship between durations and distances produced and times and line lengths shown for the time, space, and scan tasks. Please note that scales on the y axis vary from graph to graph. Significant R^2 values (p<.05) are followed by an *.

p=.001), scan ($F_{(1,32)}$=45.37, p<.001)[39], and tau$_{scan}$ ($F_{(1,32)}$=7.30, p=.011) effect conditions. These indicated that space, time, and scanning estimates all increased with increases in both space and time. No meaningful interactions were found in any of the ANOVAs conducted on space and scan data providing no evidence to suggest that the influence of time differs for different spaces or vice versa[40]. However a significant linear contrast was found for the interaction between time and space in the time data ($F_{(1,32)=}$7.58, p=.010). This suggests that the impact of space on time increases as durations increase. The kappa effect is stronger with longer times than shorter times.

Standardised effects. To test for a difference in the size of the kappa and tau effects, data from individual participants was standardised by removing their mean space produced for the space judgement conditions and their mean time produced for the time judgement conditions from each observation. These adjusted scores were then divided by the size of a 1 unit change in the dimension participants had been asked to produce. In the case where time estimates were produced (time judgement and the kappa effect conditions) 500ms was 1 unit, in the case of spatial estimates (space judgement and the tau effect conditions) 75 pixels was one unit[41]. Finally the independent variable was re-coded as an ordinal condition number from 1-9 and relabelled as target (time and space judgement conditions) or distracter (kappa and tau effect conditions) rather than the duration or number of pixels presented.

Planned linear contrasts were conducted as part of a $2_{(response)}$ x$9_{(distracter)}$x$9_{(target)}$ repeated measures ANOVA[42]. These revealed a significant main effect of distracter condition number ($F_{(1,32)}$=14.05, p=.001) but no interaction between distracter condition number and response type was found ($F_{(1,32)}$=1.811, p>.05). This suggests that there is insufficient evidence to demonstrate a

[39] Significant quadratic ($F_{(1,33)}$=8.71, p=.006) and cubic ($F_{(1,33)}$=17.38, p<.001) patterns were also found when space was used as the predictor. However these models were more complicated than a linear model and predicted less variation. They were rejected as a result of this and the fact that they were only run for the sake of completeness.

[40] A fourth order interaction between time and space was found for the linear term in space and tau comparison (p=.013) and a fifth order interaction was found in the cubic term (p=.024). These appear to be due to noise in the data as they have no clear theoretical interpretation. A Quadratic interaction was found between time and space in the scan task ($F_{(1,33)}$=4.99, p=.032). This suggested that the tau$_{scan}$ effect may have been larger at short and long intervals than intermediate durations.

[41] These were the changes from one physical stimulus to another. The same pattern of results is observed if each participant's perceived difference between conditions is used instead of the actual difference in the stimuli that have been presented or if each participant's standard deviation in the relevant dimension is used as the scaling factor (the production of z scores).

[42] The same pattern of significant differences is produced by the initial output of a Repeated Measures ANOVA. Linear contrasts are preferred as a linear trend is predicted in all conditions rather than a non-specific difference.

difference in the kappa and tau effect after standardisation[43]. Plots of the scores related to the kappa and tau effects before and after standardisation can be found in Figure 9.4. A significant main effect of target condition number was found ($F_{(1,32)}$=940.99, p<.001) as well as an interaction between response type and target condition number ($F_{(1,32)}$=15.72, p<.001). The interaction suggests

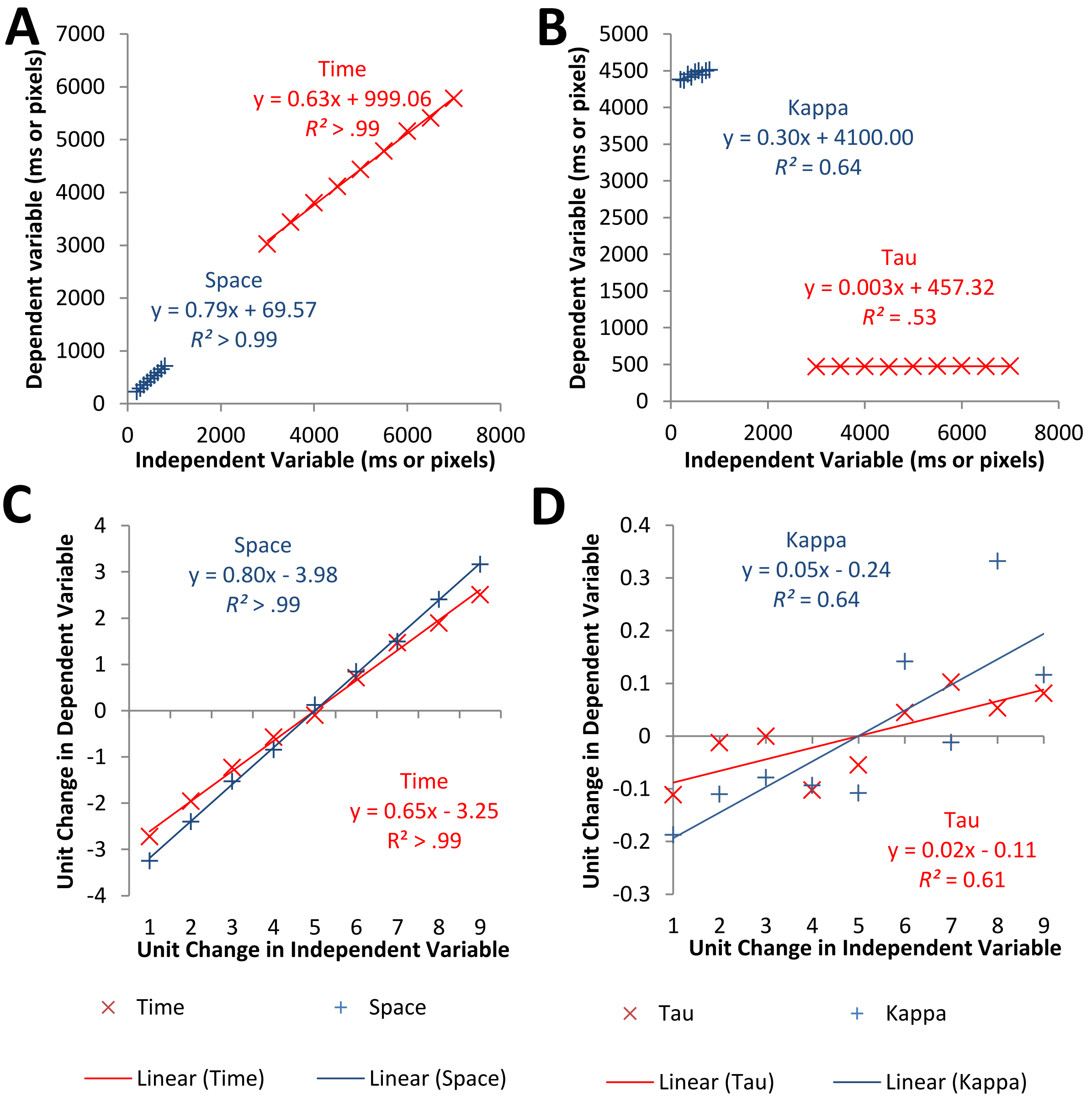

Figure 9.4. Pre-standardised time and space (A) and tau and kappa effects (B) contrasted with standardised time and space (C) and tau and kappa effects (D)

[43] While there is a significant interaction before standardisation the interaction is meaningless as the scales do not share a common unit.

that space and time remain statistically different after standardisation.[44] No other significant interactions were found.

In line with Casasanto and Boroditsky (2008) linear regression analyses were also conducted to assess the strength, stability, and slope of the relationship between the group means for the independent and dependent variable in each condition. Results of these analyses are presented in Table 9.1. These indicated that space, time, and scanning estimates all increased with increases in both space and time. The overall slope for space and time were both below 1 suggesting that, on average, participants underestimated increases in time and space. The overall slope for scanning suggested participants scanned 1 pixel every 1.88ms. The slopes for the kappa, tau, and tau_{scan} effect suggest that there is a slight increase in estimates of the relevant dimension when the irrelevant dimension is varied. Each of these relationships has a strong effect size.

Correlations comparing the size (Cohen, 1992) of the kappa and tau effect across participants were conducted. These revealed a non-significant correlation between the kappa and tau effect sizes (r=.12, p>.05) and between the slopes for the kappa and tau effect (r=-.08, p>.05). While visual inspection of a scatter plot of the slopes and correlations, see Figure 9.5, suggested the presence of outliers. These outliers were not the same participants across the comparisons and were not removed.

Table 9.1
Regression equation components and strengths for all effects

Effect Type	b	intercept	R^2	*r*
space	0.79***	69.57***	>.99***	>.99
time	0.63***	999.06***	>.99 ***	>.99
scan	1.88***	837.86***	.96***	.98
kappa	.30**	4100.00***	.64***	.80
tau	.003*	457.32***	.53**	.75
tau_{scan}	.03*	1583.50***	.55**	.74
space std	.80***	-3.98***	>.99***	>.99
time std	.65***	-3.25***	>.99***	>.99
kappa std	.05*	-0.24*	.64***	.80
tau std	.02*	-0.11*	0.61**	.74

Note: df=7 in all cases, * p<.05,**p<.01, ***p<.001.

[44] In a two-tailed test this difference becomes non-significant if perceived time and spatial units are used in standardisation rather than displayed time and spatial units. However it is significant in a one tailed test and is what would be predicted under Vierodt's law (Lejeune & Wearden, 2009). Put more simply the psychophysical constant for space and time are close to 1 (Stevens, 1957) but the scaling factor for space is larger and closer to 1 than that of duration.

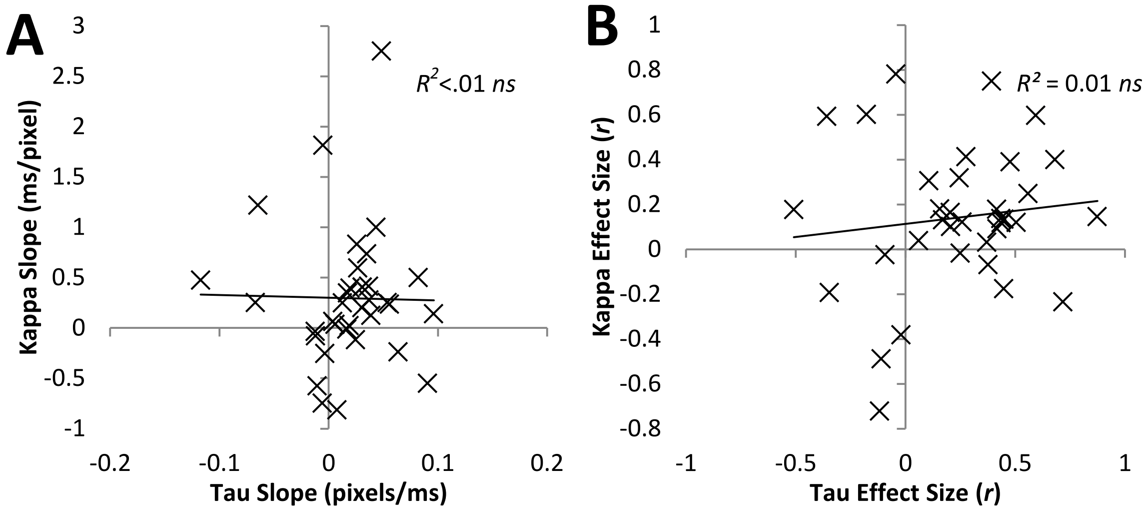

Figure 9.5. The relationship between the kappa and tau effect size. Each data point shows an individual participant. Panel A compares the slope of the effects, panel B compares the size of the effects. Non-significant R^2 values (p>.05) are followed by *ns*.

Scan-speed. To address the question of whether or not the size of kappa and tau effects are related to scan-speed, scan-speeds[45] for individual participants were compared to the correlations obtained for the kappa and tau effects for individual participants. In order to make speeds comparable between screen sizes, speeds from participants using the 15 inch screens were scaled to match the larger screen size[46]. A borderline significant[47] moderate positive correlation, shown in Figure 9.6, was found between scan-speed and the tau effect (r=.32, p=.057) giving weak evidence that faster scan-speeds may be associated with a stronger tau effect. No such relationship was found between scan-speed and the kappa effect (r=.12).

The relationship between the tau effect and scan-speed was examined further after visual inspection of the scatter plot shown in Figure 9.7. On a visual and theoretical basis (Collyer, 1977) it was appropriate to split the participants into two groups – those showing a positive correlation (a positive tau effect) and those showing a negative correlation (a negative tau effect). When

[45] Speed, being a measurement of time per unit space, was calculated by finding the slope of the scanning function for individual participants. This differs to a standard measure of speed, pixels per millisecond in this case. While some may refer to time per unit space as inverse speed, it will be referred to here as speed for the sake of simplicity. This metric was preferred over the standard measurement of speed as it is a more honest interpretation of what is happening in the scan task (changes in space are producing changes in times produced rather than the other way around).

[46] The horizontal distance of the small screen was 87.5% of the large screen. Speeds for small screen participants were divided by 0.875 to account for this. Such a correction is compatible with the findings of the combined analyses presented later.

[47] Had the size of the tau effect been expected to increase with increasing speed this would come out as significant in a one tailed test.

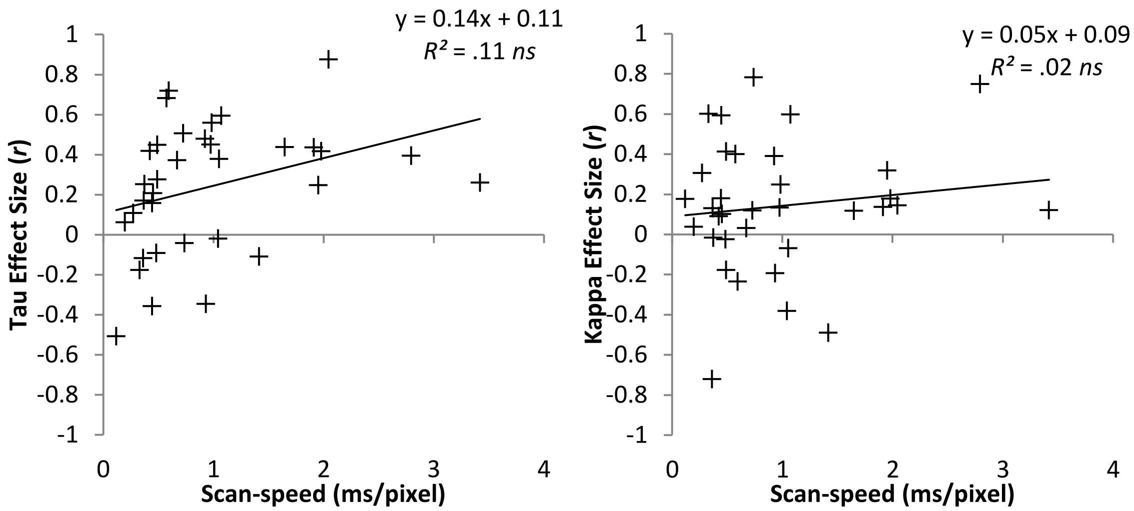

Figure 9.6. The linear relationship between scan-speed and tau effect and kappa effect size. Non-significant R^2 values (p>.05) are followed by *ns*.

considered separately the plot of the relationship in the participants with a positive tau effect appeared to have a quadratic trend. Fitting a quadratic trend led to a significant increase in R^2 compared to a linear trend in this group. There was no significant linear or quadratic relationship found in the participants with negative correlation values and the pattern in these data does not appear consistent across speed, average scanning times, and scanning slope. These changes in R^2 are summarised in Table 9.2.

Table 9.2
R^2 for all groupings when comparing scan-speed and the tau effect

Participants	Trend	R^2	ΔR^2
All	Linear	.10	
	Quadratic	.19*	.09
Positive	Linear	.04	
	Quadratic	.30*	.26*
Negative	Linear	.21	
	Quadratic	.34	.13
Absolute Tau	Linear	.04	
	Quadratic	.12	.08
Shifted Tau	Linear	.07	
	Quadratic	.32**	.25**

Note: $N_{(All)}$=34, $N_{(Positive)}$=25, $N_{(Negative)}$=9, * p<.05,**p<.01.

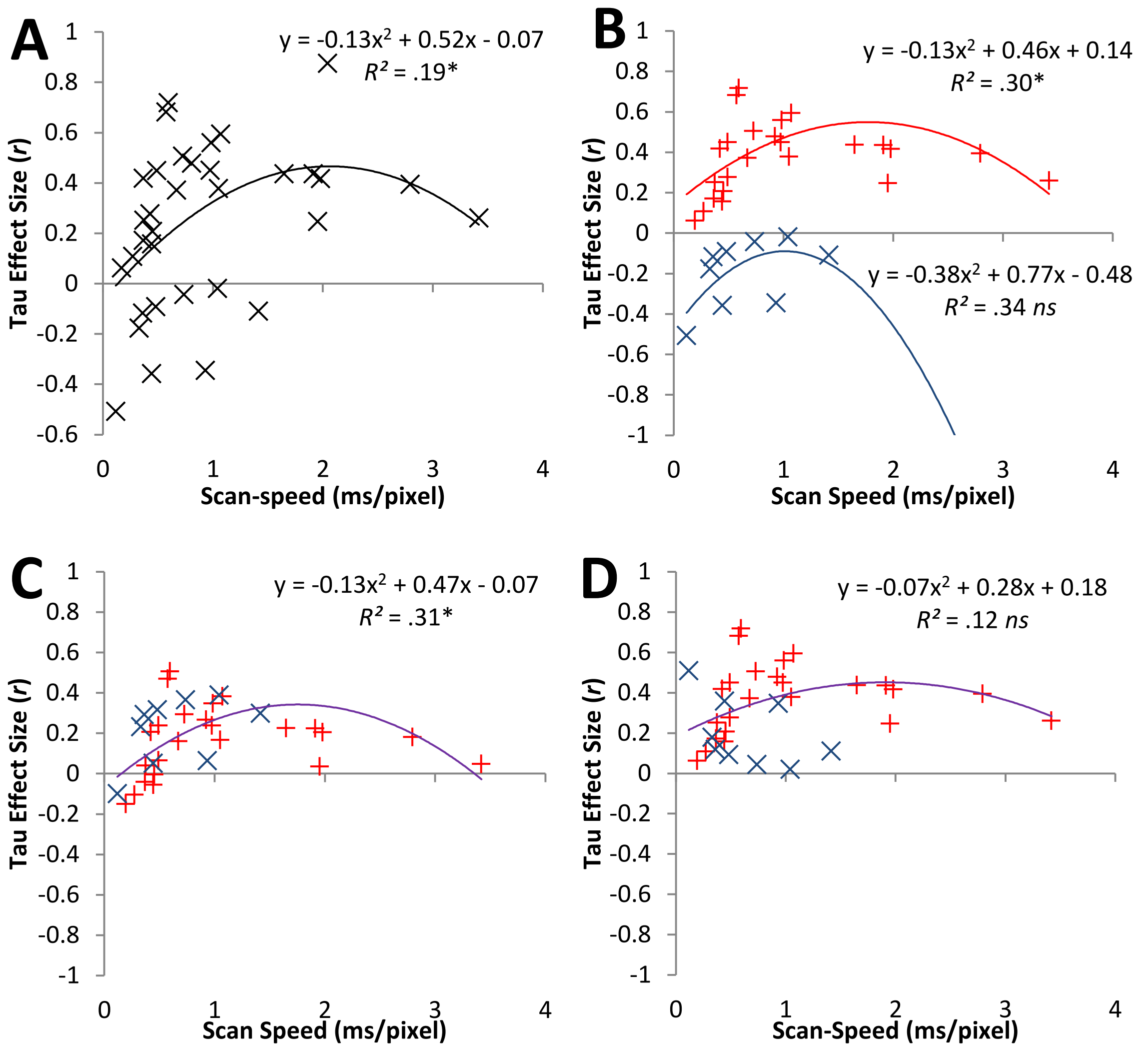

Figure 9.7. The quadratic relationship between the size of the tau effect and scan-speed. Individual data points represent individual participants. Figure A shows the data considered as a whole. Figure B shows the data considered as two separate groups: participants with a positive tau effect and those with a negative tau effect. Figure C shows the negative participants with a constant effect size of -.6 removed. Figure D shows the absolute correlation values for all participants. In panels where positive or negative tau effect identity is considered, participant's with positive tau effects are shown as red +s and those with negative tau effects as blue Xs. The trend line for the positive tau effect is red, the negative tau effect blue, and the combined trend line is purple. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

Additional analyses were conducted to test explanations of the difference between the negative individuals and the positive individuals. In both cases negative and positive participants were considered together. In one case the absolute value of the correlations was considered. This tests the idea that participants in the negative group may be reversing the tau effect. This produced no significant linear or quadratic trend. In the second case the equations for the quadratic lines of best fit were found for the negative and positive participants. The constant for each group was then removed from each member of that group. For the positive participants 0.14 was removed. For the negative participants -0.48 was removed. This is equivalent to removing a strong effect (Cohen, 1992) of -0.62 from all negative participants. When these adjusted values were analysed as a single group a significant strong quadratic relationship, similar to that produced when the positive participants are considered in isolation, was found. The results of these additional post-hoc analyses are shown in Table 9.2 and Figure 9.7.

Discussion

As expected, on the basis of most previous studies, kappa and tau effects were both found in the present experiment. The time, space, and scan effects also produced strong relationships with their target dimensions (time, space, and space respectively). Furthermore, like space, the scan task also produced a tau-like effect, referred to as the tau_{scan} effect. While there was no evidence for a relationship between scan-speed and the kappa effect there was evidence for a relationship between scan-speed and the tau effect. While the effect size was small when all participants were considered, the strength of the relationship between the size of the tau effect and scan-speed improved when participants showing a positive tau effect were separated from participants showing a negative tau effect. Making a distinction between participants with positive and negative tau effects is in line with suggestions from earlier research (Collyer, 1977).

That kappa *and* tau effects were found and their slopes did not interact after standardisation argues against, but does not disprove, the idea that there is a clear directional relationship between the concepts of time and space in the human mind. When these results are compared with those of the Casasanto and Boroditsky (2008) study they suggest that the size of the tau effect, and possibly the kappa effect, may be dependent on the magnitude of the stimuli in the irrelevant dimension and/or that it may be important to beware of crossing boundaries in the timing system when looking for kappa and tau effects. While the Casasanto and Boroditsky (2008) study may have straddled sub and suprasecond times the current study is more comfortably in the suprasecond time bracket.

An alternative interpretation would be to consider that the scanning task may have reintroduced imputed velocity. If Casasanto and Boroditsky (2008) failed to find a tau effect because their stimuli suppressed imputed velocity the hypothesised re-introduction of imputed velocity through the scan task would explain the discrepancy in the results. As the scanning task is intended to index imputed velocity and requires subjects to imagine/impute velocity this should be explored.

Scan. The presence of a relationship between scan-speed and the tau effect but not the kappa effect is understandable if initially surprising. The scanning task is essentially a replication of the line length judgement task meaning that scanning is aligned most closely with line length judgements. As a result, scan times, and therefore scan-speeds, are associated with judgements of line length rather than judgements of time. As the tau effect measures the impact of presentation duration on judgements of line length, it makes sense that an effect of time on space would be seen here and not the other way around. Essentially, as it is possible to find a tau effect in the scanning paradigm but it is not possible to find a kappa effect, it makes sense to expect a relationship between scanning and the tau effect but not scanning and the kappa effect.

An initial analysis of all participants suggested the presence of a significant quadratic relationship between scan-speed and the size of the tau effect. Such a trend suggests that there is an optimum speed where the impact of time on space is at its strongest before declining again; essentially participants who scan too quickly or too slowly do not produce tau effects that are as reliable as those who scan at the optimum speed. Further examination of the relationship between scan-speed and the tau effect suggested the presence of two groups that have been alluded to in past research (Collyer, 1977). The first group produced a tau effect in the expected direction: a positive tau effect where estimates of space increased with presentation duration (more time = more space). The second group showed the opposite pattern: a negative tau effect (more time = less space).

When participants showing positive tau effects are considered in isolation a strong quadratic pattern is found. A line of best fit suggests that the peak of the quadratic pattern is at about 2 milliseconds per pixel with the strength of the tau effect rolling off on either side. This is in line with the slope of the overall scanning results and what was expected at the beginning of the experiment. Specifically, the strength of the tau effect appears to be at a maximum when the speed at which participants expect items to travel matches the optimum scan-speed for the stimulus set. The message from the negative group is less clear and will be considered as more participants enter the group across experiments. The current best interpretation is that the underlying pattern is the same in both groups. This suggests the presence of a common process that may be followed by some form of adjustment or correction.

The proposed quadratic relationship is in line with the idea of a resonant scan-speed: a level of imputed velocity that is compatible with the lines and times presented. While it has been noted that the optimum speed is one that maximises the tau effect size produced the resonant speed could occur at the peak of the quadratic term or at the tails. Expecting the resonant scan-speed to occur at the peak is equivalent to saying that greater resonance leads to a larger tau effect. Expecting the resonant scan-speed to occur at the tails suggests the opposite: resonance is associated with a smaller tau effect.

In order to establish where resonance occurs it is necessary to consider the argument put forward by equation 8.1 on page 100. The equation is re-presented below for clarity.

$$V = s_1/t_1 = s_2/t_2$$

Here the expected imputed velocity is given by s_1 and t_1. If the imputed velocity produced by s_2 and t_2 matches that produced by s_1 and t_1 no tau effect is predicted. Under this scenario s_2 and Vt_2 both provide the same information. As time and space provide the same answer through imputed velocity *matching* across conditions this is considered the resonant scenario. In an alternate scenario, where the speed associated with s_1 and t_1 conflicts with that of s_2 and t_2, a tau effect would be produced. In this case s_2 and Vt_2 provide *conflicting* information. Large tau effects are the product of interference (a lack of resonance) whereas small tau effects are the product of strong resonance.

If the tails of the quadratic trend indicate the points of highest resonance, this suggests that the cost of failing to achieve resonance is initially very high. The relative cost then decreases as speed moves further from the resonant point and closer to the optimum speed.

While resonance producing a reduction in the tau effect may appear counter-intuitive at first it is in line with previous work and ideas around the relationship between the tau effect and imputed velocity. Specifically, when a match between time, space, and speed was produced tau and kappa effects were not found (Sarrazin et al., 2004). Furthermore, probabilistic models of these effects would suggest that an association between space and time should be less apparent when the associations match (Goldreich, 2007).

Such a finding and interpretation is in line with the idea of a common representation for space and time. Under such an account resonance between space and time produces a situation where temporal and spatial information provide a common answer (Goldreich, 2007; Walsh, 2003). Increasingly non-resonant combinations produce a scenario where temporal and spatial information are increasingly out of step. This leads to a greater contribution of time information in spatial judgements, more typically referred to as tau effects.

Best practice. While this experiment has provided insight through broad use of analysis techniques it has also highlighted that some techniques are more beneficial than others. In particular, comparisons of non-standardised measures of space and time have been shown to be unnecessary and will be excluded from analysis from this point onwards. While standardised measures are useful they will be reserved for a broad meta-analysis at the end of the prospective timing section.

In individual experiments the primary focus will be the relationship between scan-speed and the tau effect. Particular attention will be paid to the separation between participants showing a tau effect with a positive correlation and those showing a negative correlation. Henceforth these shall be referred to as positive tau effects and negative tau effects. Particular attention will also be given to the role of the scan task in the production of tau effects before broader investigations are considered.

A secondary focus will concern the presence or absence of standard effects: space, time, scan, kappa, and tau effects. Space, time, and scan effects give an indication of the validity of the responses that the participants are providing whereas the presence or absence of kappa and tau effects may provide information on the conditions required for their elicitation. Ideally an indication of how they can vary from experiment to experiment will be provided.

Further comments on associations with the kappa effect will be reserved until the final meta- analysis. If the apparent absence of a relationship between the kappa effect and scan-speed or the tau effect is due to high variability in the data, running an analysis at the end of the chapter including all participants should reduce this issue.

Conclusion

The current design produced an apparently symmetrical relationship between space and time with space and time having an equal impact on each other when investigated using standardised units and effect size (*r* value) measures. Scan-speed was related to the size of the tau effect but not the kappa effect and this relationship was strongest when considering participants who demonstrated a tau effect in the expected (positive) direction. This relationship was quadratic with a maximum at a scan slope of around 2ms per pixel. This suggested that participants who scanned at an intermediate speed experienced a larger tau effect than those who scanned faster or slower. Each aspect of the findings requires further examination. The most pressing being the suggestion that the introduction of the scan task may be responsible for one or more of the findings.

Chapter 10. Experiment 5

In order to test whether the difference between results of the previous experiment and those of Casasanto and Boroditsky (2008) were due to the introduction of the scan task the experiment was redesigned with the scan task conducted separately at the end of the experiment. The selection criteria and standard sample size of Casasanto and Boroditsky (2008) were also considered in order to replicate their study as closely as possible and to test whether difference in the results produced may be due to these factors. The results will be compared to and combined with those of experiment 4 in order to test the reliability of the findings and explanations of experiment 4 as well as test for an effect of moving the scan task.

Methods

Participants. Volunteers from the University of Auckland student population were recruited until 19 participants, sample size that matches the original Casasanto and Boroditsky (2008) static line experiment sample size, met the inclusion criteria of Casasanto and Boroditsky (2008). In total 31 new volunteers from the University of Auckland student population took part in the experiment. All participants reported normal or corrected to normal vision. While 19 met the requirements of Casasanto and Boroditsky (2008) original study[48], 22 met the current criteria. There was no difference in the pattern of results observed across criteria sets. As such the results reflect those of the larger sample of 22 participants (N_{males}=6, M_{age}=24.33, $\text{Range}_{\text{age}}$=18-30, All Right Handed; N_{females}=16, M_{age}=22.33 $\text{Range}_{\text{age}}$=18-30, 15 Right Handed, 1 Left Handed).

Materials. The materials were identical to those used in Experiment 4.

Procedure. The procedure was the same as that of Experiment 4 with the following exceptions. When explaining the task to participants the experimenter outlined the space and time tasks and stated that a third task would occur at the end of the experiment. The explanation screen, practice block, and all trials associated with the scanning task were moved to the end of the experiment such that participants completed the time and space estimation trials before receiving any information about, or practice with, the scan trials. Upon reaching the scan practice block participants were free to practice the scan trials as many times as they wished and were encouraged to ask the experimenter for clarification on the task requirements if necessary.

Results

Independent analyses. As in experiment 4 the results were considered as a function of presented time or presented line length in each condition. Target and distracter conditions are as

[48] Participants must have English as a first language, respond to the relevant stimulus dimension in each task, and produce a slope of more than 0.5 in the space and time tasks.

defined in experiment 4. All participants who successfully and accurately completed the task were considered. Accuracy of completion was defined as having a strong (Cohen, 1992) positive effect size for the time, space, and scanning tasks[49]. Similarly if participants were found to be performing the wrong task in any of the conditions they were excluded from analysis.

Standard effects. Planned linear contrasts were conducted as part of separate two way repeated measures ANOVAs for the time, space, and scanning responses. These contrasts revealed significant linear main effects in the time ($F_{(1,20)}$=672.59, p<.001), space ($F_{(1,20)}$=4114.71, p<.001), scan ($F_{(1,20)}$=22.67, p<.001), and tau_{scan} ($F_{(1,20)}$=6.15, p=.022) effect conditions. These indicated that space, time, and scanning estimates all increased with increases in their target dimension. Scanning estimates also increased with their non-target dimension (time). No significant tau ($F_{(1,20)}$=0.328, p>.05) or kappa ($F_{(1,20)}$=0.74, p>.05) effect was found. No interactions reached significance. Plots of the standard effects are shown in Figure 10.1.

Scan-speed. As for experiment 4, speeds from participants using the 15 inch screens were scaled to match the larger screen size[50]. There were insufficient data points available to demonstrate whether the quadratic relationship between scan-speed and the tau effect was retained in the present experiment in isolation. In the combined condition, determined by processes outlined in the combined analyses, a quadratic relationship trending in the expected direction accounted for 10% of the variance but did not reach significance ($F_{(2,20)}$=1.11, p>.05) while the positive and negative conditions each produced non-significant relationships in the wrong direction. Plots of these points can be seen on the left of Figure 10.3.

Combined analyses.

Standard effects. Standard effects are shown visually in Figure 10.2. Planned contrasts conducted as part of repeated measures ANOVAs reported a main effect of time ($F_{(1,54)}$=660.64, p<.001), space, ($F_{(1,54)}$=4528.93, p<.001), scan ($F_{(1,54)}$=72.09, p<.001), kappa ($F_{(1,54)}$=5.67, p=.021), tau ($F_{(1,54)}$=5.28, p=.025), and tau_{scan} ($F_{(1,54)}$=13.68, p=.001) . Scan task order produced an interaction with the tau ($F_{(1,54)}$=8.42, p=.005) effect, reflecting the fact that the tau effect was stronger in experiment 4 than experiment 5. No other interactions were significant.

[49] These criteria are used in place of those use in the Casasanto and Boroditsky (2008) study as both sets of criteria led to the same pattern of results and these criteria led to the retention of more participants (N=19 vs. N=22). The inclusion of more participants made no difference to whether any of the comparisons undertaken came out as statistically significant and gives greater confidence that the data reflect the population.

[50] The horizontal distance of the small screen was 87.5% of the large screen. Speeds for small screen participants were divided by 0.875 to account for this. Such a correction is compatible with the findings of the combined analyses presented later.

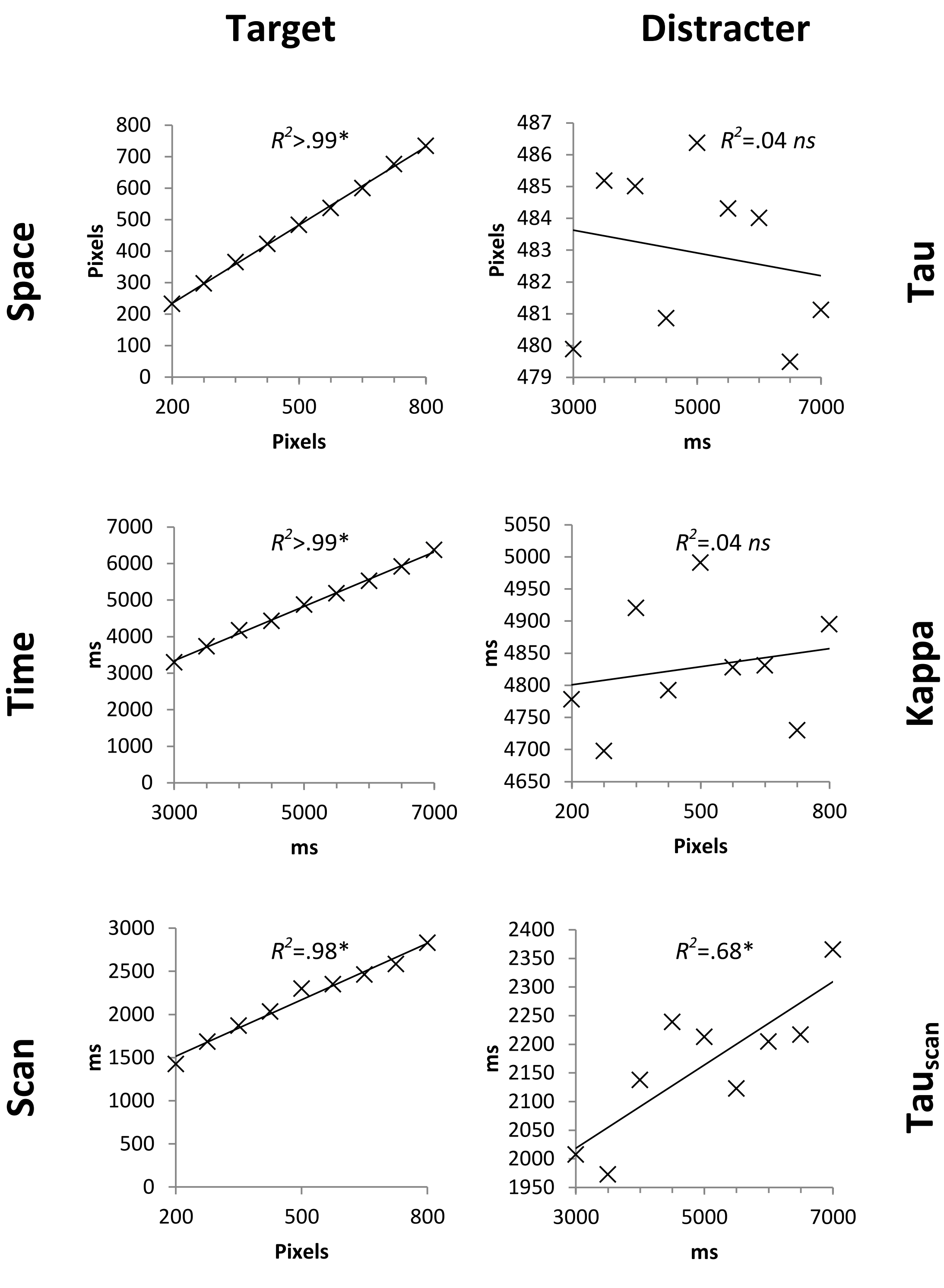

Figure 10.1. The relationship between times and spaces produced and times and spaces shown for the time, space, and scan tasks for experiment 5 only. Please note that scales on the y axis vary from graph to graph. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

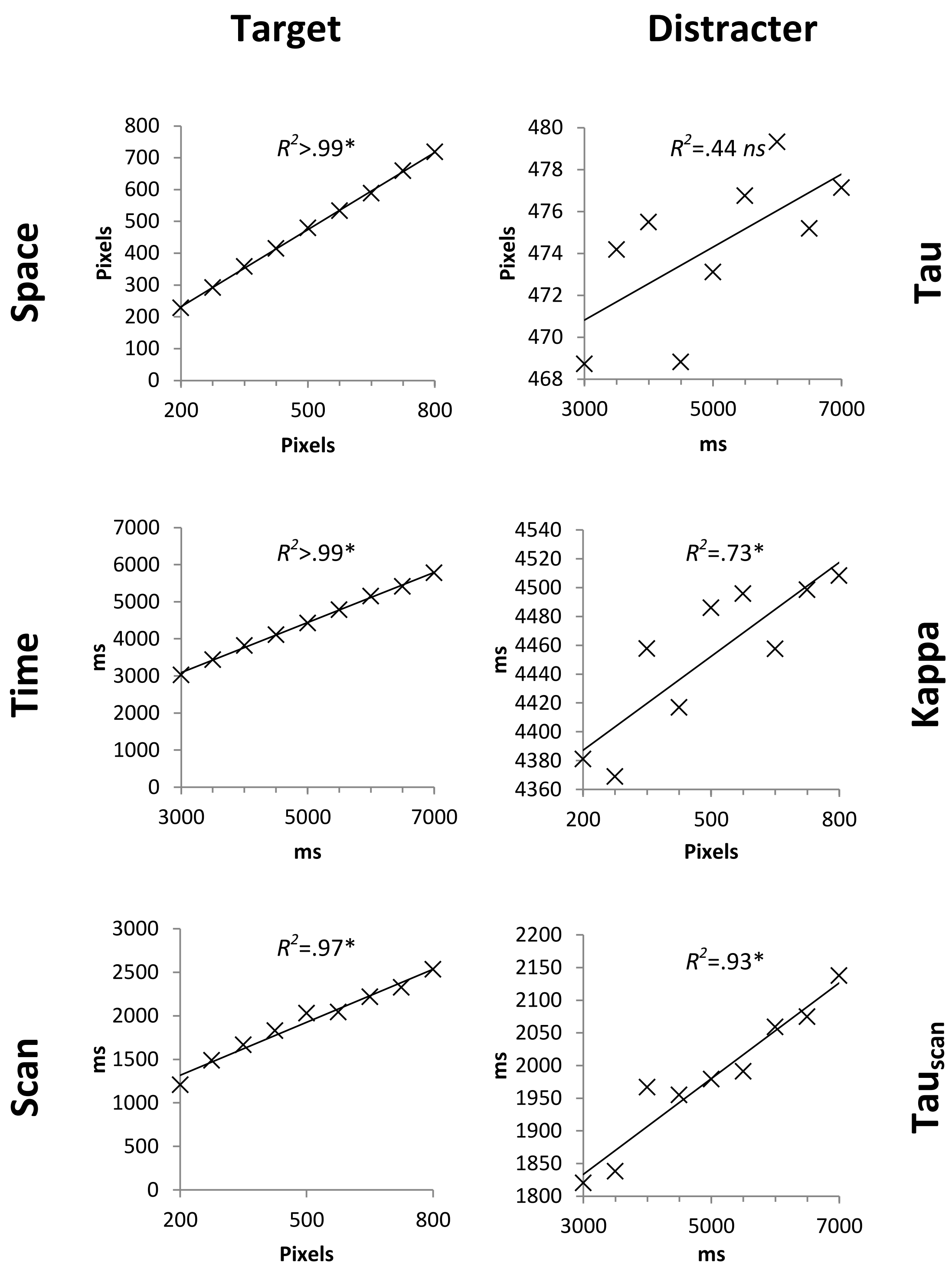

Figure 10.2. The relationship between times and spaces produced and times and spaces shown for the time, space, and scan tasks for experiment 4 and experiment 5 combined. Please note that scales on the y axis vary from graph to graph. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

Scan-speed. The combined data sets appear to show two separate quadratic trends in the relationship between scan-speed and tau effect size. When fitted separately they give the equations shown in Figure 10.3. As the trends appear similar, with one shifted down compared to the other,

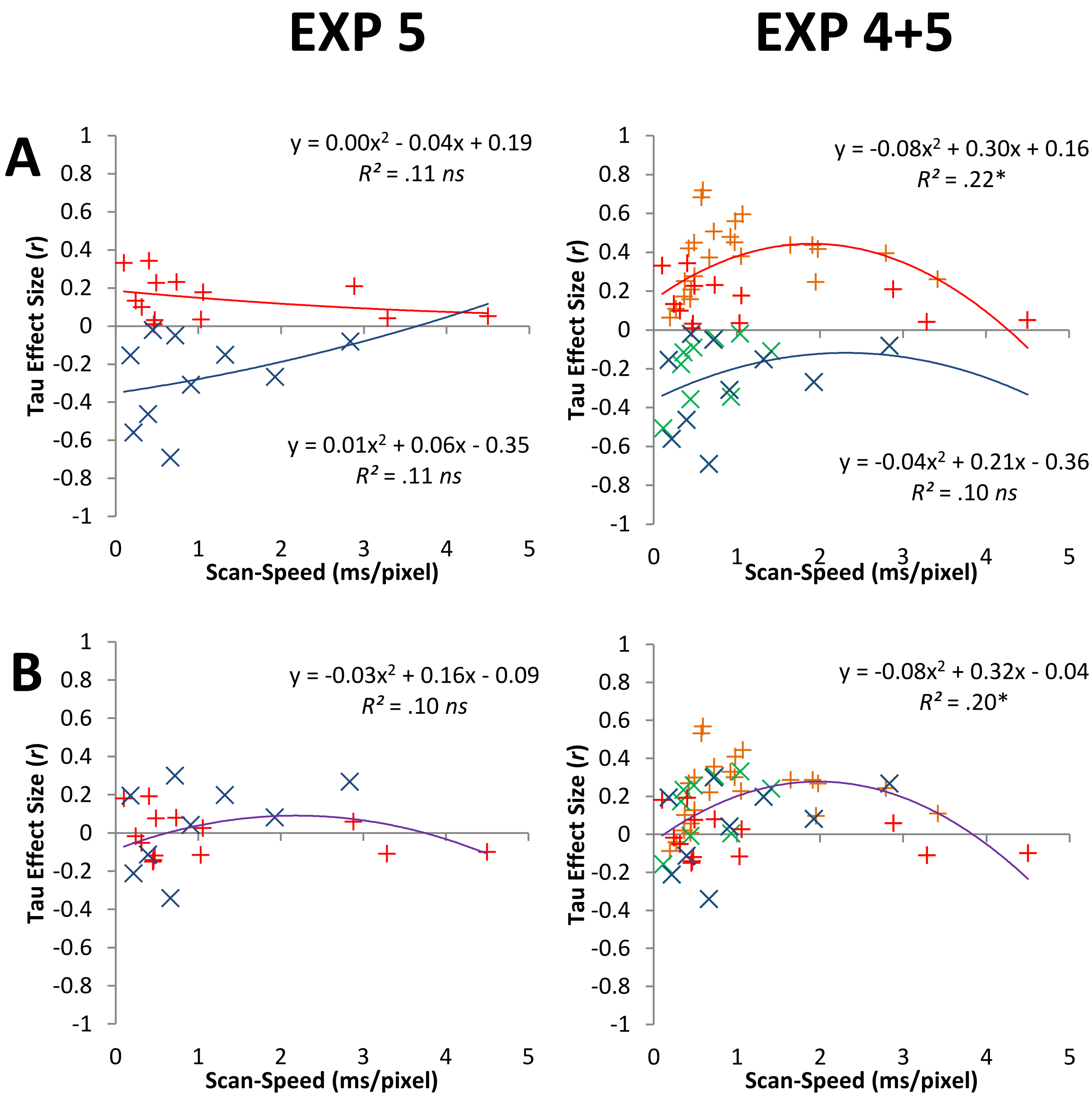

Figure 10.3. The relationship between the tau effect and scan-speed. Data from experiment 5 is shown alone on the left and overlaid with that of experiment 4 on the right. The figures in panel A show separate functions for the negative and positive groups. The figures in panel B show the positive and negative groups combined through the removal of a constants from the combined data of experiments 4 and 5 shown on the right of panel A. Participants showing positive tau effects are represented as pluses. Data from experiment 5 is in red, data from experiment 4 is shown in orange. Participants showing negative tau effects are represented as crosses. Participants from experiment 5 are shown in blue, participants from experiment 4 are shown in green. The trend line for the positive tau effect is red, the negative tau effect blue, and the combined trend line is purple. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

the data was combined by removing .16 (the constant of the associated model) from the tau effect scores of participants in the positive group and adding .36 (the constant of the associated model) to the effect size of the participants in the negative tau group. The trends were significant when participants with positive tau effects were considered in isolation ($F_{(2,35)}$=5.03, p=.012) and when all participants were included in the combined analysis ($F_{(2,53)}$=6.50, p=.003). The trend in participants producing negative tau effects did not reach significance ($F_{(2,15)}$=0.84, p>.05).

Discussion

Unlike experiment 4, no significant kappa or tau effect was found in experiment 5 when it was considered in isolation. When data from experiment 5 was combined with data from experiment 4 a significant kappa but not tau effect was found. In the combined case, the *r* values of the kappa and tau effects were in line with those found in the Casasanto and Boroditsky (2008) study.

The failure to find a kappa or tau effect within the present design may be seen as supporting the idea that a design of this type, one using static lines, can be used to suppress imputed velocity. This would also suggest that introducing the scan task during the space and time trials in the experiment, as was the case in experiment 4 but not experiment 5, reintroduces imputed velocity.

While such an argument strengthens the case for scanning being a useful index of imputed velocity it is also problematic. First, the Casasanto and Boroditsky (2008) experiments reliably produced kappa effects without the presence of a scan task and without creating a tau effect. As mentioned in experiment 4, if imputed velocity is actually suppressed by static lines there is no reason to expect to see a tau effect or a kappa effect. The production of a kappa effect in Casasanto and Boroditsky (2008) is problematic then as it does not explain why a kappa effect would not be seen in the current experiment and is incompatible with an imputed velocity account. Second, when the results of experiment 4 are combined with the results of experiment 5 the overall kappa effect becomes stronger whereas the overall tau effect becomes non-significant. This pattern of results is in line with Casasanto and Boroditsky (2008) and essentially suggests that the failure to find this pattern of results individually in experiments 4 and 5 is due to an insufficient sample size. Unfortunately an explanation based on sample size is somewhat problematic. While experiment 4 did have fewer participants than the Casasanto and Boroditsky (2008) study had overall, it had more participants than any of their individual experiments. Furthermore, experiment 5 had as many or more participants than any single experiment in the Casasanto and Boroditsky (2008) study.

Scan-speed. While experiment 5 alone shows no significant relationship between scan-speed and the tau effect, the data does add to that produced in experiment 4. Where the scatter plot in experiment 4 suggested the presence of a strong quadratic relationship for positive participants and no clear relationship for negative participants, plots of data combined across experiments 4 and 5 appears to show a similar pattern may be present in both groups.

The interpretation for the positive group remains the same as it was for experiment 4. There appears to be an optimal scan-speed and participants scanning at this speed have larger tau effects than those scanning faster or slower. However the interpretation for the negative group is strikingly counterintuitive. Despite showing the same shape as the positive group the meaning of the relationship is reversed. Those who scan quickly and slowly seem to have strong negative tau effects. Those scanning at an intermediate speed have weaker negative tau effects.

The simplest explanation is one that fits the effect size adjustment suggested in experiment 4. As the definition of a tau effect fits the positive group but contradicts the results of the negative group the difference between the positive group and the negative group could be viewed as the negative group applying some sort of correction. As the distance between the positive group's line of best fit and the negative group's line of best fit appears to be relatively constant this suggests that the same correction is applied by all members of the negative group regardless of the speed that they scan at. As the difference is around .52 this suggests that the effect of this correction is quite strong (Cohen, 1992). Furthermore, as the correction is applied regardless of scan-speed, the suggestion is that the correction is applied regardless of the size of the tau effect that is actually experienced. Essentially the positive group and the negative group may perceive the stimuli in the same way but the negative group is mistrusting and correcting their perceptions. However, as the number of participants in the negative group is still quite small this interpretation must be treated with caution.

Difference across experiments. The combination of the failure to find an overall tau or kappa effect in experiment 5 and the finding that participants from experiment 5 continued to show the relationship between tau effect size and scanning speed in experiment 4 raises some interesting questions. The obvious question is why a difference in the overall kappa and tau effect was observed. If examination is restricted to the current line of experiments the implication would be that the movement of the scan task to a separate block led to this change. In turn this could imply that the perception of imputed velocity, argued to produce kappa and tau effects, depends on the presence of a scan task when static lines are used. However such an explanation cannot explain the results observed by Casasanto and Boroditsky (2008) as their study found a kappa effect but not a tau effect.

An alternative explanation would be that it is the knowledge of how time and space relate to each other that influences the likelihood of finding a kappa or tau effect. In such a scenario experiments would be very susceptible to experimenter expectancy effects (Intonspeterson, 1983;

Intonspeterson & White, 1981) and/or the implication of a relationship between time and space given before or during the experiment. Experimenter expectancy effects essentially describe the phenomenon where participants produce the results that the experimenters are expecting to see by picking up on subtle cues that the experimenter is unaware of providing (Intonspeterson, 1983; Intonspeterson & White, 1981). These can be as subtle as differences in the way a scripted set of instructions is read to participants (Intonspeterson, 1983). Given that these tasks assess mental representations of time and space (mental imagery), have been shown to be susceptible to how people think things interact with their environment (Masuda et al., 2011), depend on the approaches individuals take to the task at hand (Collyer, 1977), and mental imagery findings have been argued to be the result of people behaving in the way that they feel they should behave (Intonspeterson, 1983; Intonspeterson & White, 1981; Pylyshyn, 2003) this sort of explanation should be considered.

Assuming that participants picked up on the expectations that Casasanto and Boroditsky (2008) held before running, experimenter expectancy effects (Intonspeterson, 1983; Intonspeterson & White, 1981) provide a good candidate for the pattern of results observed across the experiments so far. Namely, Casasanto and Boroditsky (2008) expected to find a stronger kappa effect than a tau effect and did. In experiment 4 kappa and tau effects of equal size were expected and found. In experiment 5, imputed velocity was expected to come from the scan task and its removal was expected to remove at least the tau effect but theoretically the kappa effect as well. Once again the results produced were in line with experimenter expectations.

Chapter 11. Experiment 6

In order to determine if the differences observed between experiments 4 and 5 were the result of type 1 error in experiment 4, low sample size/type 2 error in experiment 5, different experimenter expectations, or moving the scan task to the end of the experiment, a variation was conducted. In this version of the experiment participants would be randomly allocated to one of 3 conditions without the experimenter's knowledge. The first condition was the same as experiment 4 (scan task during), the second condition was the same as experiment 5 (scan task post), the third condition was the same as experiment 5 except the scan task came before the other two tasks (scan task pre). This combination should demonstrate whether the differences between experiment 4 and experiment 5 can be replicated and/or were due to experimenter expectancy effects, the running of all 3 tasks concurrently, or the assumed encouragement of imputed velocity.

If the differences are type 1 error and/or due to experimenter expectancy effects, no differences should be observed between conditions and no relationship between the scan task and the tau effect should be observed. If the differences were due to the running of 3 tasks at once, thereby increasing demand on participants, only the scan task during condition should show both kappa and tau effects. If it is due to the scan task imparting knowledge of how space and time might relate to each other, the scan task pre and during conditions should show both effects whereas the post condition, the only condition where the scan task is introduced after the space and time tasks, should not show a tau effect. If there is no difference between the tau effect in any of the conditions and it reaches significance this would suggest that the failure to find a tau effect in experiment 5 was due to a low sample size/type 2 error.

While experimenter expectancy effects have been shown to be very pervasive (Intonspeterson, 1983; Intonspeterson & White, 1981) the current design should be able to eliminate them. Unlike previous experiments, in this experiment the experimenter cannot have any clear expectations. First, the experimenter will not know which condition a given participant will be running in during the experiment, thereby preventing the experimenter from relaying any expectations that are condition specific. Additionally, even if such expectations could be relayed, the expectation of the pattern that a certain condition should produce depends on the expected cause of these differences. As all of the explanations outlined in the previous paragraph are assumed to be equally likely, no clear or consistent expectation exists.

Methods

Participants. In total 60 volunteer undergraduate and postgraduate students from the University of Auckland with normal or corrected to normal vision participated in this study. Reimbursement, in the form of a $10 petrol voucher, was accepted by 44 participants. The final number of participants considered here is 47 (N_{males}=15, M_{age}=24.93, $Range_{age}$=18-33, 14 Right Handed, 1 Ambidextrous; $N_{females}$=32, M_{age}=24.56, $Range_{age}$ =18-39, 29 Right Handed, 1 Ambidextrous, 2 Left Handed) as the data of 13 participants suggested that they had completed the task incorrectly.

Materials. The same materials were used as in experiments 4 and 5 except the screen size was standardised at 15 inches.

Procedure. After arriving in the lab and consenting to participate, participants were randomly allocated a participant number. The computer used this participant number to counterbalance the participants across 3 conditions. Participants in the 'during' condition performed the experiment in the same way as those in experiment 4. Those in 'post' condition completed the experiment in the same way as those in experiment 5. Those in the 'pre' condition completed all of the scan trials first then completed the space and time trials (similar to experiment 5 with the order reversed).

After completing the trials participants were presented with a description of standard kappa and tau effects. They were asked a) whether this description surprised them and b) whether they were aware of making any kind of correction for kappa and/or tau effects. These responses were recorded for later analysis.

Results

As in experiment 4 any participants who completed the task correctly were included in the analysis. In particular participants showing a negative correlation in the scan, space, or time condition and/or any participants showing a non-significant correlation between space presented and space reported or time presented and time reported were removed. The effect size criterion of 0.5 was maintained in all conditions, 13 of the 60 participants recruited failed to meet these inclusion criteria. A chi-square analysis showed that there was no significant contribution of experimental condition to dropout rate.

Standard effects. There was no effect of or interaction with scan task inclusion order in any of the analyses. As such the data presented here collapses across these conditions[51]. Data from these analyses are shown in Figure 11.1.

Planned linear contrasts were conducted as part of separate $9_{(times)}x9_{(lines)}$ repeated measures ANOVAs for the time, space, and scanning responses. These contrasts revealed significant linear

[51] Appendix B contains plots similar to Figure 11.1 separated by scan task inclusion order.

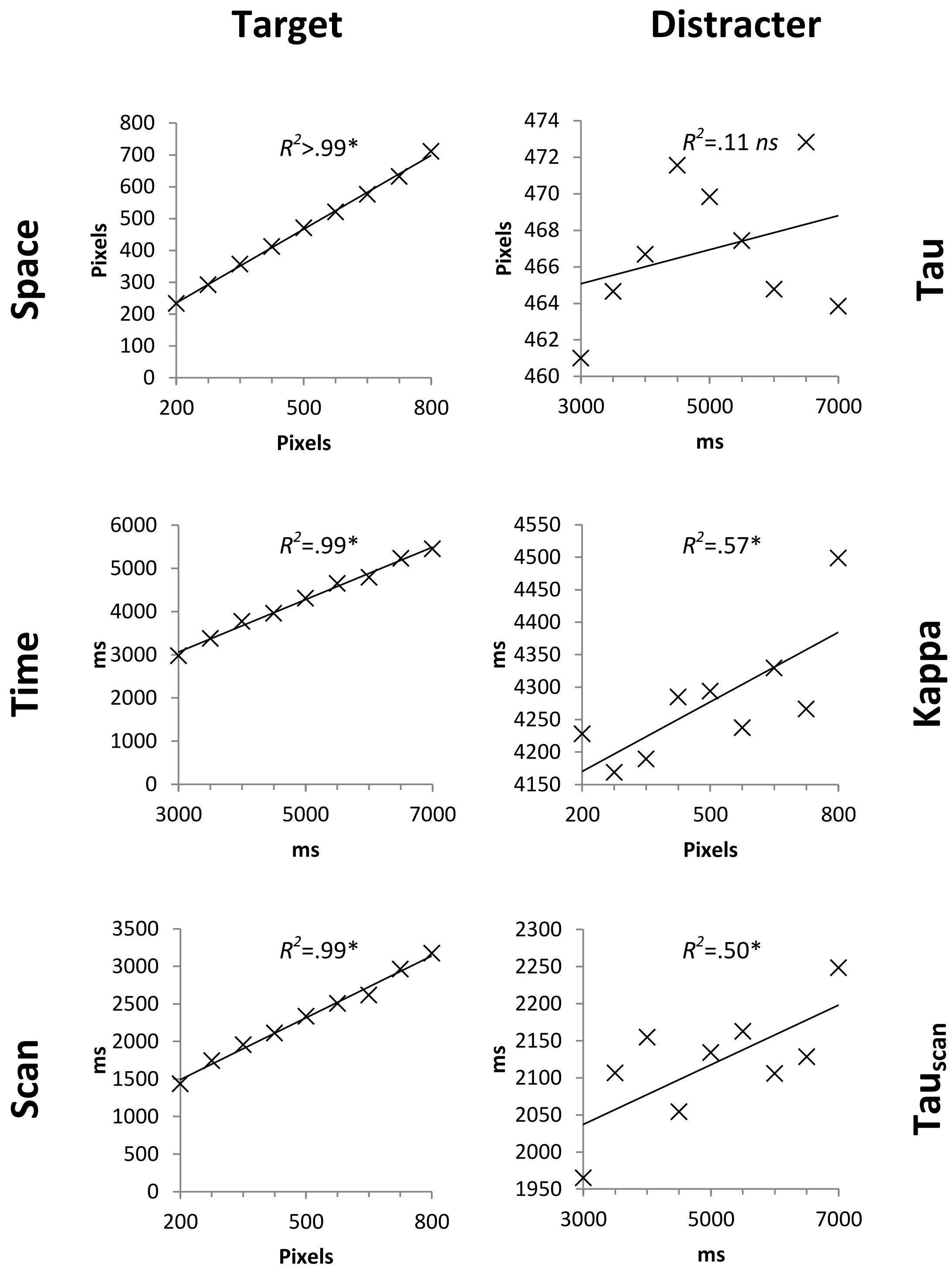

Figure 11.1. The relationship between durations and distances produced and times and line lengths shown for the time, space, and scan tasks. Please note that scales on the y axis vary from graph to graph. Data has been combined across pre, during, and post conditions. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

main effects in the time ($F_{(1,46)}$=589.64, p<.001)[52], space ($F_{(1,46)}$=1840.85, p<.001)[53], and scan ($F_{(1,46)}$=24.69, p<.001)[54] conditions, tau$_{scan}$ ($F_{(1,46)}$=12.38, p=.001), and kappa ($F_{(1,46)}$=10.13, p=.003)[55]. The tau effect was not significant ($F_{(1,46)}$=0.85, p>.05). These indicated that space, time, and scanning estimates all increased with increases in both the dimension that they are intended to measure.

Scan-speed. Visual inspection of Figure 11.2 suggested the presence of outliers. For the purpose of the current analyses any scan value of more than 4 was excluded[56]. These outliers will be reintroduced for a final analysis in the final meta-analysis. The relationship between scan-speed and the tau effect was considered. In line with findings from earlier experiments the participants were split into those showing a positive tau effect and those showing a negative tau effect.

Quadratic trends were fitted to the data shown in Figure 11.3. A weak non-significant relationship ($F_{(2,24)}$=2.37, p=.115) was found for participants showing a positive effect while a moderate significant relationship was found for participants showing a negative tau effect

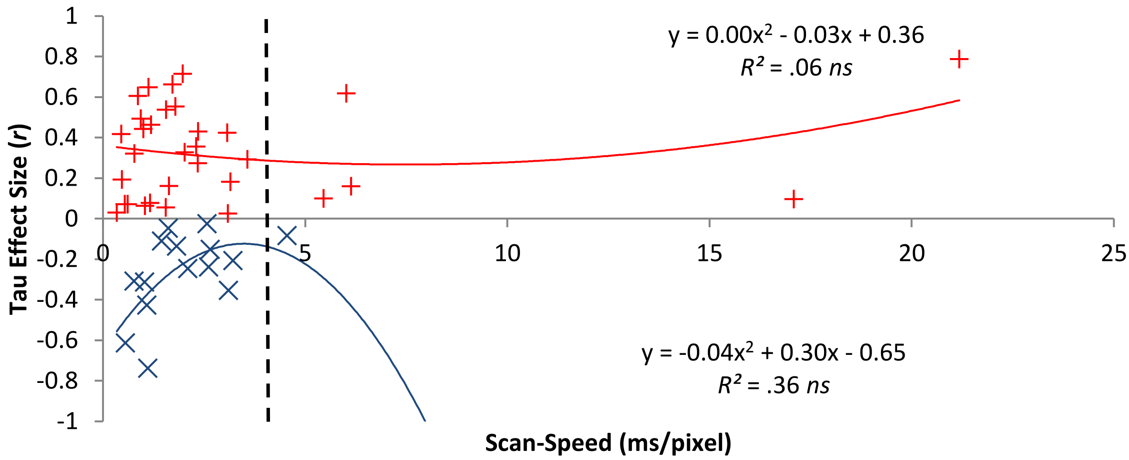

Figure 11.2. The relationship between tau effect size and scan-speed in experiment 6 before outliers have been removed. Data for participant's with positive tau effects are shown as red +s and those with negative tau effects as blue Xs. The trend line for the positive tau effect is red, the trend line for the negative tau effect is shown in blue. Non-significant R^2 values (p>.05) are followed by *ns*. The black dotted line illustrates the point of division between observations included in the scan-speed analyses (to the left of the dotted line) and those considered to be outliers (to the right of the dotted line).

---

52 A significant 7th order effect was also found ($F_{(1,46)}$=8.96 p=.004).

53 There were significant cubic ($F_{(1,46)}$=18.08 p<.001) and fourth order ($F_{(1,46)}$=7.80, p=.008), effects of space as well.

54 There was a significant cubic trend here as well ($F_{(1,46)}$=5.41, p=.025).

55 A significant 4th order was also found ($F_{(1,46)}$=5.17, p=.028).

56 Owing to the change in screen size, had an equivalent criterion been used experiments 4 and 5 any participants with a scan-speed greater than 4.57 (4/0.875) would have been excluded. Under this criterion no participants from experiment 4 or 5 would have been excluded.

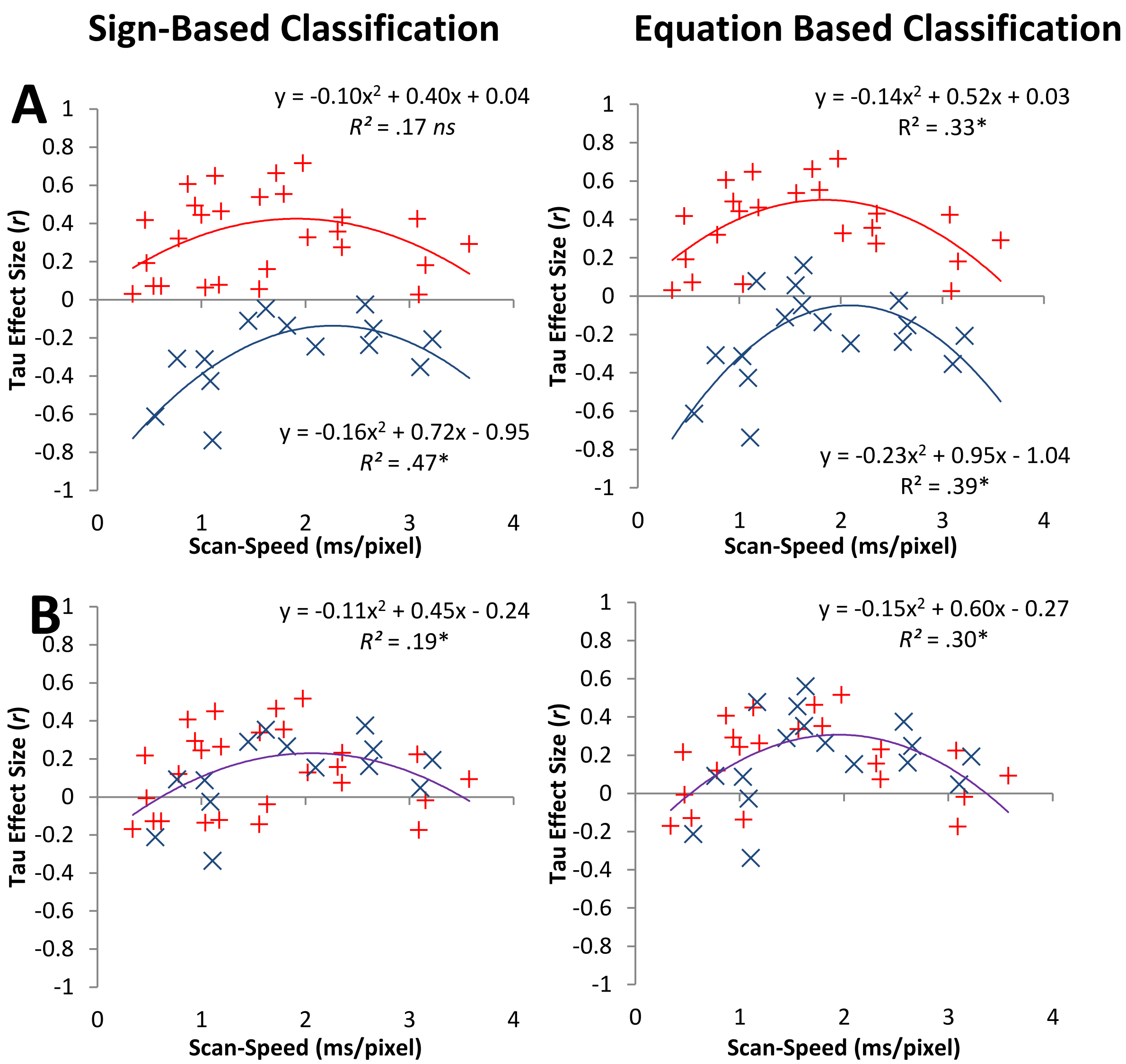

Figure 11.3. The relationship between the tau effect and scan-speed as defined by different groupings. Plots on the left show categorisation by the sign of the tau effect. Plots on the right are categorised according to which line from experiment 5 best fits the current data. Plots in row A show data separated into positive and negative groups. In these cases positive subjects are identified with a + whereas negative subjects are identified by an x. Data in row B show data combined across these groups through the removal of a common effect size. All panels show participant's with positive tau effects as red +s and those with negative tau effects as blue Xs. The trend line for the positive tau effect is red, the negative tau effect blue, and the combined trend line is purple. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

($F_{(2,11)}$=4.90, p=.030). When the two sets are combined through the process used in the combined analysis of experiments 4 and 5 a significant relationship is found ($F_{(2,38)}$=5.39, p=.009).

In a subsequent analysis, participants were categorised as members of the positive tau or negative group according to their proximity to the trend-line from data combined across experiments 4 and 5 (scaled for screen size) rather than whether they showed a positive or negative tau effect. This decreased the R^2 value for the negative group (R^2=.39, $F_{(2,14)}$=4.55, p=.030) and increased it for the positive group (R^2=.33, $F_{(2,21)}$=4.94, p=.018). The combined plot – removing the influence of group membership in the same way as the combined analysis of experiments 4 and 5– has an improved R^2 value (R^2=.30, $F_{(2,38)}$=6.83, p=.003). Formal analysis of such improvements will be left until the combined analyses.

A chi-square analysis was conducted to investigate whether there was a difference in the tendency of participants with negative or positive tau effects to report noticing (having insight into) and/or correcting for potential kappa and/or tau effects. Frequencies for these categories are presented in Table 11.1. No significant effect is found when between these categories are examined together ($\chi^2_{(3)}$=2.22, p>.05). Separate analyses of each question will be presented as part of the final combined analyses.

Table 11.1
Observed counts of participants reporting awareness of the likelihood of the experiment inducing tau and kappa effects, reporting correcting for such an effect, and those producing positive and negative tau effects.

	Insight			
	Yes		No	
	Corrected		Corrected	
Tau effect	Yes	No	Yes	No
Positive	11	17	1	3
Negative	4	10	1	0

Discussion

As expected, time, space, and scanning effects suggested the tasks were completed accurately. Overall a kappa effect but no tau effect was found. An effect of scan task inclusion order was not found as ANOVAs and planned contrasts failed to find any significant difference between conditions.

Initial analysis of the relationship between scan-speed and participants showing a positive tau effect also failed to replicate earlier findings. While visual inspection of the plot suggested that a quadratic pattern was likely to be present a cluster of participants with moderate scan-speeds and low tau effects reduced the quality of the fit.

Counter-intuitively it was the participants showing negative tau effects that showed the strongest relationship. Where participants in previous experiments had failed to produce any results on the downward slope of the function, participants in the current experiment seemed to produce more.

Given that the current experiment aimed to reduce and/or eliminate potential experimenter expectancy effects one potential explanation of the observed results would be to conclude that the manipulation worked. This would imply that the previous results were influenced to some degree by participants doing what they felt the experimenter expected of them. However, while this explanation fits the pattern of overall kappa and tau effect findings, as effects have been observed when they were expected to occur but not when they were not expected to occur, it does not explain the presence of a relationship between scan-speed and tau effect size. If experimenter expectancy effects were still influencing this particular relationship a significant trend should have been observed for the positive tau participants, not the negative tau participants.

One likely explanation for the failure to find a significant effect for the relationship between the positive tau effect group and scan-speed is the misclassification of subjects. When comparing the plot of the positive participants for the current experiment to those of experiments 4 and 5 a key difference is noticeable; the current experiment has a number of participants showing weakly positive tau effects where large tau effects should be expected. While this could suggest that something in the current experiment is producing this result an alternative explanation is that these subjects are correcting for a large tau effect. Normally this would lead to a negative tau effect. However, if the correction was insufficient and/or the initial tau effect was sufficiently large it is foreseeable that a large positive tau effect may be corrected to a weakly positive tau effect. Essentially a participant who should move to the negative group may end up polluting the positive data.

In order to test this, the combined equation produced by data from experiments 4 and 5 was fitted to the data of the current experiment. Participants were assigned to conditions according to which condition produced the lowest SSR for each participant rather than their negative or positive grouping[57]. This change produced a significant quadratic relationship for positive participants, negative participants, and the combined analysis. The quadratic function appeared to be similar in the new negative and positive groupings. Like earlier results, they appear to be separated by a constant effect size regardless of the scan-speed participants had been producing.

[57] This fit improved further when an adjustment was made for relative screen size.

Screen Size. The adjustment in screen size also appears to have had some impact on the results. As shown in Figure 11.4, fitting functions to the data indicates that a smaller screen size is associated with a tighter quadratic function. This makes some sense as a smaller screen will lead to smaller visual stimuli being displayed. If this is found to be true, it could suggest that something like an ideal or resonant scan-speed may exist and be dependent on the stimuli presented. In this case, larger tau effects would be associated with lower speeds than in previous experiments where larger screens were used. That is to say that when the duration of the stimuli is held constant but the space is reduced the optimal travel speed is also reduced. This means participants who take less time to cover the same amount of pixels, participants who scan faster, produce larger tau effects.

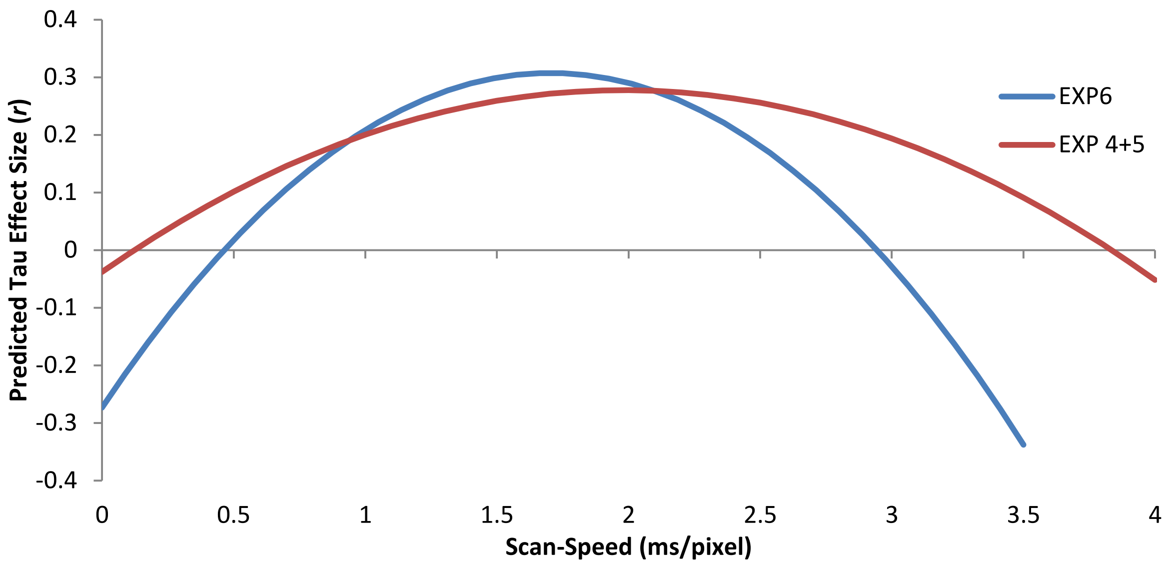

Figure 11.4. Combined group models generated for small (blue) and large (red) screens when not adjusting for screen size

Chapter 12. Experiment 7

In explaining the tendency of different scan-speeds to correlate with different tau effect sizes, previous experiments suggested that scan-speed may provide an index of imputed velocity. As shown in Figure 11.4, the change in screen size in experiment 6 may have altered the scale of the relationship. If this change shows a real effect, it would suggest that the line and time lengths used in each experiment create some sort of resonant speed specific to the stimulus set.

Screen size is somewhat compelling. It provides an objective measurement that changed across the experiments and changes in screen size, shown in Figure 11.4, were associated with changes to the model that best fit the data. In the current experiment this association will be tested through attempts to compress and stretch the quadratic relationship modelling the association between the scan-speed and the size of the tau effect. This will be achieved by varying the spatial length and temporal duration of the stimuli used. It is expected that the use of spatially shorter lines will compress the quadratic trend horizontally towards 0 whereas spatially longer lines will stretch it horizontally further from zero. Similarly the use of longer times should compress the curves towards zero and shorter times should cause them to stretch. These patterns are depicted in Figure 12.1. If these manipulations are successful participants showing the largest tau effects should be those who scan at a speed appropriate to the stimuli. The lawful variation in the scan-speed associated with the largest tau effect will be referred to as the resonance account. The alternative, referred to as the

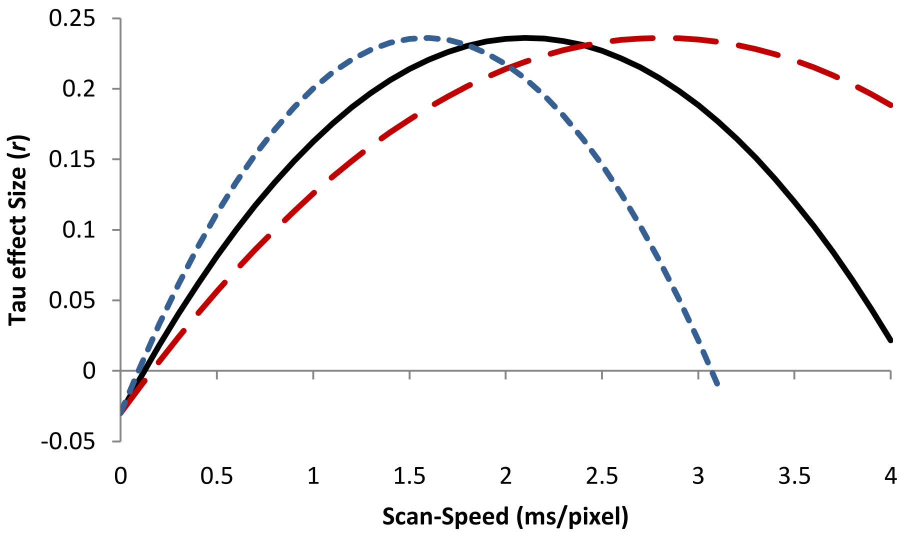

Figure 12.1. Predictions for the change in the relationship between scan-speed and tau effect size as the ratio between the spaces and times used changes. The black line represents the space-time combinations used so far and is based on the equation produced in experiment 5. The blue line with small dashes represents the expected change when shorter lines (or longer times) are used. The red line with large dashes represents the expected change when shorter times (or longer lines) are used. If changing the space-time ratio has no effect, all conditions should produce the black line.

universal best scan-speed account, would be that those who scan at a specific speed, regardless of the size of the stimuli, would always have the largest tau effects.

In order to test these hypotheses the line lengths and durations will be reduced to 75% of their original size. This 25% reduction is in addition to that created by the change in screen size in previous experiments as the smaller screen will be used. Assuming that the change in screen size did produce a shift in the quadratic trend an additional 25% reduction should be more than large enough to create another shift.

Methods

Participants. Students from the University of Auckland and some members of the general public volunteered to run in this experiment. Of an initial 60 participants 52 met the inclusion criteria outlined in experiment 4. (N_{males}=18, M_{age}=23.51, $Range_{age}$ =18-30, 15 Right Handed, 2 Ambidextrous, 1 Left Handed; $N_{females}$=34, M_{age}= 22.34 $Range_{age}$ =17-54, 28 Right Handed, 4 Ambidextrous, 2 Left Handed). All participants were offered a $10 petrol voucher to go towards reimbursing their travel costs. All participants reported normal or corrected to normal vision.

Materials. The materials were the same as those used in experiment 6 except, in some conditions, the spatial and/or time stimuli were reduced by 75%. The adjusted spatial lengths, measured in pixels, were rounded to whole numbers where necessary. They were 150, 206, 263, 319, 375, 431, 488, 544, and 600 pixels. The durations did not need to be rounded and ranged from 2250ms to 5250ms in 375ms steps.

Procedure. As in experiment 4, participants had the space, time, and scan tasks explained to them before the experiment. As in experiment 6, participants were randomly allocated a participant number which in turn allocated them to one of three counterbalance conditions. In the short time condition the spatial lengths were the same as those used in previous experiments but the durations were reduced by 75%. In the short space condition the spatial lengths were reduced by 75% but the durations were the same as previous experiments. In the short both condition both space and time were reduced by 75%. The experimental procedure was the same as the scan task during condition in experiment 6. Space, time, and scan trials were intermixed and participants were asked if they were aware of tau/kappa effects, if they corrected for those effects, and to provide their thoughts on what had happened in the task.

Results

Standard effects. Separate mixed design ANOVAs were conducted for the time, space, and scan responses when the underlying stimulus values were not corrected for. Stimulus size was entered as a between subjects variable. The temporal and spatial lengths of the stimuli were entered as within subject variables. Significant space ($F_{(1,49)}$=2721.15,p<.001)[58], time ($F_{(1,49)}$=636.98,p<.001)[59], scan ($F_{(1,49)}$=58.64,p<.001)[60], tau ($F_{(1,49)}$=18.42,p<.001), kappa ($F_{(1,49)}$=42.19p<.001), and tau_{scan} ($F_{(1,49)}$=16.10, p<.001) effects were found. Significant interactions were found between stimulus size and time ($F_{(2,49)}$=13.12, p<.001) and space ($F_{(2,49)}$=22.14, p<.001) effects. As well as time and the kappa effect ($F_{(1,49)}$=4.98, p=.03)

When the underlying values were scaled to the stimuli displayed the same pattern of results holds with the exception of the interaction between stimulus scale and the time and space effects. In each of these cases the interactions become non-significant. Specifically, significant space ($F_{(1,49)}$=540.92, p<.001)[61], time ($F_{(1,49)}$=631.02, p<.001)[62], scan ($F_{(1,49)}$=66.06, p<.001)[63], tau($F_{(1,49)}$=17.11, p<.001), kappa ($F_{(1,49)}$=41.23, p<.001), and tau_{scan} ($F_{(1,49)}$=16.24, p<.001) effects were found. No other comparisons reached significance. These standard effects are shown in Figure 12.2.

Scan-speed. As the manipulations in the current study are assumed to impact on the relationship between scan-speed and the tau effect the analysis was conducted in two ways. The first ignored the differences in times and spaces in the stimuli. The second took these into account in the calculation of scan-speed. In particular, speeds were multiplied by 1/0.75 in the small space condition, 0.75 in the small time condition, and 1 in the small space and small time condition (as the two multipliers cancel out). In the uncorrected and corrected cases participants were first separated into the positive and negative groups according to their tau effect value and then their proximity to the trend line from the combined data, scaled for screen size, of experiments 4 and 5. In line with the approach taken in experiment 6 any participant with a scan-speed greater than 4 was considered an outlier and excluded from analysis in the current experiment. Due to the scaling of scan times this meant that some participants were included in the corrected analysis but excluded from the uncorrected analysis and vice versa. The data pattern is shown in Figure 12.3.

[58] A significant quadratic effect was also found ($F_{(1,49)}$=5.80, p=.02).

[59] A quadratic term is also significant ($F_{(1,49)}$=6.26, p=.016).

[60] Significant quadratic ($F_{(1,49)}$=6.17, p=.016) and cubic (F=$_{(1,49)}$4.08, p=.049) relationships were also found.

[61] A significant quadratic effect was also found ($F_{(1,49)}$=5.12, p=.028).

[62] A quadratic term is also significant ($F_{(1,49)}$=6.35, p=.016).

[63] A significant quadratic ($F_{(1,49)}$=8.12, p=.006) relationship was also identified.

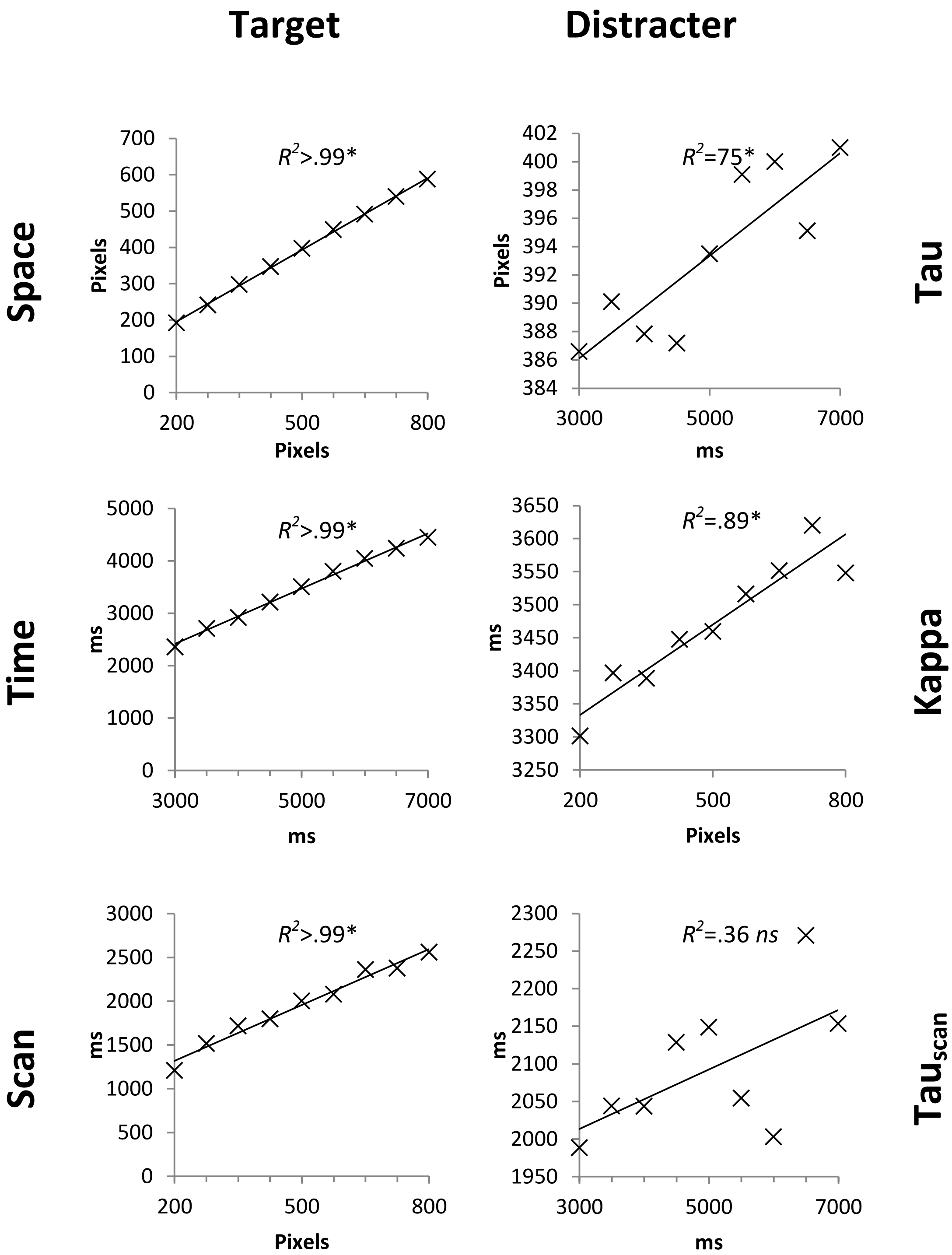

Figure 12.2. The relationship between durations and distances produced and times and line lengths shown for the time, space, and scan tasks. Please note that scales on the y axis vary from graph to graph. Data has been combined across scale conditions after correction for differences in scale. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

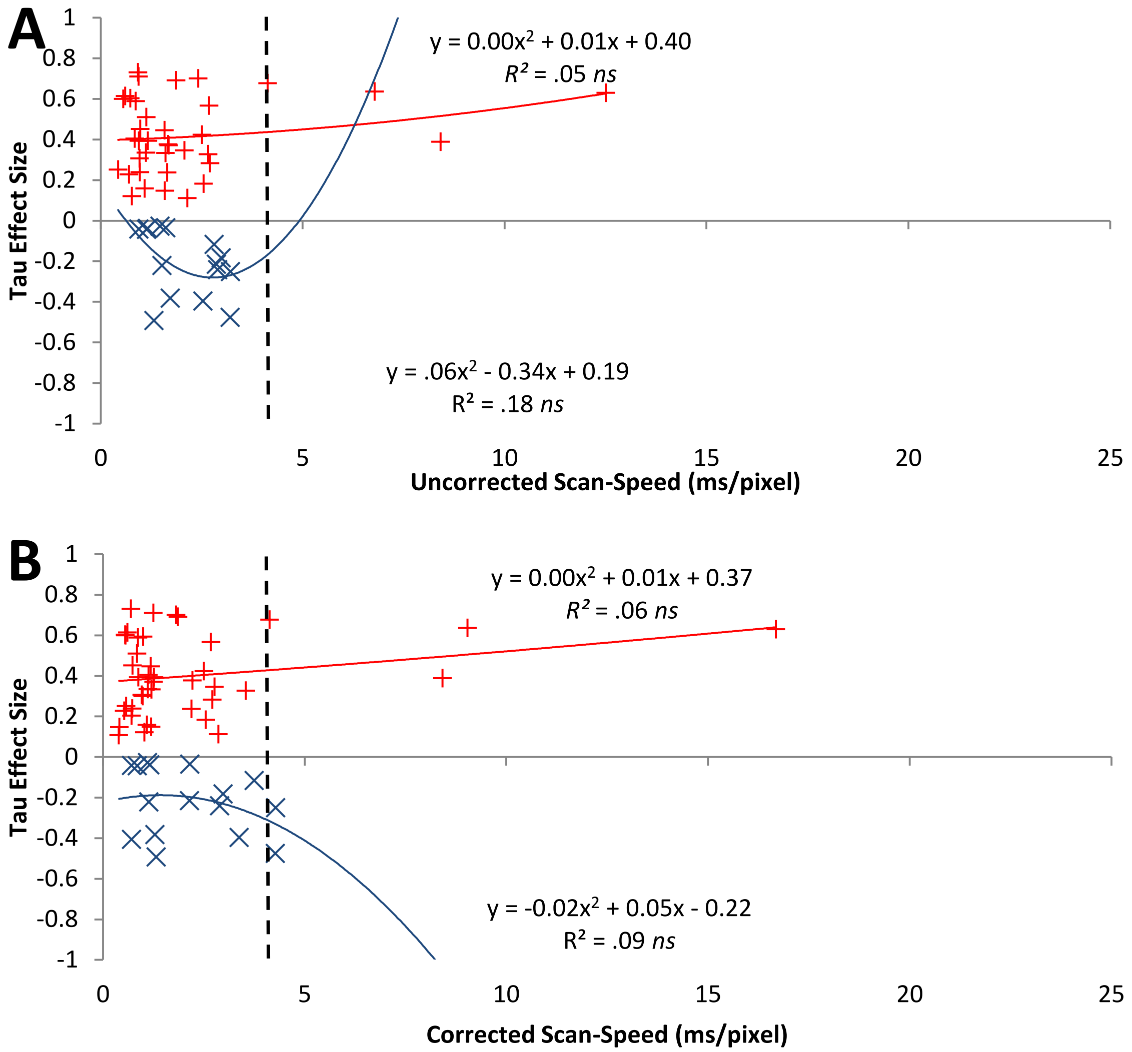

Figure 12.3. The relationship between tau effect size and scan-speed in experiment 7 before outliers have been removed. Groups have been divided by the value of tau effect. Data without a correction for stimulus size is shown in panel A. Data with a correction for stimulus size is shown in panel B. Both panels show participant's with positive tau effects as red +s and those with negative tau effects as blue Xs. The trend line for the positive tau effect is red, the negative tau effect trend line is blue. Non-significant R^2 values (*p*>.05) are followed by *ns*. The black dotted line illustrates the point of division between observations included in the scan-speed analyses (to the left of the dotted line) and those considered to be outliers (to the right of the dotted line).

As shown in Figure 12.3 and Figure 12.4, when the differences between conditions are not considered, there is no apparent relationship between the tau effect and scan-speed. When the differences between conditions are accounted for, separate lines of best fit do not produce a clear quadratic relationship[64]. However, participants with a positive tau effect do begin to visually replicate the expected data pattern.

When the groups are defined according to the method used in experiment 6, a borderline significant effect is found for the corrected positive condition ($F_{(2,32)=}$3.27, p=.051) but not the negative condition. This can be seen in Figure 12.4. When the groups are combined by removing the constant values from the data in experiment 4 and 5, a significant effect remains (R^2=.19, $F_{(2,44)}$=5.07, p=.010). This pattern suggests that the tau effect increases with scan-speed until reaching a peak at a scan-speed of approximately 1.1. After this point tau effect sizes decrease with increasing scan-speed.

A chi-square analysis was conducted to investigate whether there was a difference in the tendency of participants with negative or positive tau effects to report noticing (having insight into) and/or correcting for potential kappa and/or tau effects. Frequencies for these categories are presented in Table 12.1. No significant effect is found when between these categories are examined together ($\chi^2_{(3)}$=2.18, p>.05). Separate analyses of each question will be presented as part of the final combined analyses.

Table 12.1
Observed counts of participants reporting awareness of the likelihood of the experiment inducing tau and kappa effects, reporting correcting for such an effect, and those producing positive and negative tau effects.

	Insight			
	Yes		No	
	Corrected		Corrected	
Tau effect	Yes	No	Yes	No
Positive	12	15	1	8
Negative	5	8	1	1

[64] Appendix C shows data separated by stimulus size.

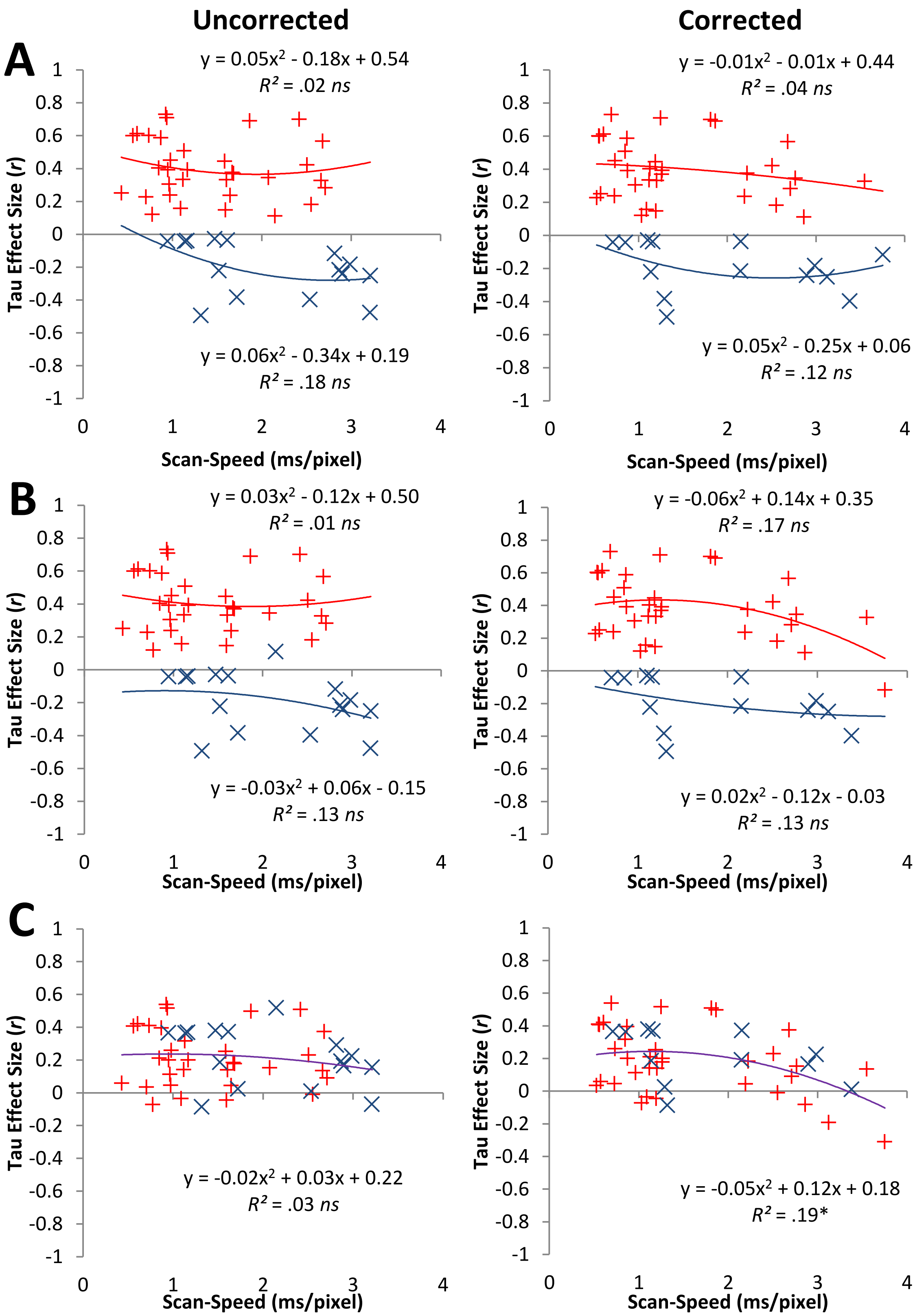

Figure 12.4. The relationship between the tau effect and scan-speed when stimulus size is uncorrected (left column) and corrected (right column). Participant's with positive tau effects are shown as red +s and those with negative tau effects as blue Xs. The trend line for the positive tau effect is red, the negative tau effect blue, and the combined trend line is purple. Group identity is determined by whether individual tau effects were positive or negative (row A) or according to proximity to the lines of best fit from the combined data of experiments 4 and 5 (rows B and C). In row C the negative and positive groups are combined by removing the constants associated with the positive and negative groups found in the combined data from experiments 4 and 5. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

Discussion

Participants performed the space, time, and scan tasks well and produced significant kappa and tau effects across the experiment. Inspection of the uncorrected scatter plot comparing scan-speed with tau effect size suggested it was unlikely that a relationship between the two existed. However, the appearance of the scatter plot improved substantially after correcting for stimulus size and trends in the expected direction were produced. Despite this improvement the relationships did not reach significance when the data was split into positive and negative observations. However, if the groups are defined according to their proximity to the pattern observed in earlier work the combined group does reach significance.

Given the weak nature of the data pattern observed in previous experiments, it is fair to suggest that expecting a clear replication of the target pattern after correcting for stimulus size may have been overly optimistic. This is particularly true given that the correction creates a horizontal shift in the data. For a quadratic data pattern, if a horizontal shift is off even slightly the result can be highly detrimental in terms of the production of the associated sum of squares.

Given the apparent replication in the direction of the trends, and the improvement observed in the R^2 value of the combined data after correction, the results of the current experiment are encouraging. In addressing the question of whether scaling improves the fit of the model it will be beneficial to consider these results in conjunction with those of the previous experiments in a combined analysis.

Chapter 13. Combined Analyses

While a fairly consistent picture has emerged across the series of experiments it was of interest to combine the data to increase statistical power and provide a better test for some of the more speculative claims presented so far. In particular, pooling data should allow for a definitive test of the contribution of screen/stimulus size to kappa and tau effects, and allow for a better account of outliers in the relationship between scan-speed and tau effect size. The contribution of scan task inclusion order, as described in experiments 4, 5, and 6, will also be reviewed by combining the participants in experiment 5 with those in the post condition of experiment 6. Standardisation analyses trialled in experiment 4 will also be re-explored in order to address the question of the relative size of the kappa and tau effects. The results of the combined analyses will contribute to the general discussion of the findings across experiments.

Methods

Participants. All 155 participants meeting the selection criteria[65] are considered here ($N_{females}$=111, M_{age}=23.22, $Range_{age}$ =17-54, 96 Right Handed, 9 Ambidextrous, 6 Left Handed; N_{males}=44, M_{age}=23.77, $Range_{age}$=18-33, 39 Right Handed, 4 Ambidextrous, 1 Left Handed). Break downs of the statistics for specific experiments can be found in the relevant chapters. Participants who failed to meet these criteria and completed the task are considered in a separate analysis.[66]

Materials. Data from each participant in each of experiments 4, 5, 6, and 7 is considered here.

Procedure. Experimental data was coded by scan task inclusion order (pre, during, and post), and stimulus size. Analyses were conducted with and without corrections for experimental manipulation of stimulus size being employed. Scan-speeds of participants who used the 17 inch screen were scaled to match those who used the 15 inch screen.[67]

[65] An r value of at least 0.5 in the space, time, and scan tasks.

[66] A basic analysis of the standard effects seen in the excluded participants can be found in Appendix D.

[67] This is the opposite of the scaling used in experiments 4 and 5. Scaling to a 15 inch screen has been used here as most participants ran using a 15 inch screen and the boundary point for outliers has been set from results from a 15 inch screen. Scaling to a 17 inch screen was maintained in experiments 4 and 5 as most participants ran using a 17 inch screen in these cases. Retaining this scaling also helps explain why stimulus scale was considered an important variable for consideration in experiment 7. If the scan-speeds are not scaled for screen size for the combined analyses scaled the observed pattern of results is largely unchanged.

Results

Preliminary analyses found no effects of or involving the first language, handedness, or gender of participants. As such these measures are excluded from the current analysis. Stimulus size and task order showed some interactions in some analyses and are included below. The analyses presented here focus on the results of planned contrasts conducted as part of mixed design ANOVAs.

Standard effects. Stimulus size and task order were included in the analyses to determine if either of these variables had an effect on the results. Contrasts were conducted as part of separate mixed design ANOVAs on space, time, and scan task responses. Stimulus size and task inclusion order were included as between subjects factors. The temporal and spatial lengths of the stimuli were included as within subjects factors. Separate analyses were conducted for the uncorrected and corrected data.

Uncorrected. Significant linear trends were found for time ($F_{(1,148)}$=1080.29, p<.001), space ($F_{(1,148)}$=3858.11, p<.001), kappa ($F_{(1,148)}$=22.06, p<.001), tau ($F_{(1,148)}$=13.10, p<.001), scan ($F_{(1,148)}$=93.38, p<.001), and tau_{scan} ($F_{(1,148)}$=28.97, p=.001) effects. These effects are shown in Figure 13.1.

An interaction was found between scan task inclusion order and time ($F_{(1,148)}$=4.21, p=.017). Interactions were also found between stimulus size and time ($F_{(1,148)}$=6.99, p<.001), space ($F_{(1,48)}$=9.28, p<.001), and scan ($F_{(1,148)}$=5.24, p=.006).

An interaction was found between time and space positions in the time effect ($F_{(1,148)}$=4.68, p=.032) and the tau effect ($F_{(1,148)}$=4.35, p=.014).

Corrected. After corrections for stimulus size were made the pattern of results remained largely the same. The key difference was that interactions involving stimulus size no longer reached significance, suggesting that the corrections had worked. Significant linear trends were found for time ($F_{(1,148)}$=1129.37, p<.001), space ($F_{(1,148)}$=2751.32, p<.001), kappa ($F_{(1, 148)}$=23.07, p<.001), tau ($F_{(1,148)}$=12.74, p<.001), scan ($F_{(1,148)}$=99.51, p<.001), and tau_{scan} ($F_{(1,148)}$=27.70, p<.001) effects. These effects are shown in Figure 13.2.

Interactions with inclusion order were found for tau ($F_{(1,148)}$=4.20, p=.021), and scan ($F_{(1,148)}$=3.96, p=.021) effects. This suggests that when participants are required to respond to the spatial size of a stimulus their size estimates decrease as the scan task is included later in the experiment for the tau and scan effects.

Contrasts were used to address the interactions. Theoretical reasons outlined in experiments 4, 5, and 6 predict a difference between the post condition and the pre and during

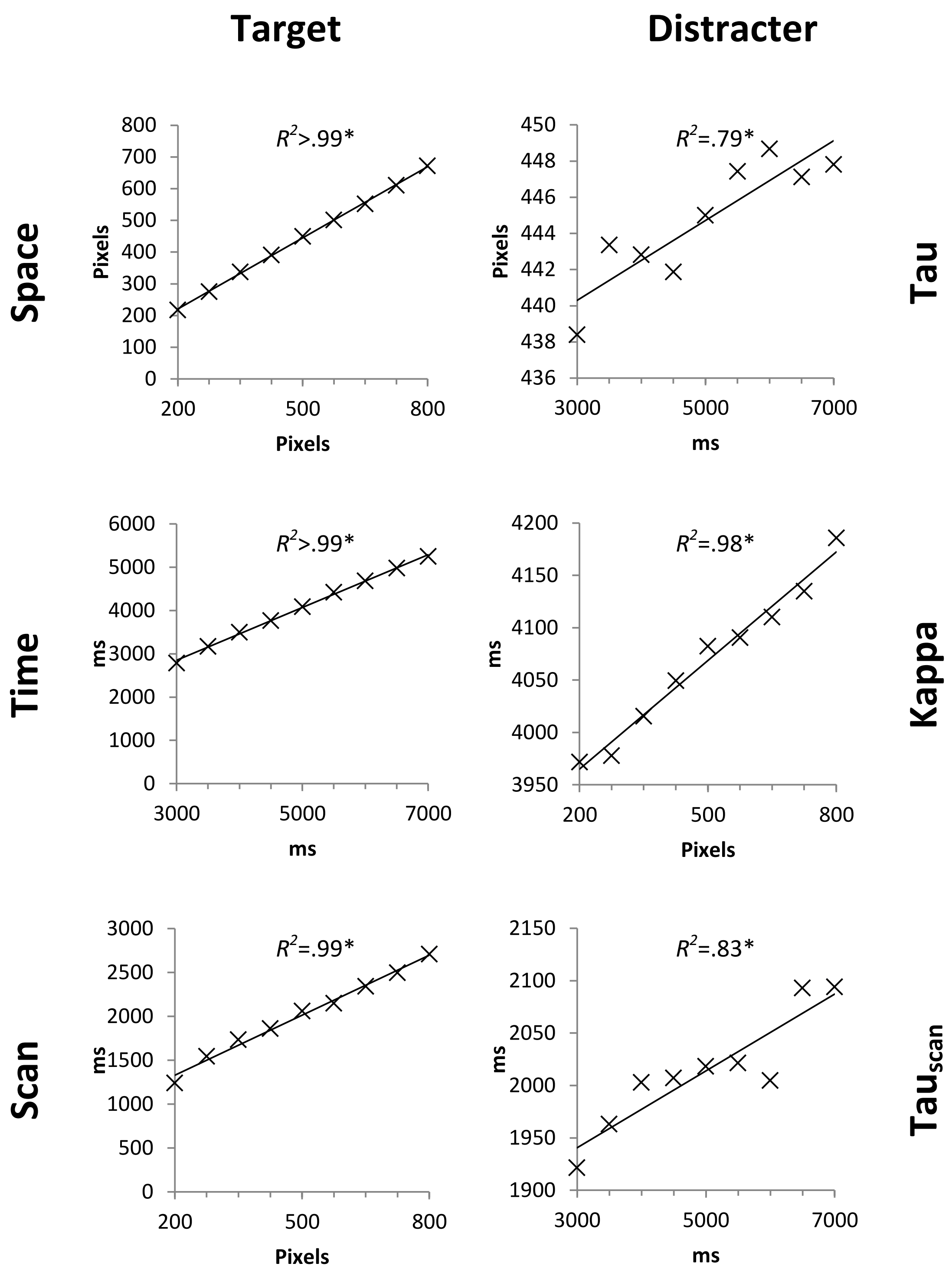

Figure 13.1. The relationship between durations and distances produced and times and line lengths shown for the time, space, and scan tasks. Please note that scales on the y axis vary from graph to graph. Data has been combined across different stimulus scales and is uncorrected for stimulus scale. Significant R^2 values (p<.05) are followed by an *.

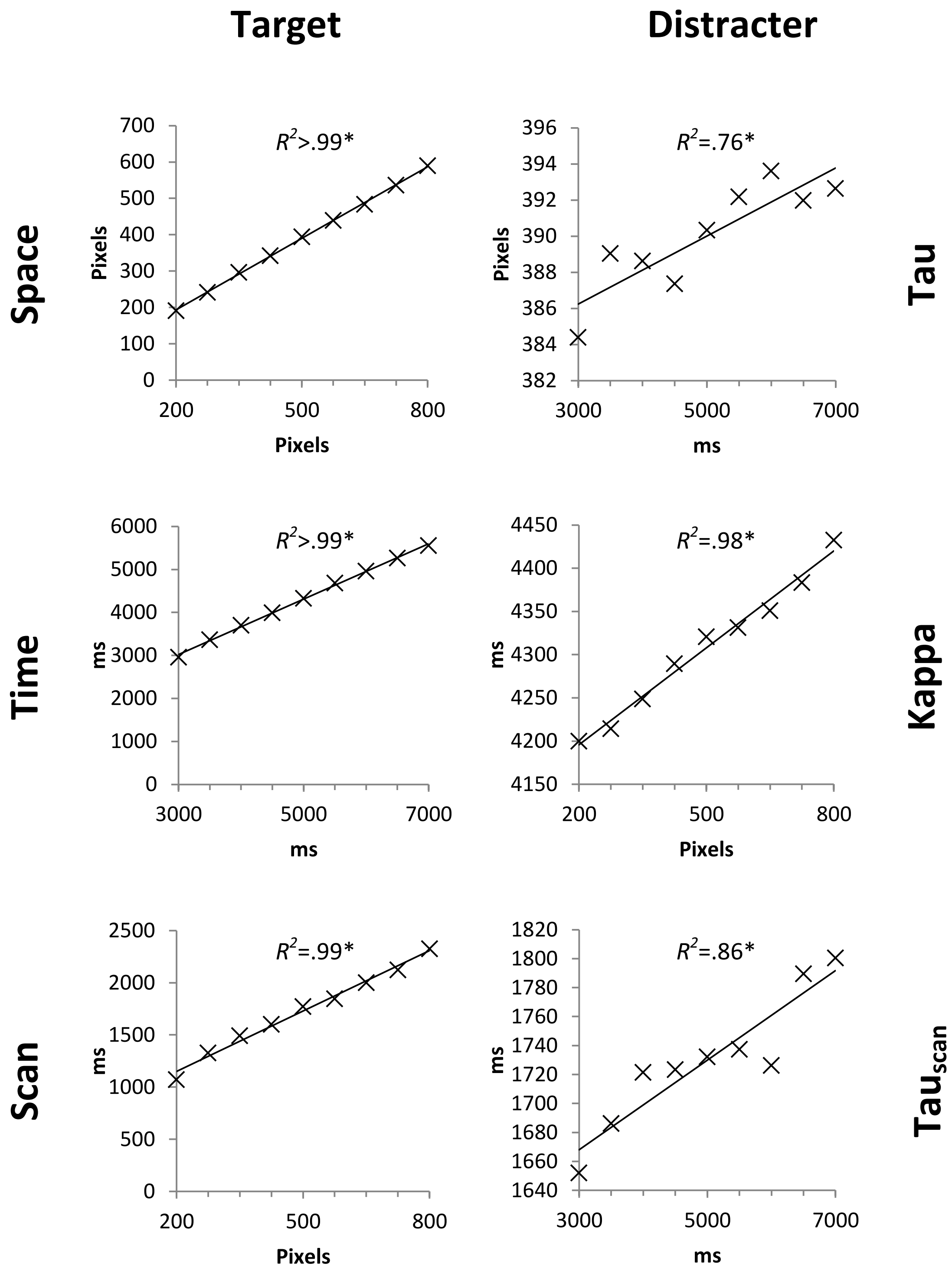

Figure 13.2. The relationship between durations and distances produced and times and line lengths shown for the time, space, and scan tasks. Please note that scales on the y axis vary from graph to graph. Data has been combined across different stimulus scales and is corrected for stimulus scale. Significant R^2 values (p<.05) are followed by an *.

conditions for the tau effect. The post condition was found to have a weaker tau effect than the pre and during conditions ($F_{(1,149)}$=8.13, p=.005). No difference between the pre and during conditions was found ($F_{(1,149)}$=0.36, p>.05). The data pattern is presented in Figure 13.3.

Visual inspection of the interaction between inclusion order and scan task performance, shown in Figure 13.3, suggested that the pattern in the scan task was in line with fatigue. This was confirmed by contrasts comparing the pre condition to during and post conditions, then comparing the during condition with the post condition. The slope of the scan task was steeper in the pre condition than the during and post conditions ($F_{(1,149)}$=12.29, p=.001). There was no difference in scan slope between the during and post conditions ($F_{(1,149)}$=0.31, p>.05).

Scan-speed and the tau effect. For the purposes of these analyses 10 participants who had a scan-speed greater than 4 were excluded as they were deemed to be outliers. This division is shown in Figure 13.4 by a dotted black line. For the sign-based classification analyses, the sign of each tau effect (positive or negative) was used to allocate observations to the positive or negative group. For the model-based classification analyses, tau effects were allocated to the group that minimised the RSS to the equation from experiment 5[68]. When calculating the combined group values, the constant

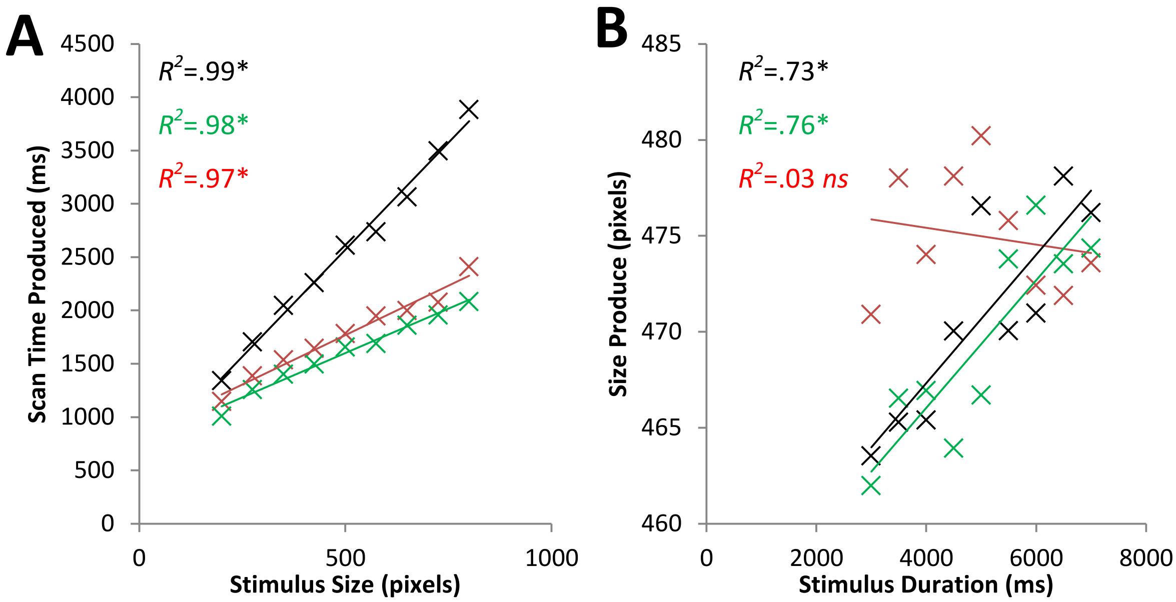

Figure 13.3. Plots of the interactions involving scan task inclusion order in the corrected data. Scan task pre is shown in black, during in green, post in red. Panel A shows the scan effect, panel B shows the tau effect. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

[68] The full complement of group vs. correction combinations was considered but the presented combinations gave the best fit of the uncorrected and corrected data in that they provided the highest R^2 and lowest AICc values. AICc values are explained in the section "Model selection".

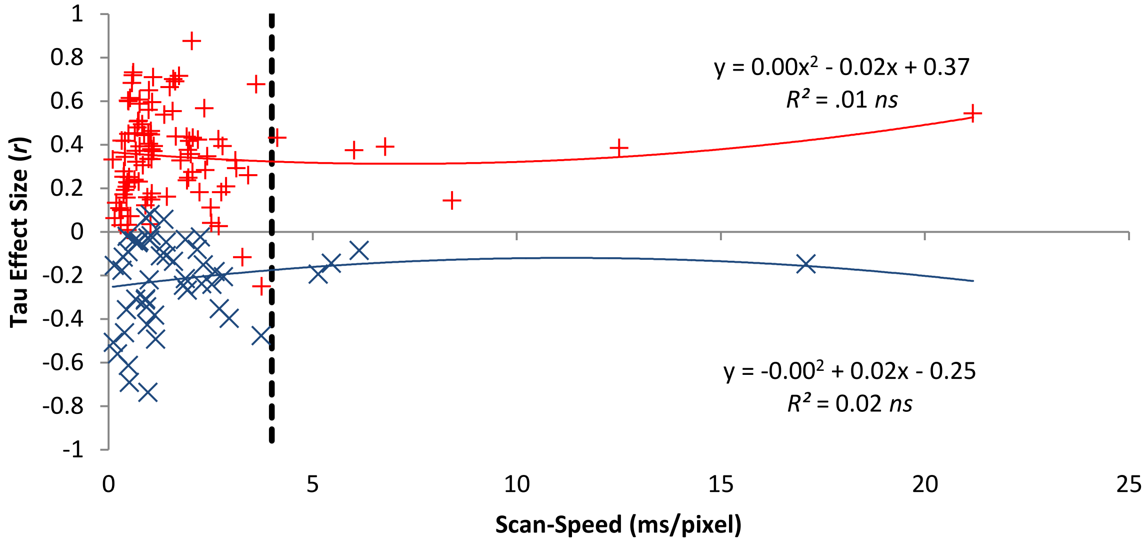

Figure 13.4. Tau effect sizes and scan-speeds, adjusted for stimulus size, for participants from experiments 4-7. The black dotted line illustrates the point of division between observations included in the main combined analyses (to the left of the dotted line) and those considered to be outliers (to the right of the dotted line). Participant's with positive tau effects are shown as red +s and those with negative tau effects as blue Xs.

resulting from fitting separate quadratic trends to the positive and negative groups were removed. The trends associated with the resulting groupings are shown in Figure 13.5.

The presence of a main effect of scan task inclusion order was tested through inclusion in regression analyses. A significant benefit was found for the inclusion of a main effect of inclusion order in forward regression for the positive ($F_{(1,91)}$=6.67, p=.011) and combined ($F_{(1,140)}$=8.40, p=.004) groupings. While the main effect of inclusion order is not significant in the negative group no significant interaction was found between inclusion order, group identity, scan-speed, or squared scan-speed. As a result inclusion order is included in models for the positive, negative, and combined groupings.

Terms associated with an interaction between inclusion order and scan-speed/scan-speed squared were included by SPSS when running backward regression in place of the main effect. If the interaction terms are entered instead of the main effect a similar R^2 value is produced ($R^2_{\text{interactions}}$=.25, $R^2_{\text{main effect}}$=.24). As the interaction model includes an extra parameter (2 interaction terms vs. 1 main effect term) the main effect model is preferred[69].

[69] Preference can be established from AICc values ($\text{AIC}_{\text{main effect}}$=-477, $\text{AICc}_{\text{interaaction}}$=-476) or an F change statistic ($F_{(1,139)}$=0.87, p>.05). AICc values are explained in the section "Model selection".

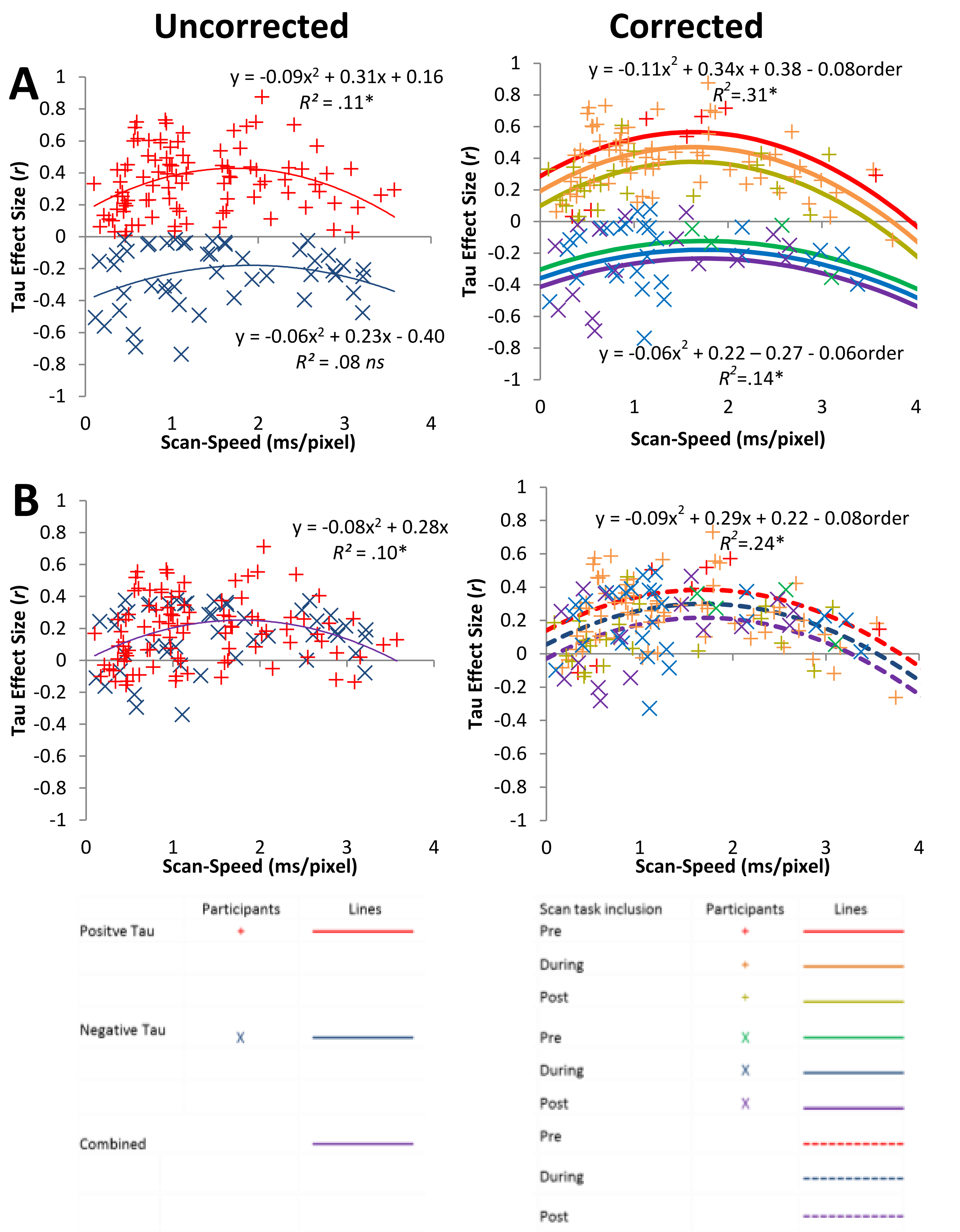

Figure 13.5. The relationship between the size of the tau effect and scan-speed for the uncorrected and corrected models when positive and negative values are considered separately (panel A) and combined (panel B). Plots on the left show the data before corrections for stimulus size and scan task inclusion order are implemented. Plots on the right who data corrected for stimulus size. Different scan task inclusion orders are represented by different colours and different lines of best fit. Outliers are excluded in all cases. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

Sign-based classification. Significant quadratic trends were found for the positive group ($F_{(2,95)}$=5.87, p=.004) and the combined data ($F_{(2,142)}$=7.57, p=.001). The trend in the negative group reached significance for a one-tailed test only ($F_{(2,44)}$=1.83, p>.05).

Model-based classification. Significant quadratic trends were found for the positive ($F_{(2,94)}$=15.01, p<.001), negative ($F_{(2,45)}$=3.59, p=.036), and combined ($F_{(2,142)}$=17.46, p<.001) groups.

Associations with the kappa effect. The correlation between the kappa effect and the tau effect, shown in Figure 13.6, is positive and significant (r=.43, p<.001). As the strength of a participant's tau effect increases, on average, the strength of their kappa effect also increases. If the

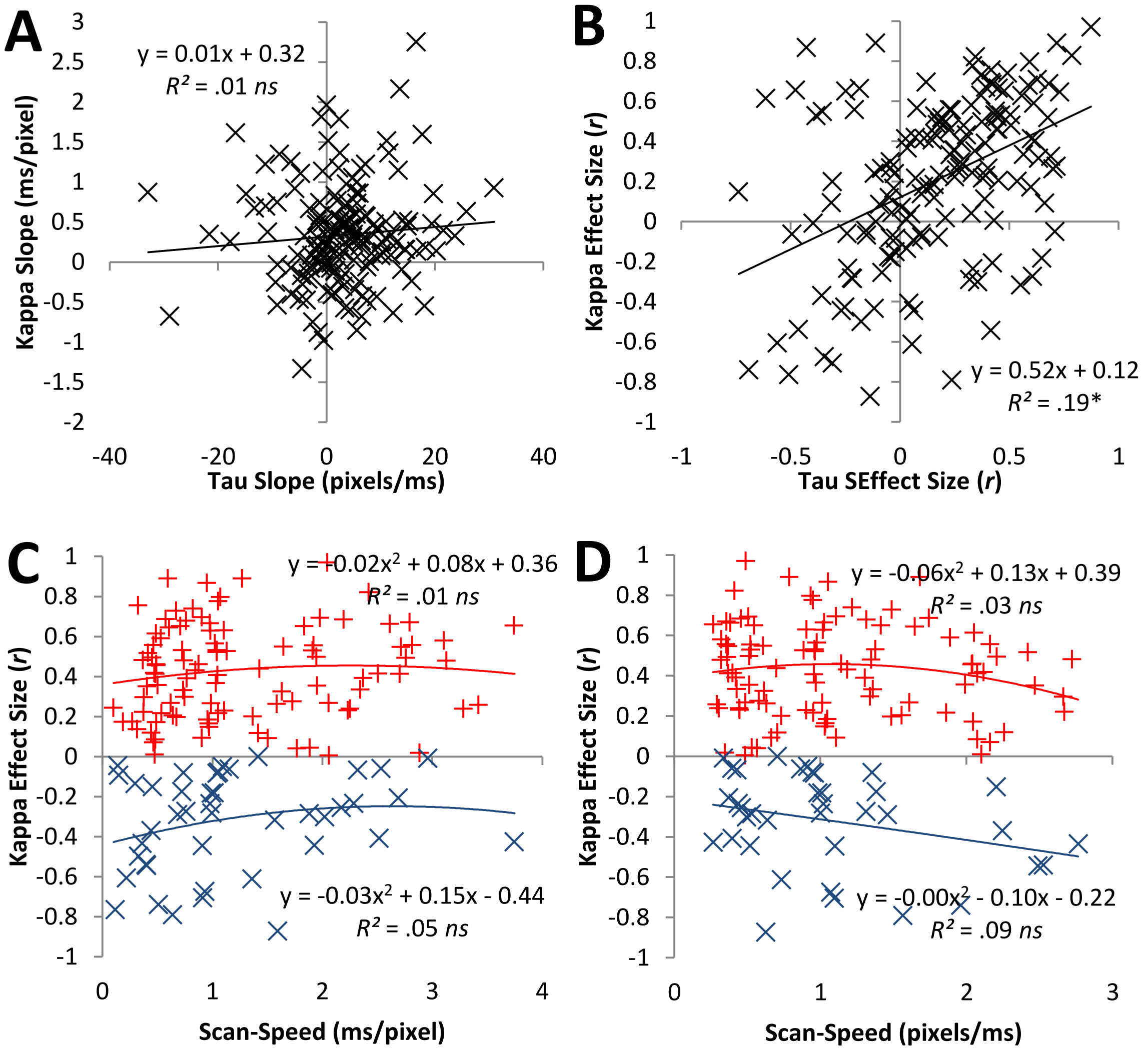

Figure 13.6. The relationship between the kappa effect and the tau effect slopes (A) and sizes (B) and scan sped (C+D) observed in individual participants. Speed has been calculated in 2 different ways. In Panel C speed is shown as pixels per millisecond. In Panel D speed has been calculated as milliseconds per pixel. In panels C and D participant's with positive tau effects as represented by red +s and those with negative tau effects as blue Xs. The trend line for the positive tau effect is red, the negative tau effect is blue. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

kappa effect is included in the model of the tau effect it uniquely accounts for 1% of the variance and the R^2 value increases from .24 to .25 but the increase is not significant ($F_{(1,139)}$=1.98, p>.05). The association between the kappa effect and scan-speed, shown in Figure 13.6, remains non-significant.

Associations with the tau$_{scan}$ effect. As shown in Figure 13.7, the correlation of the tau effect, the kappa effect, and scan-speed with the tau$_{scan}$ effect is non-significant.

Accounting for outliers. Figure 13.8 shows the association between tau effect size and scan-speed for all participants, including outliers. Previous work suggested the association between the tau effect size and scan-speed reflected resonance between time and space. Two key points in this relationship were the resonant point, where tau effects approach zero, and the optimal scan-speed, where maximal tau effects are observed. Resonance implies the potential for a harmonic

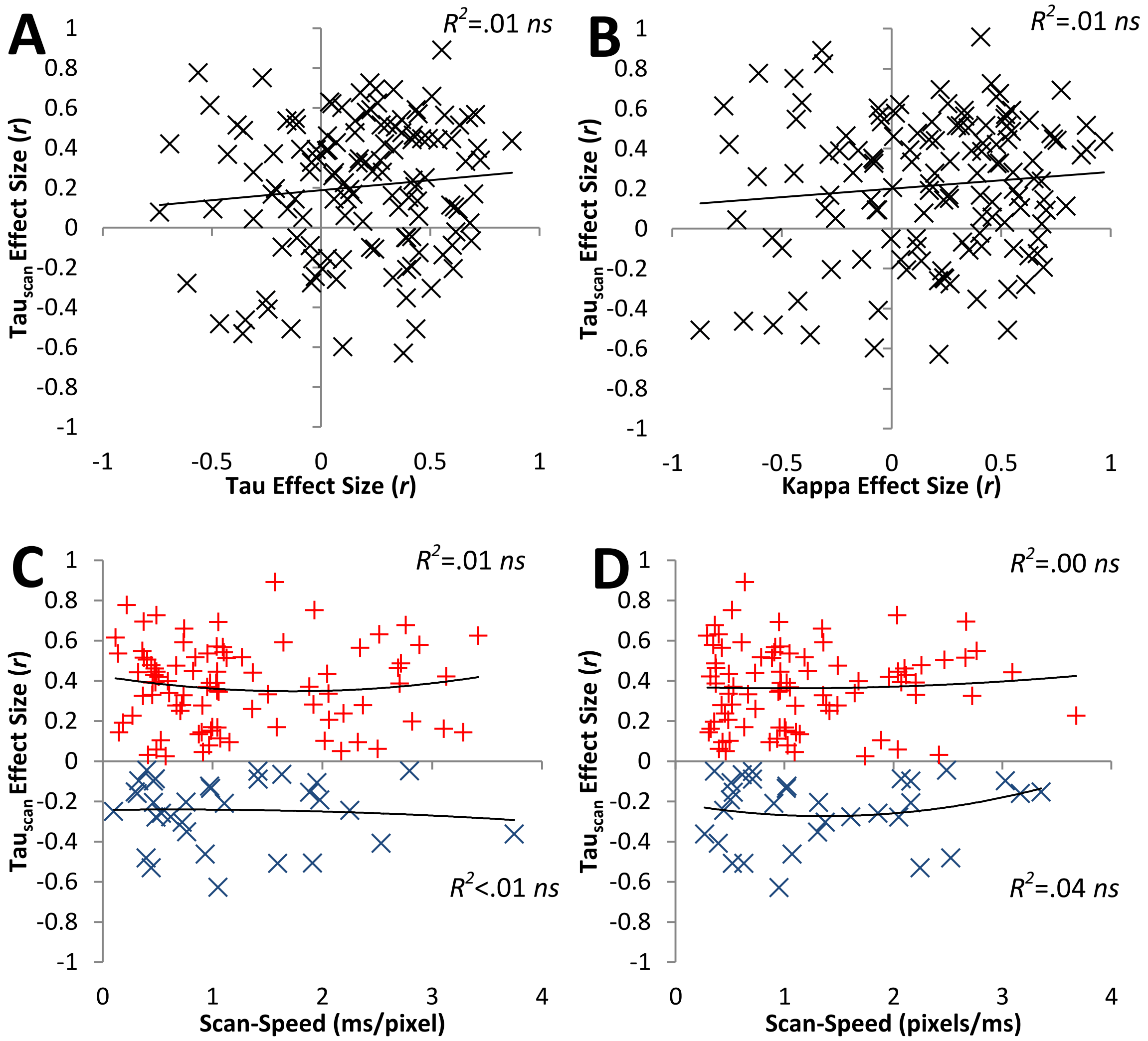

Figure 13.7. Associations of the tau effect (A), the kappa effect (B), and scan-speed (C and D) with the tau$_{scan}$ effect. In the case of scan-speed, outlying scan-speeds have been excluded and separate quadratic trends have been fitted to the positive and negative tau$_{scan}$ participants. In panels C and D participant's with positive tau effects as represented by red +s and those with negative tau effects as blue Xs. The trend line for the positive tau effect is red, the negative tau effect is blue. Non-significant R^2 values (p>.05) are followed by *ns*.

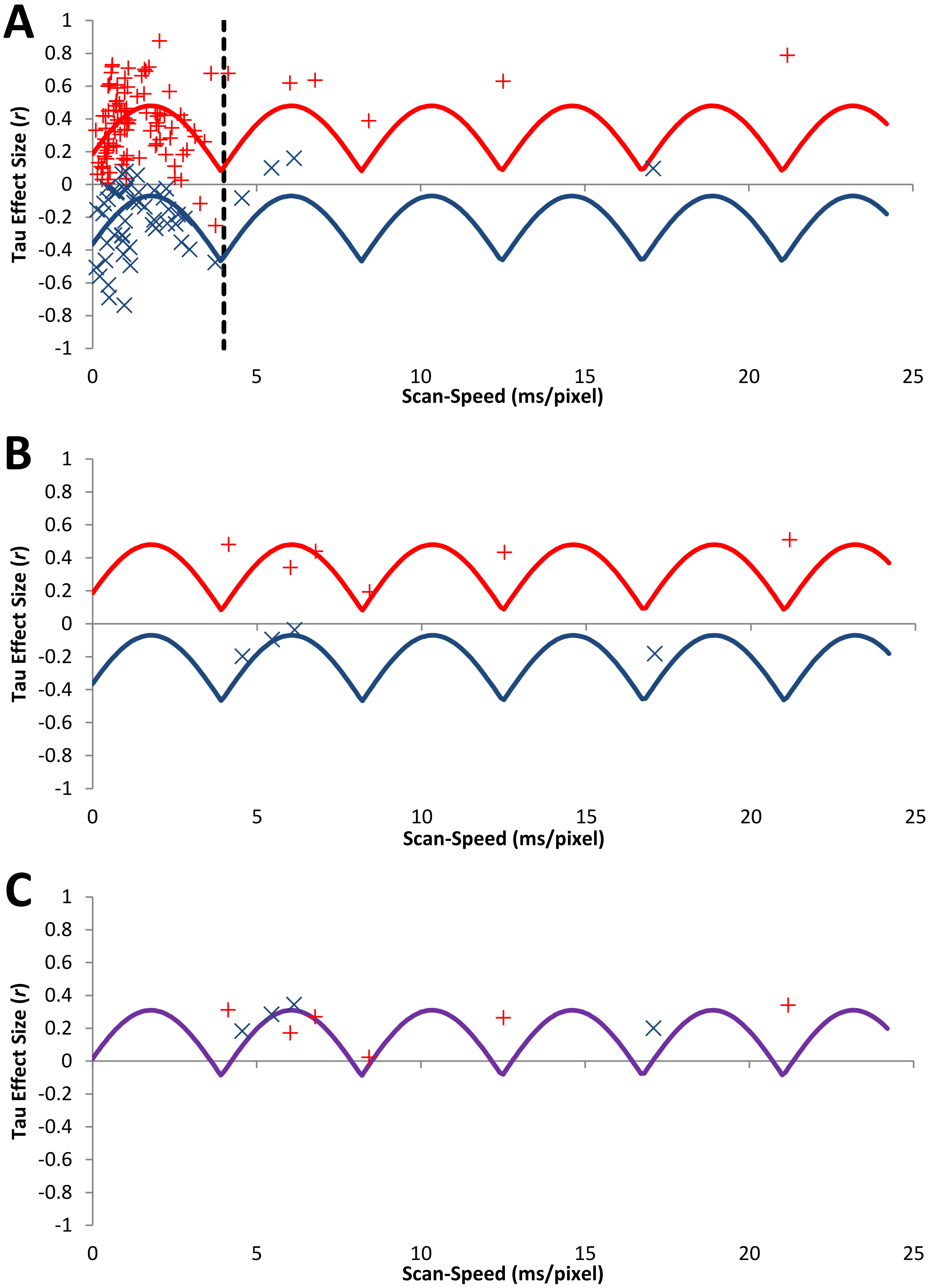

Figure 13.8. Tau effect and scan-speed values for participants who were deemed to be outliers. Panel A shows the outliers prior to group corrections. Panel B shows all outliers divided by positive and negative group identity. Panel C shows the data with corrections for group identity applied. All panels show participant's with positive tau effects as red +s and those with negative tau effects as blue Xs. The trend line for the positive tau effect is red, the negative tau effect blue, and the combined trend line is purple.

relationship, which implies the repetition of the expected pattern. This was modelled with a sine wave that was constrained to produce an absolute value[70]. The sine wave was produced by fitting a sine wave to the participants with scan-speeds below 4. The equation produced was then used to predict the tau effect sizes associated with the scan-speeds of the outlying participants. Outlying participants were free to fall on the sine wave that best fit their observed tau effect. For simplicity the data in Figure 13.8 is shown after corrections for stimulus size, group identity, and scan task inclusion order have been made. The mean was a better predictor of the variance than the model.

Model selection. In order to find the best model for the association between scan-speed and tau-effect-size the corrected Akaike Information Criterion (AICc) (Burnham & Anderson, 2002) was calculated for each candidate model. The AICc provides an objective measure of whether additional parameters in a model account for enough variance to warrant their inclusion. Models with lower AICc values are said to provide a more efficient fit than models with higher AICc values. As the AICc does not consider the absolute value of a number, the model with the value closest to negative infinity, rather than the one closer to zero, is preferred. Unlike the standard Akaike Information Criterion the AICc takes sample size into account and is a more conservative measure. The parameters (K) associated with each model are presented in Appendix E. The Akaike values, presented in Table 13.1, indicated that the increase in R^2 value associated with fitting separate lines to the positive and negative groups did not justify the added complexity in any case. Simpler quadratic models are also preferred over sine wave models when outliers are excluded[71].

Table 13.1
AICc values associated with different models

Correction	Model	Combination	RSS	R^2	K	AICc
Uncorrected	Quadratic	Combined	5.11	0.10	5	-470.33
		Separate	5.06	0.11	7	-467.44
	Sine	Combined	5.11	0.10	6	-468.15
		Separate	5.06	0.11	9	-462.92
Corrected	**Quadratic**	**Combined**	**4.72**	**0.24**	**7**	**-477.38***
		Separate	4.71	0.25	10	-456.04
	Sine	Combined	4.74	0.24	8	-474.52
		Separate	4.71	0.25	12	-466.12

Notes: In all cases *N*=143. The best model is shown in bold with an *.

[70] Absolute values were determined by the sine component of the model only. The constant then shifts the values away from zero. If this constraint is not used, pure sine wave models perform even worse.

[71] If outliers are included the corrected sine models are deemed better than the quadratic models but this is due solely to the influence of the outliers. Inclusion of kappa was not justified in any model. The best value produced by a model including kappa was -477.04 when added to the corrected, quadratic, combined model.

In absolute terms the AICc suggests the use of the corrected quadratic model over the uncorrected quadratic model. In relative terms the AICc suggests that the uncorrected quadratic model is 0.03 times as likely to be as good a choice as the corrected quadratic model.

As the AICc cannot indicate whether a model is significantly better than another, and as the comparative AICc values were so close to one another, the *F* change statistic was calculated in order to compare the best corrected model to the conceptually simpler uncorrected model. The resulting *F* statistic ($F_{(1,138)}$=11.40, p<.001) suggests that the corrected model is significantly better than the uncorrected model. This is in line with the higher R^2 for the corrected model, doubling with the inclusion of the extra parameter.

Standardised kappa and tau. In line with the analysis conducted in experiment 4 data was standardised by removing each participant's mean response and dividing the remaining value by a 1 unit change in the stimulus (e.g. 75 pixels in the uncorrected spatial task). This allowed the data to be expressed in standardised units. Data was then reconceptualised in terms of target and distracter dimensions rather than space or time in order to make meaningful comparisons and plotted in Figure 13.9. As this procedure essentially corrects the data for difference in stimulus scale, no division between corrected and uncorrected data occurs.

Preliminary analyses considered the role of stimulus scale and scan task inclusion order. No effects of, or interactions with, stimulus scale were found and it was excluded from further

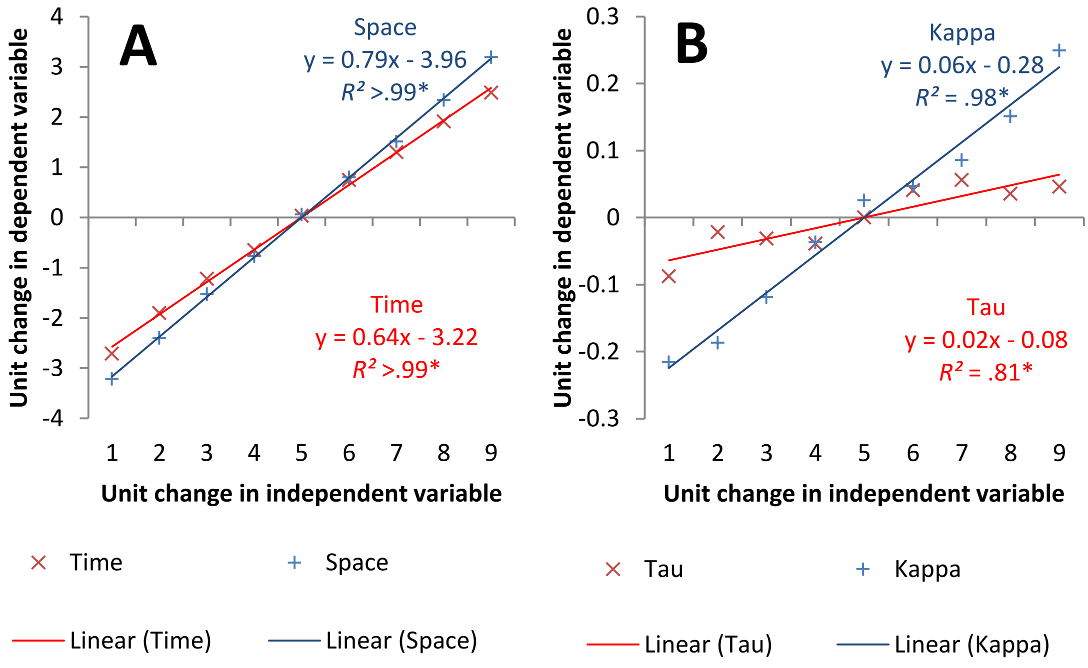

Figure 13.9. The relationship between standardised time and space (A) and standardised kappa and tau effects (B). Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

standardised analyses. Scan task inclusion order produced some expected effects and interactions but also lead to some expected effects occurring in higher order relationships (i.e. quadratic). As it failed to replicate some patterns observed in the unstandardised analyses and is the result of substantial data manipulation a simpler analysis, collapsed across inclusion order, was preferred. The results of the analyses that include scan task order can be found in Appendix F.

When collapsing across inclusion order and stimulus scale significant linear main effects were found for target condition number ($F_{(1,153)}$=5601.56 , p<.001) and distracter condition number ($F_{(1,153)}$=60.52, p<.001). Significant linear interactions were found for the interaction between stimulus type and target ($F_{(1,153)}$=90.62 , p<.001) and distracter ($F_{(1,153)}$=21.88 , p<.001) condition number. This suggests that the space effect is stronger than the time effect and the kappa effect is stronger than the tau effect. A significant linear interaction was also found between target and distracter condition number ($F_{(1,153)}$= 11.25, p<.001). This suggests that the target dimension in a task has a stronger impact on a participant's response than the distracter dimension. Put another way, space and time effects are stronger than kappa and tau effects.

A significant linear interaction was found between stimulus type, target condition number, and distracter condition number ($F_{(1,153)}$= 5.27, p=.023). This suggests that the difference between space and time is larger for the target condition than the distracter condition.

Testing the distribution of positive and negative tau effects. The distribution of scan-speed among participants with positive and negative tau effects is shown in Table 13.2 and Figure 13.10. For the purposes of this analysis group identity was determined by proximity to the quadratic lines of best fit. As a result all outliers were allocated to the positive group. Presence of a difference in the means of the distributions was tested with an independent samples *t*-test with equal variances assumed. No significant difference was found between the mean scan-speed of the negative participants (M=1.44, SE=0.15) and the positive participants (M=1.91, SE=0.27, $t_{(153)}$=1.36, p>.05). A chi-square test comparing the distribution shown in Table 13.2 also failed to produce a significant result ($\chi^2_{(8)}$=8.57, p>.05).

Table 13.2
Observed values and categories used in the Chi-square analysis of the distribution of scan-speeds for positive and negative participants.

	Scan-Speed									
	0-0.5	0.5-1	1-1.5	1.5-2	2-2.5	2.5-3	3-3.5	3.5-4	>4	total
Positive	22	28	15	12	9	7	3	1	10	107
Negative	10	12	8	5	4	5	1	3	0	48

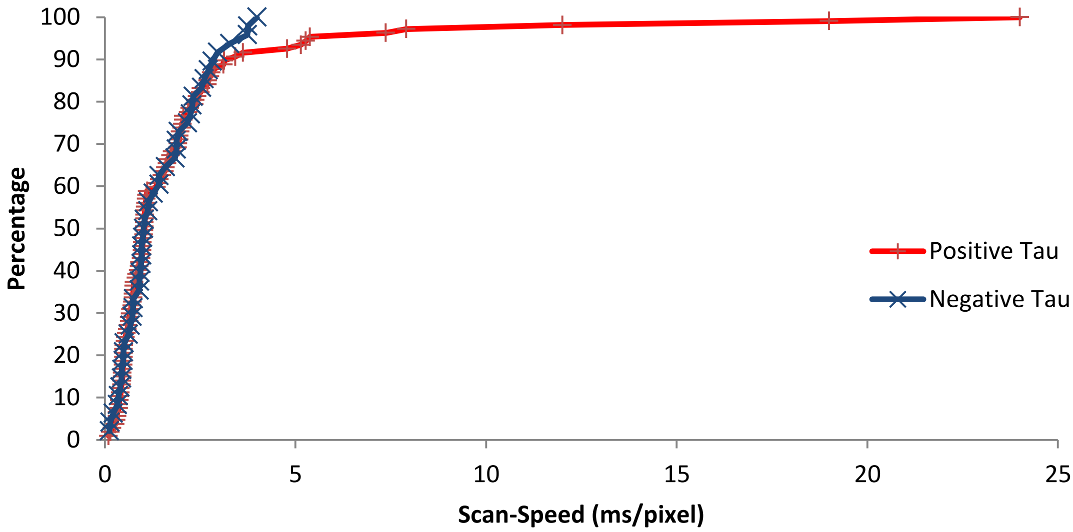

Figure 13.10. The cumulative distribution of scan-speed for negative and positive tau effects. Group identity is assigned by proximity to the quadratic trend lines.

Data from experiments 6 and 7, shown in Table 13.3, was used to test the distribution between positive and negative tau effects for participants who reported correcting for the tau effect and participants report noticing (having insight into) what the experiment was investigating (kappa and tau effects). No significant difference was found when both questions were considered simultaneously ($\chi^2_{(3)}$=4.29, *p*>.05) or separately (correction: $\chi^2_{(1)}$<0.01, *p*>.05; surprised: $\chi^2_{(1)}$=1.26, *p*>.05).

Table 13.3
Observed counts of participants reporting awareness of the likelihood of the experiment inducing tau and kappa effects, reporting correcting for such an effect, and those producing positive and negative tau effects.

	Insight			
	Yes		No	
	Corrected		Corrected	
Tau effect	Yes	No	Yes	No
Positive	23	32	2	11
Negative	9	18	2	1

Discussion

Summary of effects. For the most part the results of the meta-analysis confirmed the findings of the earlier experiments. As expected the overall results show time, space, scan, tau, kappa and scanning tau effects. The previously observed relationship between the tau effect and scan-speed was also observed and was shown to be best explained by a model that took the relative scale of the stimuli into account while using a common model for participants producing positive and negative tau effects. A similar association between scan-speed and the kappa effect could not be found. This discrepancy was mirrored in the presence of a significant difference between tau and kappa slope when a standardised measure was used.

Standard effects. The data was fairly reliable across experiments. The fact that interactions with stimulus size are found with the uncorrected data but not the corrected data gives some support to the use of the corrections. However, interactions between inclusion order and scan-speed, and inclusion order and tau effects remained in the corrected data.

In the case of the scan task, inclusion before the time and space trials is associated with longer scan times and a steeper slope. As such it could be a measure of participant fatigue, practice effects, or simply due to the scan task pre condition having the fewest participants, thereby providing the least reliable estimate. As these explanations and the observed difference in scan-speed is unimportant to wider interpretations they are treated as a curiosity.

The tau effect was found to be weaker when the scan task was included at the end of the experiment rather than before or during the space and time trials. The production of a weaker tau effect when the scan task is included at the end of the experiment is consistent with the predictions made in experiments 5 and 6. The implications of this finding will be discussed as appropriate when considering the relationship between tau effect size and scan-speed as well as when considering the results of Casasanto and Boroditsky (2008).

Tau-effect-size / Scan-speed relationship. The consistent finding of a quadratic relationship between the strength of the tau effect and scan-speed strengthens the idea that a relationship between these elements does exist. Furthermore, as the best model was found to be one that takes the stimulus size and scan task inclusion order into account while assuming that positive and negative participants are separated by a constant value, there is reasonable evidence to claim that this relationship generalises across all participants.

Group identity. The first element of such a model is the constant that describes the difference between the positive and negative groups. The steady shift of 0.56 is in line with a large effect size (Cohen, 1992) and is compatible with the notion of a deliberate correction occurring in one of the groups. However, positive or negative tau group identity was not associated with conscious awareness of the use of a correction, suggesting the term deliberate is inappropriate. Furthermore, it is currently difficult to conclusively assign one group to the role of corrector. However, as negative tau participants are less common and demonstrate an implausible negative tau effect it appears reasonable to assign them the role of the correctors. As such, unless otherwise specified, further discussion of the relationship between scan-speed and the tau effect will refer primarily to the positive tau group.

Inclusion order. The second element of the quadratic model to consider is the main effect of inclusion order. As shown in Figure 13.5, when the scan task is included earlier in the experiment tau effects can be expected to be more positive than when the scan task is included later in the experiment. On the surface this is compatible with predictions from experiment 5 and 6 as well as Casasanto and Boroditsky (2008). Specifically, if the scan task is included after the space and time tasks, imputed velocity, and therefore the tau effect, should be suppressed when using static line stimuli. Importantly though, the existence of a main effect of inclusion order, rather than an interaction, suggests that imputed motion and the tau effect are not completely suppressed. Rather, much like the correction proposed between groups, there is simply a downard shift in tau effect size strength as the scan task is inlcuded at later time points in the experiment. At an individual level then it is less correct to say that tau effects are suppressed toward an effect size of zero. It is more correct to say that the use of static lines in the absence of a scanning task produces a downward shift of the quadratic relationship between scan-speed and tau-effect-size.

Quadratic trend. Under the quadratic element of the model a given stimulus set is associated with at least two resonant scan-speeds and one optimum scan-speed. As outlined in experiment 4, the resonant scan-speed is the speed at which time and space do not interfere with one another. Consistent with accounts offered by Jones and Huang (1982) and Goldreich (2007), in these cases the temporal and spatial elements of the stimuli are supplying the same answer. In these cases a participant's scan-speed matches the imputed velocity derived from the stimulus set. The optimum scan-speed is where time and space produce conflicting information and interfere with each other most. The interference resulting from the inconsistency leads to the production of larger tau effects. The resonant and optimal scan-speeds for a given stimulus set are proportional to the combination of times and spaces used. The corrected data for the current dataset peaks at 1.54ms per pixel and loosely fits the empirically derived equations provided on the next page.

$$\text{Optimum scan-speed} = \text{Mean Time}_{(ms)}/\text{Mean Space}_{(\text{degrees of visual angle})} \times 4.68\text{x}10^{-3}$$

$$\text{Optimum Scan-Speed} = 5000/13.84 \times 4.68\text{x}10^{-3}$$

$$\text{Optimum Scan-Speed} = 1.69\text{ms per pixel}$$

$$\text{First Resonant Scan-Speed} = 0$$

$$\text{Second Resonant Scan-Speed} = \text{Optimum scan-speed x2}$$

$$\text{Second Resonant Scan-Speed} = \text{Mean Time}_{(ms)}/\text{Mean Space}_{(\text{degrees of visual angle})} \times 9.36 \text{ x}10^{-3}$$

$$\text{Second Resonant Scan-Speed} = 5000/13.84 \times 9.36 \text{ x}10^{-3}$$

$$\text{Second Resonant Scan-Speed} = 3.38\text{ms per pixel}$$

(13.1)

One simple way to think of this resonance relationship is to think of a car driving around a circular race track. Imagine that the driver will be given a penalty if they do not cross the start/finish line within a certain time window. The catch is that they must travel around the circuit at a constant speed. In this scenario, resonant speeds are those that minimise the number of penalties that the driver can accumulate in any particular multi-lap race. Such an example nicely demonstrates key points of the quadratic relationship: the high cost of imprecisely matching the resonant speed and the interdependence between space, time, and speed.

In the race car example selecting the resonant/correct speed ultimately means fewer penalties are collected. However if the speed is close to correct penalties may be avoided for the first few laps. In these cases, as more and more laps go by, the time the car arrives at the finish/start line drifts further from the time window in which penalties do not apply. As a result the cost of inaccuracy is high and the cost increases as proximity to the resonant scan-speed increases. Put more simply, a difference of half a speed unit from the resonant speed leads to a larger increase in penalties than a half unit movement away from resonance in a non-resonant speed. This is in line with the pattern produced by the quadratic trend: a clear lack of penalties at the resonant point with relative penalty values declining towards the optimum scan-speed.

The interdependence between space, time, and speed becomes apparent when identifying the resonant speed for different tracks. On a track 1 spatial unit long where the car can be penalised 1 point every 1 second the resonant speed is 1 unit per second. If the time is changed such that one penalty can be awarded every half second the resonant speed becomes 2 units per second. If the spatial distance is reduced to half, penalties can be minimised while travelling at 0.5 units per second. These predictions are in line with the results of experiment 7 and the combined analyses. Different sized stimuli, or different tracks, are associated with different resonant scan-speeds.

Finally, a speed account helps to explain why the first resonant scan-speed would be zero. Remembering that the measure of speed used here is milliseconds per pixel, a speed of 0ms per pixel is equivalent to saying participants are scanning infinitely fast. If a race car were able to travel infinitely fast the driver would always be crossing the start/finish line. As a result they would never be in a position to collect any penalties. As a result, in the case of the tau effect, travelling infinitely fast should be associated with the absence of a tau effect.

The proposed resonant property also implies that a harmonic relationship may exist. Any speed that is a whole number multiple of the resonant speed should produce the same result. For example if the resonant speed is 1 unit per second a car travelling at 2, 3, or more times that speed will still cross the finish line at the right time to avoid the penalty, it will simply complete more laps of the circuit between points. If a harmonic relationship exists it may have been reflected in the outliers. However, the outliers in the present dataset do not support the presence of a harmonic relationship.

Discrepancies and associations between kappa and tau effects. While the relationship between the tau effect and scan-speed supports the case for imputed velocity linking perceptions of time and space, the continued lack of a relationship between the kappa effect and scan-speed appears to undermine the relationship. However this need not be the case. While the dominant theory around tau and kappa effects suggests that they are both the product of imputed velocity (Jones & Huang, 1982) separate forms of imputed velocity for kappa and tau effects have been suggested (Jones & Huang, 1982) or implied (Casasanto & Boroditsky, 2008) in previous studies. Perhaps the lack of an association with the kappa effect suggests that the kappa effect operates off a different type of imputed velocity than the tau effect.

If the kappa and tau effect are the result of different types of imputed velocity then there is little reason to expect an association between kappa and tau effects to be found. Furthermore, if the respective forms of imputed velocity are separate, variance in the tau effect explained by the kappa effect should be separable to that explained by imputed velocity associated with the tau effect. Evidence against both of these claims has been found. First, the R^2 values for the association between the tau effect and scan-speed is similar to that between the tau effect and the kappa effect. Second, when the kappa effect is considered alongside the scan-speed model, the kappa effect only accounts for an additional 1% of variance in the tau effect. This suggests that most of the variance demonstrating a relationship between the kappa and the tau effect can be explained through imputed velocity. This is in line with the idea that a common form of imputed velocity produces the tau and kappa effect but contradicts the fact that there is no relationship between the kappa effect and scan-speed.

The failure to find a consistent story suggests one of two things. On one hand, one or more of the associations with the kappa effect is due to chance. Either the tau effect and kappa effect are not related or the kappa effect is related to scan-speed. Alternatively, both findings are correct and the variance the kappa effect explains is somehow redundant with that which scan-speed explains. As the current results do not have sufficient evidence in favour of either account, a conclusion will be left for future work.

Standardised results. In order to determine whether there was an asymmetrical relationship between time and space as suggested by Casasanto and Boroditsky (2008), the results were converted to standardised units. Unlike experiment 4, analysis of the combined results showed an interaction between stimulus classes in the distracter condition. At face value this suggests that space impacts on time more than time impacts on space. While such an interpretation is in line with the results of Casasanto and Boroditsky (2008) it is one that should be treated with caution.

Most importantly, while the process used should standardise measurements of time and space, an interaction between stimulus type in the target condition was also found. Specifically, increases in the spatial size of stimuli were associated with larger and more accurate increases in spatial estimates than comparable associations between the temporal durations of stimuli and a participant's time estimates.

Such a difference is not problematic for wider theory. In fact, it is in line with the idea that changes in time should be less salient than changes in space (Bottini & Casasanto, 2010; James, 1890; Riemer et al., 2012) and has common experimental support. In particular, space produces unit changes in line with what would be expected from a scaling constant of 1 (Stevens, 1957), whereas time, as described by Vierodt's law (longer intervals underestimated, shorter intervals overestimated) (Lejeune & Wearden, 2009), produces results consistent with a scaling factor of less than 1 (Eisler, 1981).

Where it does become problematic is in building a case for the symmetry or asymmetry of kappa and tau effects. While the tau effect was smaller than the kappa effect, the effect of time was also smaller than the effect of space in general. As such it becomes difficult to tell what led to the asymmetry between the kappa and tau effects. In line with Casasanto and Boroditsky (2008), it might be the case that time has less of an impact on space than vice versa. Alternatively, perhaps the tau effect is smaller than the kappa effect because changes in time are perceived as smaller than changes in space. Ultimately, the current results strengthen the argument for the need to equate time and space before making claims about one having more influence over the other.

Tau and tau_{scan}. One of the most problematic findings is the failure to find a relationship between the tau effect and the tau_{scan} effect. In each case participants are measuring space and in each case their estimates of space are being influenced by time. In essence the two tasks should be the same. Why then are different results being found?

The two tasks differ in how responses are produced. In the case of the tau effect, spatial responses are being produced by clicking two separate locations on the screen. In the case of the tau_{scan} effect, responses are produced by clicking a fixed location on the screen, waiting for a determined duration to pass, then clicking the same location again. While scanning has proven to be reliably related to spatial length here and elsewhere (Kosslyn et al., 1978), this does not mean that scanning and standard spatial estimation tasks are providing the same information at a qualitative level. In particular, in terms of production, the scan task is more like the time task than the space task. As a time task cannot show a tau effect, perhaps this is why no relationship exists between the tau and tau_{scan} effects. Perhaps the tau_{scan} effect is not like a tau effect at all.

Reasons to expect a qualitative difference to occur in the production of time and space exist. In the time task subjects have no control over the pacing of their response. They must react to time as it passes. Furthermore, time always passes unidirectionally: participants always start with 0 time and can never reduce their estimate. In the space task subjects are free to produce their estimate at their own pace and can move bidirectionally: an overly long initial movement can be corrected by moving the mouse back before clicking. By converting the space task to a time task the limitations imposed in the time task are imposed on the spatial response. Imposing such restrictions has been shown to change the nature of responses to pitch and brightness (Riemer et al., 2012). When bidirectional movements in these dimensions were not possible (estimates could not be corrected) the quality of the responses changed: response became more time-like and a bias similar to Vierodt's law was imposed. The scan task creates the same limitations in its estimates of spatial lengths.

An alternative would be to consider the proposed role of scanning as an index of imputed velocity. If scan-speed is accepted as a fair index of imputed velocity then scan-speed is a mediating factor in the production of the tau effect. In the equation detailing the relationship provided by Jones and Huang (1982), equation 8.1 re-presented below from page 100 for clarity, this would make scan-speed V, time t, and space s.

$$V= s_1/t_1 = s_2/t_2$$

In the case of a standard tau effect s_2 is influenced by t_1 through V. In the case of a tau_{scan} effect the implication is that s_2 is converted to a time estimate of s_2, $t_{s2,}$ through a scan-speed approximating V. The tau_{scan} effect occurs when t_{s2} is influenced by t_1 through V. This highlights an issue with the expectation of a relationship between the tau and tau_{scan} effects: velocity is argued to influence a measure that velocity itself has created. This is equivalent to suggesting that time produces the tau effect or space produces the kappa effect. Interestingly the conclusion here is the same as the previous paragraph. As a time task cannot show a tau effect perhaps this is why no relationship exists.

An alternative is to focus less on the tau_{scan} effect and more on the scan task itself. The scan task is assumed to be the result of the conversion of spatial information into time information. Where the tau effect is an unintentional leak of information from a temporal representation to a spatial representation, the tau_{scan} effect is associated with a deliberate push of information from a spatial representation to a temporal representation. This can be considered to be like the difference between a tap that is leaking, the unintentional tau effect, and a tap that is turned on, the deliberate scanning task. Importantly, there is no reason to believe that the leak rate and the maximum flow rate of a metaphorical or actual tap would be related.

Such distinctions between intentional and unintentional processes have been found before. One example comes from a word suppression task (Debner & Jacoby, 1994). Participants were shown a word on screen either above or below awareness. Next they were shown a word stem. Participants were asked to complete the word stem without using the word that had just been presented to them. Presented words were successfully suppressed more frequently when presented above awareness than when presented below awareness. Put another way, when presented above awareness the word-specific tap could be intentionally turned off. When words were presented below awareness the information leaked through unintentionally.

Taking the tap metaphor further, in order for the scan task's push from a spatial representation to a temporal representation to produce accurate results the existing temporal information in the time 'sink' must be emptied out and all of the spatial information must be pumped in. Under this push account, tau_{scan} errors could come from one of two places. The push, whose direction would imply an association with the kappa effect, or the emptying of the time sink. As the association between the tau_{scan} effect and the kappa effect is as poor as that between the tau_{scan} effect and the tau effect, this suggests that the emptying of the time sink may be the candidate.

Crucially, a sink based account of the tau_{scan} effect is compatible with positive and negative tau_{scan} effects. Failure to empty pre-existing temporal information from the sink would result in a positive tau_{scan} effect. More time in the sink at the beginning would mean more time in the sink at the end. Conversely, emptying the sink too much could result in a negative tau_{scan} effect. In order for this to occur the emptying would need to occur, or continue to occur, after the push began. This is equivalent to removing the plug while/after filling the sink.

Such removal of the plug is compatible with a correction similar to, or the same as, that proposed for the tau effect. Such similarity suggests that this correction process is likely to occur separately to the accumulator and may be common across tasks. Assuming that the correction mechanism is the same between tasks the failure to find a clear relationship between the negativity/positivity of the tau effect and the tau_{scan} effect would suggest that use or non-use of the correction is task specific. At some level a choice is made to insert or remove the plug according to task demands.

Theoretical Implications

***Discrepancies with Casasanto and Boroditsky* (*2008*).** While tau effects have been found in half of the experiments in the current design they were absent in all of the Casasanto and Boroditsky (2008) experiments. This suggests there may have been something in the experiment design itself that lead to the suppression of tau effects. Casasanto and Boroditsky (2008) suggest that the way that their stimuli were presented, either as static lines or as dots of controlled speed, prevented imputed velocity and thereby prevented a tau effect.

The current results do provide some support for this statement. Late scan task inclusion order, equivalent to the design used by Casasanto and Boroditsky (2008), resulted in the absence of a group tau effect. However inspection of the scan-speed tau-effect-size relationship does not produce the pattern of results that suppression would predict. Rather than tau-effect-sizes being closer to 0, the result of a flatter quadratic term, observed tau effect sizes are shifted dwon, the result of a quadratic term being shifted down but not changing shape. Instead of suppressing a tau effect the use of static lines appears to encourage the production of greater negativity in tau effects.

A tendency to produce tau effects of greater negativity can explain the failure to find a positive tau effect at the group level. Most obviously, a constant negative shift in a generally positive effect will lead to a reduction in the size of the effect. Less obviously, such a shift alters the balance between positive and negative effect sizes. In addition to postive effects becoming weaker negative effects become stronger.

While a tendency to produce effects of greater negativity can explain the Casasanto and Boroditsky (2008) results it is not the only design element worth considering. The finding that stimulus size contributes to the scale of the scan-speed and tau effect size relationship suggests another possible explanation for the discrepancies between the current results and those of Casasanto and Boroditsky (2008). Specifically, it is possible that Casasanto and Boroditsky (2008) inadvertently selected a stimulus set that would lead to a reduced or negative tau effect.

In equation 13.1, presented on page 162, shorter times are associated with larger optimal scan-speeds. As shown in Figure 13.10 and Table 13.2, most participants produce scan-speeds of 3ms per pixel or less and, as shown in Table 13.2, approximately one third of participants apply a correction, leading to a negative tau effect, regardless of their scan-speed. Assuming that the

Casasanto and Boroditsky (2008) spatial stimuli occupy the same visual angle as those used in the current experiments[72] their stimuli would be associated with a shift of 1/0.6 placing the peak of the relationship at a speed of approximately 3.33 ms per pixel. As most participants scan quickly, this will lead to a large number of participants producing weakly positive or strongly negative tau effects. This will lead to an imbalance between the average size of a negative tau effect that opposes the imbalance in the number of positive vs. Negative effects. In addition, not having a scan task before or during the time and space trials leads to a further shift toward -1. Essentially, despite there being fewer negative participants these participants are expected to produce stronger negative effects that will be more likely to cancel out the weaker positive effects.

Critically, unlike the scan task inclusion order explanation, the stimulus size explanation simplifies the account across multiple studies. In particular, it helps to explain why tau effects are more volatile at lower time ranges: sometimes producing negative effects, sometimes positive ones, sometimes no effect at all (Jones & Huang, 1982).

Wider theoretical implications

The association between time and space. The pattern of resonance observed in the association between scan-speed and tau effect size suggests that there is interplay between the dimensions of space and time that is specific to each stimulus set. Whether this interplay is due to a shared representation for time and space, confusion over which dimension of a stimulus to respond to, or the contribution of an intermediary such as imputed velocity is unclear. At the moment the only thing that can be said for sure is that an association between scan-speed, treated as a measure of imputed velocity, and the tau effect exists.

Jones and Huang (1982) vs. Sarrazin et al. (2004). The association between scan-speed and the tau effect is more in line with the suggestions of Jones and Huang (1982) than the findings in the third experiment of Sarrazin et al. (2004). That being said Sarrazin et al. (2004) were unable to explain the findings of their third experiment and felt that the imputed velocity account championed by Jones and Huang (1982) did predict and account well for their results in the other conditions. As Sarrazin et al. (2004) did not offer a competing interpretation for their results a resolution is not immediately forthcoming.

[72] No information on screen size, distance from the screen, or visual angle is supplied in the Casasanto and Boroditsky (2008) article. Given that the current experiments use small screens positioned toward the far end of the recommended ergonomic viewing distance it is likely that the screens in the original experiment were the same size or larger and at a similar distance or closer. This should mean that their stimuli should appear to be the same size or larger than those in the current experiment. Larger stimuli have the same effect as shorter durations, in effect amplifying the proposed shift in the relationship between scan-speed and the tau effect making the proposed explanation more likely rather than less likely.

In attempting to reconcile their results with that of the current experiment there are two clear options. The first is to focus on the differences betweeen the studies, particularly the use of 7 combined stimuli per trial in the Sarrazin et al. (2004) study. Given that experiment 3 in Sarrazin et al. (2004) was the only experiment where both dimensions, space and time, varied on every trial a few things may have occured that lead to a null finding. The first is that the pairings somehow prevented the formation of kappa and tau effects by disrupting resonance between the dimensions to such an extent that the relationship between them brokedown, enhanced it to a point that kappa and tau effects disappeared, or changed it in such a way that positive and negative kappa and tau effects cancelled each other out in the results.

A simpler explanation would be that the trials in the third Sarrazin et al. (2004) experiment simply required participants to remember too much information. Keeping in mind that the limit of general working memory has been reported as 7+/- 2 items (G. A. Miller, 1956) and has been suggested to be as low as 4 items in spatial working memory (Baddeley, 2003) this seems quite likely. In most of the Sarrazin et al. (2004) experiments participants would be shown 7 time-space pairs and only need to hold 2 or 8 pieces of information. In the case where all the times and all the spaces were the same (CC) they would only need to remember 2 pieces of information per trial, the time and the space as appropriate. In the cases where only one dimension varied (CV, VC) they would need to remember 8 pieces of information, that related to whichever aspect of the stimulus was varying and the single piece of non-varying information. In experiment 3, where both elements changed (VV_{exp3}), they needed to remember 14 pieces of information and this information needed to be bound together. In line with such an explanation this condition was the least accurate and had the highest standard deviation score. Furthermore, unlike other conditions, performance on this task did not improve throughout the experiment. Support comes from a finding that integral dimensions, a classification that has been shown to apply to space and time (Dutta & Nairne, 1993), are not represented independently in working memory (Fougnie & Alvarez, 2011). Unlike separable dimensions integral dimensions failed in a co-dependent rather than independent manner and were associated with poorer performance. When this is applied to the Sarrazin et al. (2004) design the high load on working memory virtually guarantees some information will be lost. Exacerbating the problem the suggestion is now that if one element of the space-time pairing is lost so too is the other.

While working memory can explain why the VV_{exp3} condition in Sarrazin et al. (2004) is associated with poor performance it cannot explain why the VV condition is not. In both cases there are 14 stimuli to remember: 7 times and 7 spaces. The simplest explanation is that imputed velocity is maintained in VV but not VV_{exp3}. This essentially means that successful performance in VV can be attained by remembering variation in one dimension only. This reduces the load from 14 items to 7 meaning the task is no more difficult than the VC or CV conditions.

Reverse Tau Effects and the Suppression of Tau effects. The greatest contribution of the current results is the provision of an explanation for why tau effects do not have a consistent direction in the literature. Until this point the existence of tau effects, reverse tau effects, and null tau effects at the sample and individual level were problematic for the idea of a tau effect existing at all. After all, if sometimes the effect is positive, other times negative, and sometimes non-existent the field may be the result of type 1 error combined with a bias towards publishing significant results.

Until now the most notable and accepted account of negative tau effects was provided by Collyer (1977). Where Collyer (1977) suggested that negative tau effects were the result of an inversion of the underlying relationship between space and time, the current results have shown that they are due to an additional correction. Where the current results are best described by a downward shift Collyer's (1977) account predicts a mirror of the positive participant results in the negative participants.

In line with the explanation for the difference between the results of the current experiment and that of Casasanto and Boroditsky (2008) it is notable that reverse and null tau effects occur most frequently at lower time ranges (Jones & Huang, 1982). As outlined earlier this should bias tau effect production toward larger negative effects and smaller positive effects. As a result a mixture of positive, null, and negative group effects becomes likely.

Directions for future research.

Malleable vs. fixed scan-speed preference. While the current line of experiments have shown an association between scan-speed and the size of the tau effect it is not known whether scan-speeds or tau effect sizes are fixed for individual participants. If tau effect sizes are fixed for individual participants this would suggest that an individual's imputed velocity is malleable across experiments. If speeds are fixed, this would suggest that stimuli of different sizes would elicit different tau effects in different participants. Finally if both are able to change this may suggest that they are both susceptible to variations in external stimuli.

Importantly the current interpretation of the association with the Casasanto and Boroditsky (2008) study would imply that participants have a relatively fixed scan-speed. If participant's scan-speeds were malleable and adjusted to the stimulus set they were presented with, there is little reason to expect the Casasanto and Boroditsky (2008) study to produce a smaller tau effect than the current experiments. Support for this set preference also comes from the fact that kappa and tau effects become less reliable at lower time ranges. If participant's scan-speeds were malleable this lack of reliability should not occur.

Positive vs. Negative group identity. Although there appears to be two distinct groups of participants, those showing positive tau effects and those showing negative tau effects, group membership for participants has not been able to be predictable apriori. Results from the current study only suggest that participants are not consciously aware of their group identity, suggesting the proposed correction is not a conscious strategy. Due to the nature of the error-space, group identity is one of the largest predictors of the variance in tau effect size. If additional candidate explanations for group membership were to be explored they could add a lot to the understanding of the relationship between scan-speed and the tau effect.

The first step in such an investigation might be to check whether participants remain in the positive or negative group over multiple experiments. If they do, this may suggest that group identity is something that is relatively stable and may even be participant specific. If it is found to vary this will suggest that it could be something related to the task itself.

One potential candidate for shifting participants between groups comes from Yates, Loetscher, and Nicholls (2012). In this experiment a small change in task design and response requirements led to a reversal of the kappa effect. When participants were asked to judge which of two stimuli lasted longer a positive kappa effect was observed. When participants were asked to judge whether two stimuli were the same or different, a negative kappa effect was observed. However this was not the only change. In the second task a condition in which the stimuli had the same duration was also added. As such it is difficult to know which, if either, change led to the reversal in the kappa effect. However such a clear reversal with such a small change does suggest a promising direction for future work.

Investigating harmonic relationships by increasing outlier observations. Although the outliers in the current experiment failed to provide evidence supporting a harmonic relationship with the resonant scan-speed this does not mean that no such relationship exists. While simply recruiting more participants should produce more outliers such a strategy would require prohibitively large numbers of participants.

One option would be to alter the scan-speed participants produced. As outliers preferred to imagine the speck travelling more slowly than most participants it would be desirable to find a way to slow scan-speeds down. If participants are found to dial-in their scan-speed to suit the stimulus set it may be possible to alter a participant's scan-speed by training them on a separate set of stimuli to entrain a slower scan-speed. If participants are found to have a set scan-speed alterations to clock tick speed, through the use of experimental manipulations such as click trains, may provide a way to alter the association between speed, time, and space within individuals. Finally, if scan-speed sets a participant's imputed motion rather than simply indexing it, instructing participants to scan at slower speeds may produce the desired effect.

An alternative would be to adjust the resonance point of the stimulus set. Experiment 7 and the combined analyses showed that using smaller line lengths, or longer durations, leads to the compression of the quadratic term. While this will not increase the number of participants producing scan-speeds higher than 4, it should be possible to increase the number of participants producing a scan-speed that is larger than the second resonance point.

Subsecond timing. The current experiment focussed on suprasecond times. While suprasecond times are more consistently associated with structures involved in spatial processing subsecond times should also be considered. As previous experiments have shown the presence of kappa and tau effects with subsecond stimuli (Bill & Teft, 1972) the expectation would be that the quadratic pattern observed here would be replicated.

If the pattern of results were to be replicated with subsecond times the next step would be to determine which timing system was being used. While the obvious answer is the subsecond system this may not be a left lateralised network. As has been established elsewhere (Lewis & Miall, 2003b) the best way to establish which system is being used is to consider the requirements of the task at hand. It is possible that this experiment could utilise the right lateralised explicit/cognitive system even at shorter durations.

The results of experiment 7, and the finding that the scaling factor is beneficial in the overall results, also have practical implications for the use of subsecond times. If the same pattern of results is to be observed the current data suggests that the size of the lines themselves will also need to be reduced. If this does not occur the curve may be stretched so as to appear as a straight line. As a result if future work investigating the relationship between scan-speed and the tau effect finds a significant linear effect this should not be taken as evidence of a difference in the association at sub and supra second levels. Instead it should first be interpreted as a need to try another version of the experiment with smaller spatial stimuli.

Associations with the kappa effect. The most problematic finding of the current work in light of the imputed velocity account is the failure to find an association between the kappa effect and scan-speed. As the current experiment used a large sample size the issue is unlikely to be related to a lack of power at an experiment level. If a lack of power does exist it must be at the participant level through excessive variance in the measurements. While it is possible that this is the case, as each trial condition was only presented once per participant, the most likely explanation is that the scanning task cannot measure a form of imputed velocity that is relevant to the kappa effect.

The fact that the scan task shows an association with the tau effect may be due to the fact that they essentially involve the same task. In the scan task participants replicate space using time. Similarly, in producing a tau effect participants mistakenly incorporate time information into a spatial response. What may be required is an inverse scanning task where participants respond to time using space. In this case space will be considered in a spatial response.[73]

One further argument in favour of such an explanation is effect scale. While the scan task converted spatial estimates to time estimates the speed produced by this process would be problematic if it were used to convert time estimates to spatial estimates. As the screen was 1024 pixels wide, and the longest duration used was 7000ms, if an inverse scan task were to be used participants would need to scan at a speed of at least 6.84 pixels per millisecond or 1/6.84ms per pixel in order to accurately convert time estimates to on screen spatial estimates. Under this criterion only 2 of the current participants would be included in the analysis. All other participants would be producing line lengths that extend beyond the limits of the screen. This suggests that the failure to find an association between scan-speed and the kappa effect may be a limitation of the scan task.

Paradigm Optimisation. As touched on in the previous section the current paradigm has room for improvement. The current 243 trial design was inherited from the Casasanto and Boroditsky (2008) study and was carried through for the sake of consistency. Unlike other kappa-tau studies the number of trials in each condition was relatively low. As an example, Sarrazin et al. (2004) had 60 trials per condition whereas 9 trials per condition were used here. This is likely to have increased variability in the current results. While it has been suggested that this could have obscured potential associations with the kappa effect it may have also dampened associations with the tau effect. Future experiments may benefit from using fewer conditions thereby achieving more trials in the same amount of time.

Sinks, taps, and leaks. While the failure to find a relationship between the tau effect and the tau_{scan} effect was surprising it did lead to some interesting speculation about the way in which representations of time and space may impact on each other under different task types. The notion of a difference between an unintentional leak of information from one form of magnitude and a deliberate push of information from one magnitude representation to another is worth investigating

[73] Such a task was piloted as an optional extension to experiment 6. Evidence was inconclusive due to low N.

further. If such a difference can be established in other task types it would help to support the current explanation. If it failed it would suggest that there is something wrong with the current account. The most obvious candidate would be the inverse scanning task mentioned earlier. If such a task were created and showed a relationship between speed and the kappa effect but not between the kappa effect and a $\text{kappa}_{\text{inverse scan}}$ effect that would suggest that the tap vs. leak relationship could be common across representations of magnitude. Furthermore if the tau_{scan} and $\text{kappa}_{\text{inverse scan}}$ effects were shown to have a relationship similar to that of tau and kappa effects this would further strengthen the model.

Conclusion

This series of experiments supports the idea of interplay between representations of prospective time and space in the human mind through the production of kappa and tau effects. This has been built on by the inclusion of a scanning task and the finding of an association between scan-speed and the tau effect. While the basic shape of the relationship appears consistent, and has been shown to be related to the stimuli presented, a number of questions remain. These include whether scan-speed is task or participant specific, what contributes to participants correcting for the tau effect, what the outlier data means and whether more can be produced, whether similar associations can be found with the kappa effect, how the current paradigm can be improved, whether the tau and tau_{scan} effect should be related, and whether there is a difference between information pushed through imputed velocity and that which leaks through. Yet, while the current experiments have opened up a number of areas for future research they have also united previously disparate findings in existing research. The principle contribution in this regard is the illustration of the nature of negative tau effects and the distribution of negative and positive effects.

Chapter 14. Overall Discussion

While time and its perception may seem simple at first, these and other experiments have shown otherwise. The results of the current experiments point towards expected and unexpected refinements to the understanding of the relationship between the concepts of space and time. In addition, they offer up some useful adjustments to commonly used measures of signal that will be useful in this area of research and others. Ultimately, while they leave the big picture largely unchanged, they have contributed to the understanding of how things work.

Retrospective

As a whole, the message from the retrospective studies is not immediately clear. The return of a null result is notoriously difficult, if not downright dangerous, to interpret. On the face of it though, and with due caution, these results do not suggest that time and space are equivalent in memory. In some respects this makes sense. The current implied division of what, where, and when information in episodic and episodic-like memory (Tulving, 2002) certainly points to a separation between temporal and spatial information. Furthermore retrospective timing has been referred to as timing without a timer (Ivry & Hazeltine, 1992).

While more work on retrospective timing is required, retrospective timing work may benefit from a more specific model to test from. The current models of retrospective time simply propose that remembered time increases with increases in the complexity/and or number of the events to be remembered (Block & Zakay, 1997; Ornstein, 1969). While the current results have not produced significant findings, and retrospective time research typically is not particularly informative, other research into memory may be able to contribute to understandings of the representation of the time of events in memory. In particular, work suggesting that memory is more of a reconstruction of events than a replay of events (Addis & Schacter, 2008; Addis et al., 2007; Hassabis & Maguire, 2007) is worth considering.

Reconstruction accounts of memory are able to account for changes that occur in memories over time and have been argued to be more evolutionarily beneficial than replay models of memory by allowing for the simulation of potential future events through using knowledge from past events (Addis et al., 2007; Suddendorf & Busby, 2005). Under such an account retrospective timing would be the product of a reconstruction of an event rather than a replay of an event. As such, rather than being based off an accurate representation of the to-be-remembered event, retrospective time judgements would be based off knowledge about how aspects of the event typically play out. Importantly, such an account would explain the current results and previous patterns in the findings of retrospective timing while making specific testable predictions.

In considering the current results this account would likely suggest that time was not able to act as an effective memory cue. This is because the temporal aspect of each trial appeared to be the same and/or generally was not noticed by participants. As such, when remembering the event this aspect would not differ from one reconstruction to the next. Rather than remembering a fixation cross with a specific duration, participants would remember a fixation cross with a generic duration. Tentative support for this idea comes from the fact that, when asked, participants did not report noticing any variation in the duration of the fixation crosses or wrote the variation off as an issue with the computer.

In explaining past findings, specifically more complex events being perceived as having taken longer (Block & Zakay, 1997), the reconstruction account would argue that more complex memories have more components. Each component would be associated with a specific duration presumably leading to an additive effect overall. This is also in line with the idea of retrospective timing being timing without a timer (Ivry & Hazeltine, 1992). Under such an account a generic time linked with a given stimulus is imposed on the reconstruction rather than a time specific to that occurrence.

The predictions the model makes are where things truly become interesting. Importantly, under such an account, participants should respond as if the reconstructed interval is the real interval even if the reconstruction is based on false information. In particular, if participants were trained to perceive a given distracter as taking a certain amount of time during a training period only to have this distracter last a different amount of time during the actual experiment, this could be manipulated to impact on retrospective judgements. Specifically, participants trained to think the distracter lasted longer should produce longer durations than those trained to the think the distracter was shorter, even when retrospectively timing events of equal duration. This would expand on the current understanding of retrospective timing by showing that events of equal number and complexity can produce different reported durations.

Applying this to the case of the directed forgetting study, training with other cues to indicate time may be beneficial. For example, short durations could be paired with a red fixation cross and long durations could be paired with a blue fixation cross. These colour cues could then be removed at the test phase to ensure that any effect was due to the paired duration rather than the paired stimulus. In this example this would ensure that temporal information rather than colour information was leading to any effect observed.

However it is also possible that the directed forgetting paradigm could produce evidence in favour of this model without introducing new stimuli. The proposed model implies that, in the directed forgetting task, time would not be attended to at study or test. As a result it may be encoded as a standard or average interval. This would suggest that at the test phase an effect linked

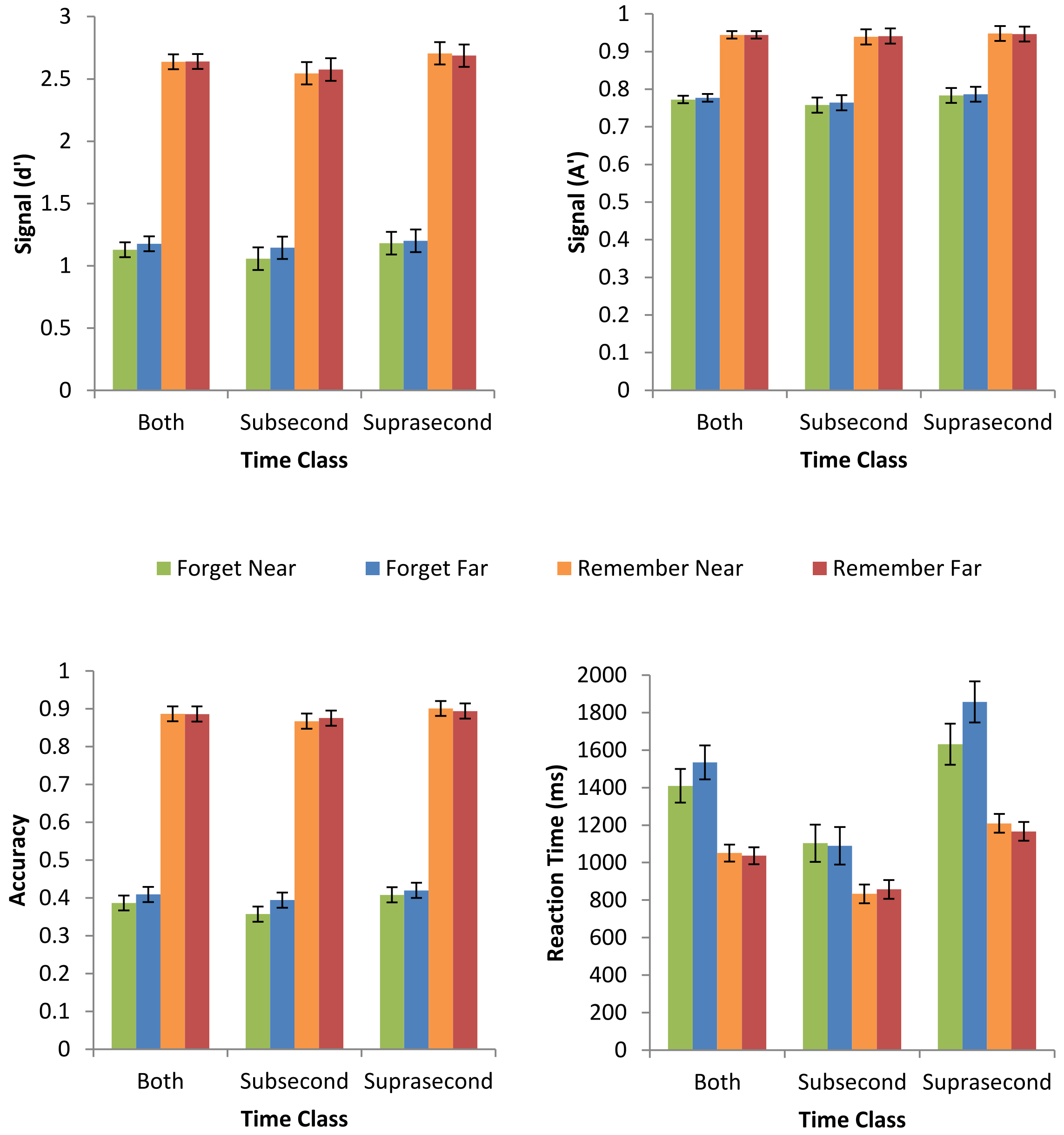

Figure 14.1. Data from the time based directed forgetting experiments reclassified into test time close to and far from the average duration at study. Within subjects error bars (Masson & Loftus, 2003) represent the planned contrasts between temporal locations within memory cue conditions e.g. near vs. far temporal locations for remember cued words.

to the average fixation cross duration at study rather than the specific fixation cross duration at study may emerge. When the data is re-categorised[74] into test phase fixation cross durations that

[74] These values and tests relate to data where the inclusion criterion is used. If the inclusion criterion is not used the pattern is largely the same except a main effect of nearness is observed in the suprasecond time range. This is consistent with the argument that participants who do not show a significant difference

are near the average and durations that are far from the average, as shown in Figure 14.1, a significant interaction involving nearness is observed for the reaction time data. However the interaction is only significant in the suprasecond condition ($F_{(1,58)}$=4.67, p=.035) and is not significant if all of the data is entered together (although a 3 way interaction involving time class, memory condition, and nearness does approach significance ($F_{(1,99)}$=3.65, p=.059)). Planned contrasts suggest that the interaction is the result of a borderline significant difference between near and far in the forget words ($F_{(1,58)}$=3.68, p=.060) but not the remember words ($F_{(1,58)}$=1.56, p>.05). This suggests that there is a benefit for matching the average fixation time at study with the fixation time at test.

The combination of significant findings in the suprasecond condition in this re-analysis and the original analysis in experiment 2 suggest that there may be merit to continuing to refine the memory paradigm. This especially true when considering its key benefit over standard retrospective timing tasks: the ability to include multiple intervals in a single experiment for a single participant without drawing attention to time.

Prospective Time – Tau and Kappa Effects

In contrast the relationship between space and time in prospective time appears well documented and well researched. In this case it was of interest to investigate the nature of the relationship between time and space, essentially presupposing its existence. Of particular interest was a set of recent findings (Casasanto & Boroditsky, 2008) which slightly contradicted previous research with the relatively simple and well established kappa and tau effects. The current experiments reinforced the imputed velocity account of Jones and Huang (1982) by identifying potential explanations for the contradictions produced in Casasanto and Boroditsky (2008).

This reliance on imputed velocity suggests that the representations and/or processing of time and space probably initially exist separately. Such a notion is reinforced by the idea that the VV_{exp3} condition in Sarrazin et al. (2004) could be due to surpassing the limit for working memory. Importantly, this limit is only surpassed if time and space have separate, rather than shared, representations.

In terms of the time as space theories, time can simply be thought of as an additional dimension, the fourth dimension. Under such an account information from 1 dimension would leak to the other dimension; something that has been demonstrated in a Garner interference task (Dutta

---

between conditions behave as if the forget condition is the remember condition. This is also compatible with the argument that participants in the Hourihan et al. (2007) study show an effect in the forget words due to forget words being contaminated with remember word performance.

None of these effects are significant if the data is presented according to the time at test rather than the near-far categorisation.

& Nairne, 1993). This is compatible with the notion of a common spatial metric or a more generic notion of magnitude proposed under ATOM. In these cases imputed velocity, or more precisely the difference between the imputed velocity in the stimulus set and the participant's ideal imputed velocity, establishes the strength of the leak.

Based on work investigating areas activated by prospective timing, it seems logical to suggest that kappa and tau effects are produced by the cognitive timing system. This system shares a similar network to spatial working memory and is commonly activated in tasks that require time to be attended to and/or require the timing of longer intervals (Coull & Nobre, 2008; Lewis & Miall, 2003b). This provides a common substrate for the representation of space and time and means for leaks to occur from one dimension to another.

Implications for a Unified Theory

Scalar timing theory. The suggestion then is that temporal working memory *is* spatial working memory. This allows for a modification to the STT depicted in Figure 14.2; one that merges it with ATOM or space as time models. Working memory can be refined to the visuo-spatial sketchpad. Leak-rate can be included and indexed by the relationship between the degree of resonance between the imputed velocity inherent in the stimuli and the expected imputed velocity, currently measured by scan-speed, in each participant. Where there is a match between the resonant speed of the stimuli and the imputed velocity subjects produce, leaks from time to space should be reduced; when there is a mismatch, leaks from time to space should be relatively higher. In the case of kappa and tau effects, and perhaps other space, time, and (theoretically) number associations, an additional correction may then be applied before a response is finally produced.

Importantly this refined model can also inform retrospective timing tasks. A retrospective timing task should activate the self-projection network (Buckner & Carroll, 2006), drawing on knowledge of similar past events to create a simulated reconstruction of the target event (Hassabis & Maguire, 2007). It should be expected that the simulation will run on previous experience. As a result the target intervals should be free of kappa or tau like effects unless they had been consistently paired with stimuli of a particular size in the past. Retrospective tasks involving reproduction of timed intervals should then activate the prospective timing model during production of the task, providing an opportunity for current kappa or tau like effects to occur. Alternatively, those involving verbal report, the response method typically employed in retrospective timing tasks, should not produce such an effect.

Finally, the refined model would be useful in explaining the episodic time division proposed in the introduction as well as a recent study investigating time based prospective memory (Waldum & Sahakyan, 2013). Time based prospective memory describes the ability to remember to perform a

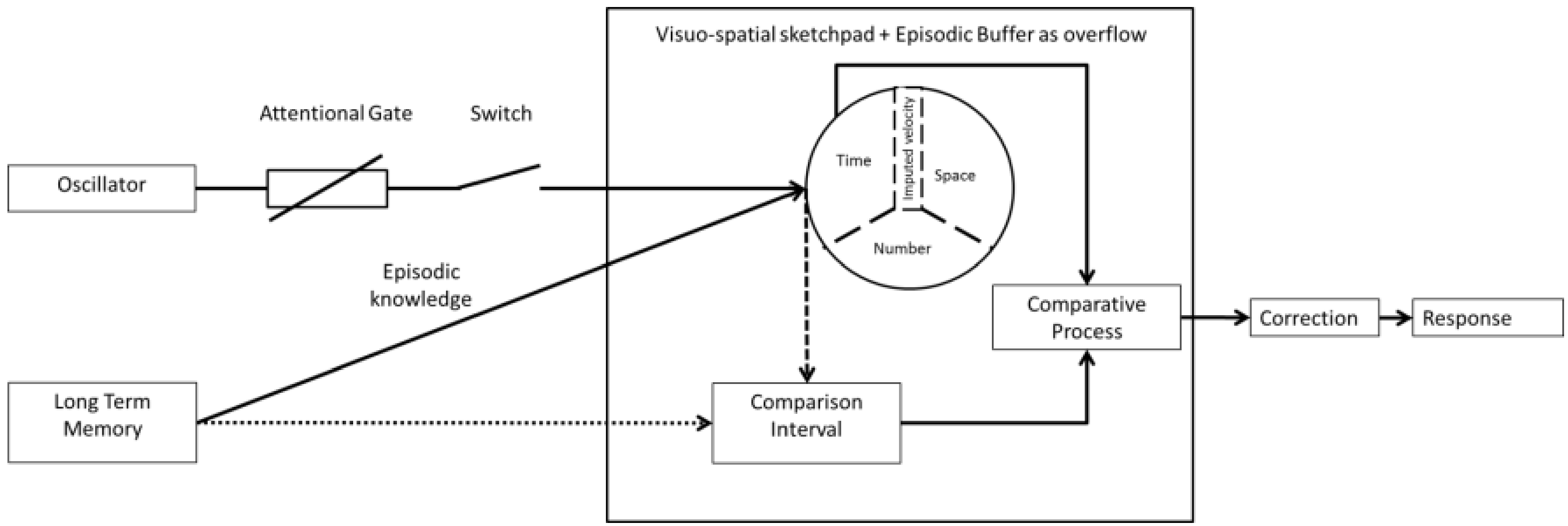

Figure 14.2. Revised internal clock model incorporating ATOM, imputed velocity, a potential correction factor, and a role for episodic knowledge. In this model it is not known whether episodic knowledge overwrites or simply adds to the amount of time accumulated.

task in the future. Typically it has been associated with a purely attentional account making use of standard internal clock models (Block & Zakay, 1997). However Waldum and Sahakyan (2013) demonstrated that time based prospective memory estimates were also influenced by background stimuli; specifically, estimates were influenced by the number of songs that played in the background during the task. When more songs played in the background time was perceived to have passed more quickly than when fewer songs played.

While distracters typically decrease the rate that time is thought to pass in prospective tasks (Predebon, 1996) participants are usually asked to attend to distracters. In Waldum and Sahakyan (2013) participants are not required to attend to the songs playing in the background. As the estimates were tied to song number the suggestion would be that participants were using knowledge about the standard duration of a song as a means of checking, overriding, or topping up their prospective estimate. In terms of the model proposed in Figure 14.2, this would be the contribution of episodic knowledge to the accumulator and therefore the representation of prospective time. Whether a check, top-up, or over-ride was occurring in Waldum and Sahakyan (2013) was unclear as attention was not measured or manipulated. However their results do suggest that retrospective timing and prospective timing can operate together in prospective tasks.

Alternative spectral account. While STT is the most obvious fit a case can also be made for a spectral account of time. While spectral accounts have been suggested to best apply to times in the subsecond range (Karmarkar & Buonomano, 2007) the notion that spatial representation systems may be used when representing time has a spectral element to it. In order for this to be true, spatial representation would need to change lawfully with time.

There is evidence to suggest that rats ‘move’ through mental maps in their mind while resting (Logothetis et al., 2012). Similarly, comparable brain regions associated with spatial navigation are also active at rest in humans (Spreng, Mar, & Kim, 2009). If it is assumed that this ‘movement’ occurs at a constant speed, tentative evidence of an intrinsic relationship between spatial representation and speed is formed. Under a spectral model the claim would then be that the relationship between spatial representations and a tendency to ‘move’ through them at rest could then be hijacked for the representation of time. Such a concept of spatial representations being paired with an underlying speed of movement fits with the idea of imputed velocity. The scan task may be an index of the speed at which an individual participant’s spatial representation system ’moves’ through an imagined space or environment. Again this could apply equally to retrospective and prospective timing.

Directions for Future Work

The most fruitful area for future research depends on whether focus is turned to the areas with the most to explore or the areas with the most practical ways to explore them. In absolute terms retrospective timing remains the greatest mystery while prospective timing has the most established forms of investigation.

Anyone investigating retrospective timing in future would benefit most from finding a paradigm that is fit for purpose. As was noted earlier, current investigations into retrospective timing typically include very few trials for individual participants and tend to involve time judgements on a scale that exceeds that of most prospective work. As such the first step should be to identify a paradigm that can include multiple trials.

In this respect something like the recognition memory task used here provides a good candidate in terms of trial number. Importantly if the current results did convey any message it was that the memory benefit is not restricted to the forget words. An effect was definitely seen in the remember words for the spatial replication and an effect may have been seen in the suprasecond temporal version. As such future work would probably benefit from removing the remember/forget division and simply focussing on running as many trials as possible in each remaining temporal location condition.

Time salience is also an issue. The difficulty with retrospective work is that the time element must not draw attention in order to prevent the task becoming a prospective timing task. Building on the model of retrospective timing here, a training period with target durations before the experiment may be beneficial. This will also have the advantage of enabling the testing of retrospective timing outlined earlier; specifically, the mis-pairing of stimuli and intervals during the study phase. If time information is encoded into retrospective memory and reconstructed without a timer, the stimulus, rather than the actual interval, should lead to the experienced interval.

Prospective timing contains many questions. Those specifically related to the studies conducted here have been covered in the previous chapter. However, when it comes to wider theory the association with space calls out for further attention. One large extension not considered earlier is a move to identify overlaps between time and space with larger scale stimuli. In particular, previous work has shown a scaling of time associated with scaled spaces (Delong, 1981), and work with both space and time suggest a common network continues to be activated at longer timescales. As spaces for representation become more complex, moving from rectangles to 3D environments, representation moves from spatial working memory (Baddeley, 2003) to the self-projection network (Buckner & Carroll, 2006), essentially recruiting the hippocampus. As times move beyond the suprasecond scale representation moves from the right lateralised cognitive timing system outlined by Lewis and Miall (2003b) to include the hippocampus (Richards, 1973; Vidalaki et al., 1999). As the self-projection network has been noted to overlap with episodic memory (Addis & Schacter, 2008) it would be particularly interesting to consider whether overlaps of prospective timing and retrospective timing might become more frequent as durations lengthen.

Conclusion

While associations between time and space seem obvious and unavoidable this may not always be the case. Retrospective timing failed to produce any reliable space-like effects whereas prospective timing may produce an apparent relationship via the intermediary of imputed velocity. This association is compatible with current models of timing and suggests some alterations to the most popular model: STT. In particular, refinement of working memory to the visuospatial sketchpad helps to capture time-space associations. The inclusion of the divisions proposed in Figure 8.1 (page 99) in the accumulator and imputed velocity as the means of association between time and space also makes sense. Future research would benefit from refinements to retrospective timing designs and may also benefit from testing a more complicated model. Prospective time research may also benefit from testing ideas associated with a model that includes prospective and retrospective elements. Support for such a model comes from the results of the experiments presented in this thesis as well as outside sources. These include memory based prospective timing and spatial navigation tasks.

Appendix A

Original Words

	List 1	List 2		List 1	List 2
1	barbeque	band	33	mat	hay
2	beer	basement	34	meadow	jail
3	blanket	beach	35	moose	jelly
4	bread	bee	36	mustard	kitchen
5	bus	bicycle	37	needle	lemon
6	canoe	blackboard	38	office	log
7	cat	boat	39	pail	mirror
8	cellar	bowl	40	paper	moon
9	chicken	broccoli	41	pavement	mouse
10	children	brush	42	pencil	mower
11	closet	bush	43	piano	noodles
12	cloud	butter	44	pickles	oven
13	cook	cabinet	45	port	pasture
14	cow	candy	46	river	pie
15	desert	car	47	rock	pillow
16	duck	cement	48	sandwich	pin
17	field	cheese	49	sauce	rail
18	fly	couch	50	shampoo	road
19	folder	country	51	ship	rabbit
20	food	crow	52	shrimp	seagull
21	forest	deer	53	sofa	sheet
22	fruit	desk	54	street	sleep
23	fudge	donkey	55	sun	slide
24	garden	elephant	56	thief	soda
25	highway	fish	57	thorn	tie
26	hill	floor	58	town	tomato
27	housecoat	flower	59	trail	towel
28	ice	fog	60	turtle	truck
29	knife	fox	61	tuxedo	valley
30	ladder	glasses	62	vinegar	water
31	lamp	goat	63	wasp	wine
32	leaf	hammer	64	wood	wrench

Buffer Words

1	milk	5	soap
2	clock	6	bag
3	lips	7	bell
4	chain	8	pocket

Additional Words

	List 1	List 2		List 1	List 2
1	apple	bridge	5	sand	sink
2	chair	dog	6	sister	stick
3	gate	hotel	7	stone	tap
4	lock	rose	8	tin	village

Appendix B

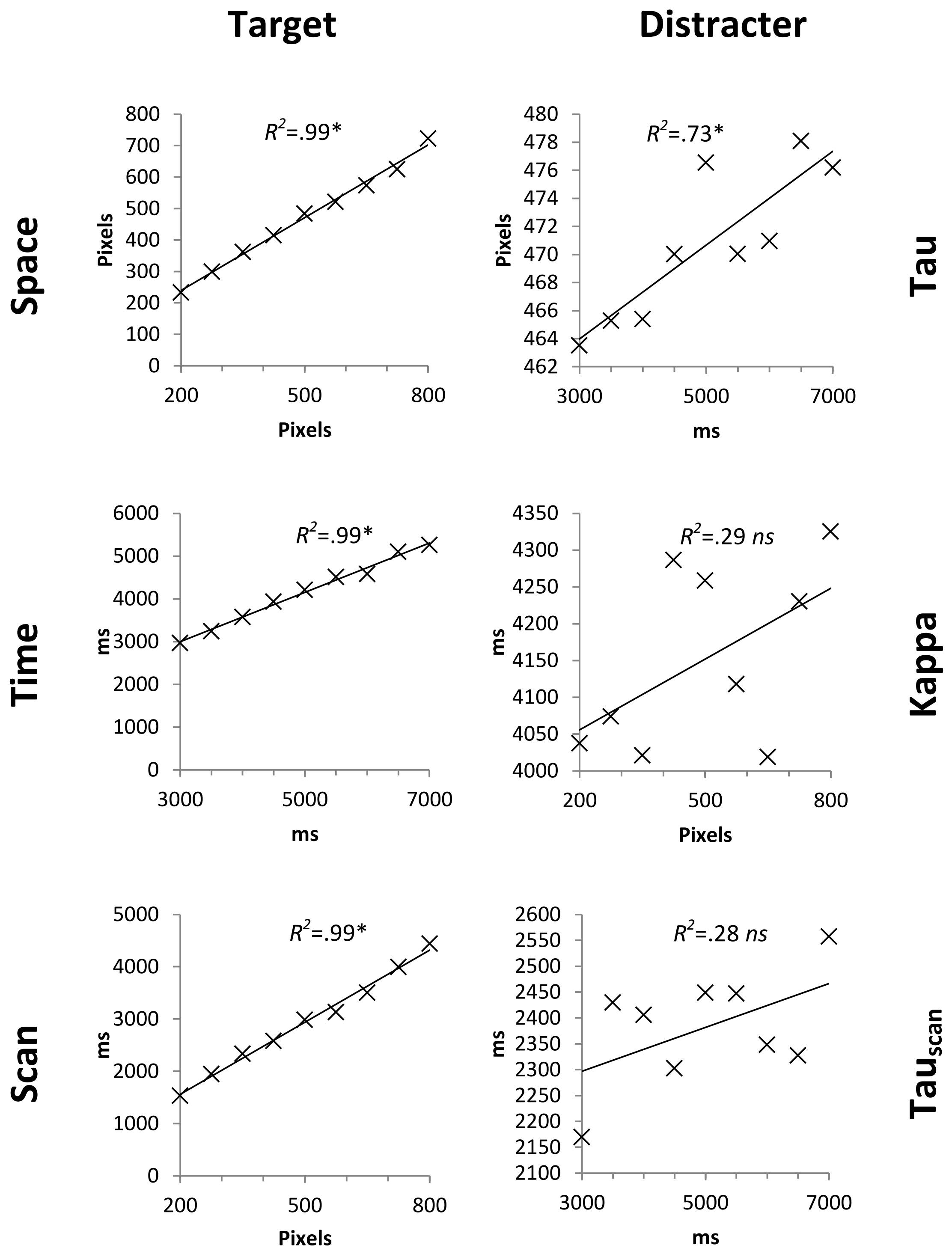

Figure B.1. The relationship between durations and distances produced and times and line lengths shown for the time, space, and scan tasks. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*. Data presented is from the scan-task-pre condition of experiment 6.

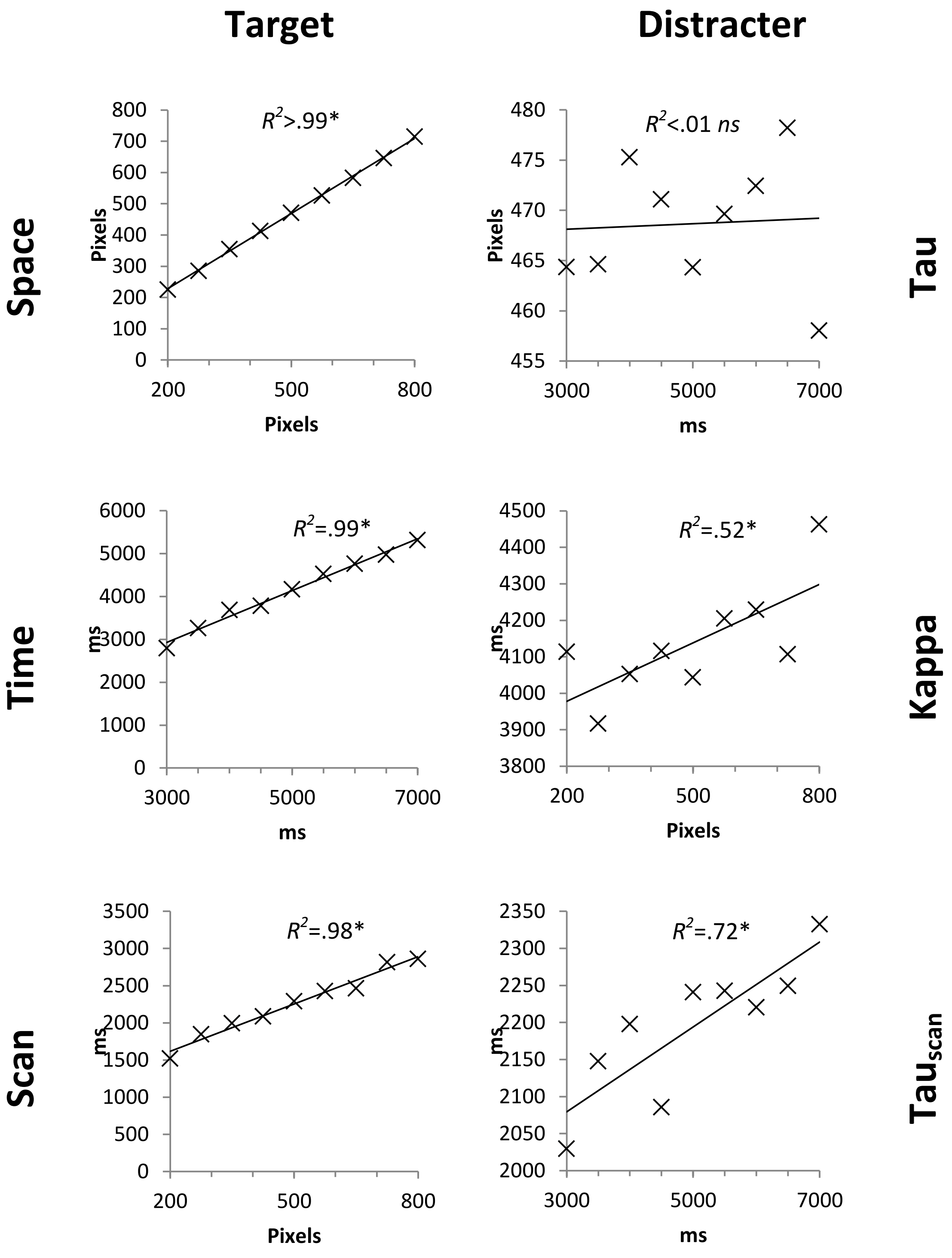

Figure B.1. The relationship between durations and distances produced and times and line lengths shown for the time, space, and scan tasks. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*. Data presented is from the scan-task-during condition of experiment 6.

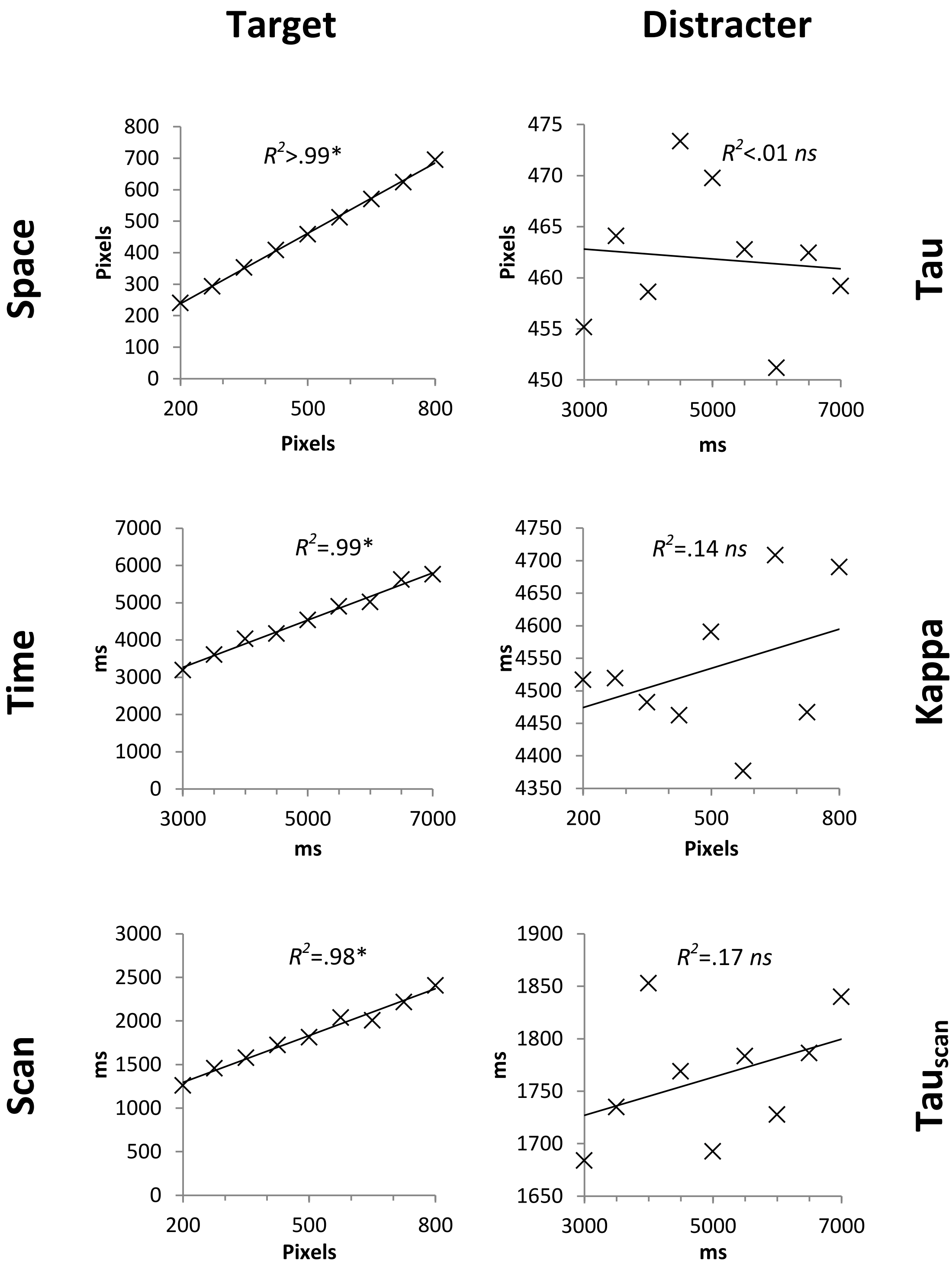

Figure B.1. The relationship between durations and distances produced and times and line lengths shown for the time, space, and scan tasks. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*. Data presented is from the scan-task-post condition of experiment 6.

Appendix C

Experiment 7

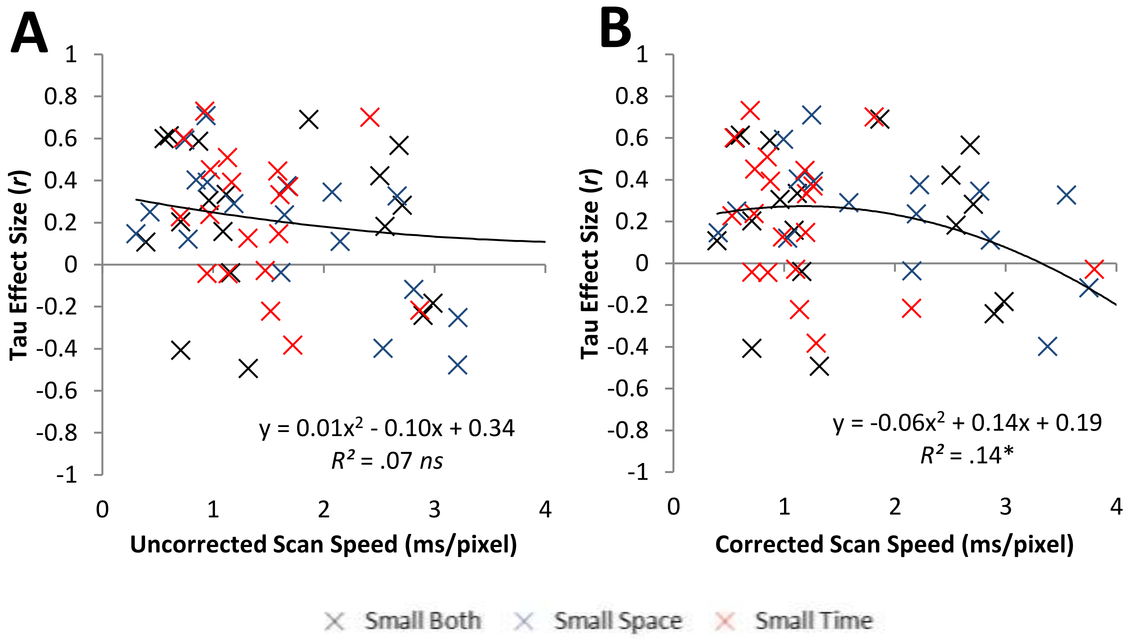

Figure C.1. The relationship between tau effect size and scan-speed in experiment 7 after the outliers have been removed and a correction for negative/positive group membership has been made. Different stimulus size classes are represented by different coloured crosses. Black depicts reduced time and space, blue reduced space, and red reduced time. Panel A shows the relationship when scan-speeds are not corrected for stimulus size. Panel B shows the relationship when scan-speeds are corrected for stimulus size. Trend lines are fitted to all data points and do not consider positive/negative group membership. Significant R^2 values (p<.05) are followed by an *. Non-significant R^2 values (p>.05) are followed by *ns*.

Appendix D

Basic Analysis of Participants Excluded from the Combined Analyses

Significant effects for scan ($F_{(1,38)}$= 19.02, p<.001), tau_{scan} ($F_{(1,38)}$=8.89, p=.005), time ($F_{(1,38)}$=107.73, p<.001), kappa ($F_{(1,38)}$=9.35, p=.004), and space were found ($F_{(1,38)}$=117.95, p<.001). The tau effect was not significant ($F_{(1,38)}$=0.78, p=.38).

The relationship between the tau effect and scan-speed is also present when the positive and negative values are combined, with group assignment based on earlier equations, and participants with a scan-speed of less than 0 are excluded ($F_{(2,30)}$=4.18, p=.025). If the groups are defined on the basis of their tau effect sign the relationship is not significant ($F_{(2,30)}$=2.26, p=.12).

Appendix E

Table E.1
Parameters (K values) associated with quadratic models used in calculating AICc values for the different models outlined in Chapter 13: Combined Analyses.

		x^2	x	constant	Group Shift	error	Scaling	order	Total
Uncorrected	Combined	1	1	1	1	1	0	0	5
	Separate	2	2	2	0	1	0	0	7
Corrected	Combined	1	1	1	1	1	1	1	7
	Separate	2	2	2	0	1	1	2	10

Table E.2
Parameters (K values) associated with sine models used in calculating AICc values for the different models outlined in Chapter 13: Combined Analyses.

		offset	Amp.	Phase shift	Freq.	Group Shift	error	Scaling	Total
Uncorrected	Combined	1	1	1	1	1	1	0	6
	Separate	2	2	2	2	0	1	0	9
Corrected	Combined	1	1	1	1	1	1	1	7
	Separate	2	2	2	2	0	1	1	10

Appendix F

Linear contrasts were conducted as part of a mixed design ANOVA. Stimulus type (time or space), target number, and distracter number were entered as within subjects variables. Stimulus size and scan task inclusion order were entered as between subjects factors. No main effect of or interactions involving stimulus size reached significance.

Significant linear contrasts were found for the main effects of target ($F_{(1,152)}$=3013.91, p<.001) and distracter condition number ($F_{(1,152)}$=24.43, p<.001). A significant interaction was found between distracter condition number and scan task inclusion order ($F_{(1,152)}$=5.96, p=.003).

Significant interactions were found between stimulus type and target ($F_{(1,152)}$=50.68, p<.001) and distracter ($F_{(1,152)}$=7.86, p=.006) condition number.

Significant quadratic trends were found for the interaction between, stimulus type, target number, and distracter number ($F_{(1,152)}$=8.00, p=.005), as well as target and distracter ($F_{(1,152)}$=4.76, p=.031).

No other significant interactions were found at the linear or quadratic level.

Analyses excluding inclusion order were preferred for three key reasons. First, the interaction between stimulus type, target number, and distracter number as well as target and distracter produce significant linear effects. These are theoretically more meaningful than quadratic effects. Second, the pattern produced when excluding participants with the scan task was included last is more similar to that produced when ignoring scan task inclusion order. Third, the pattern of interactions produced when scan task inclusion order is included is logically inconsistent.

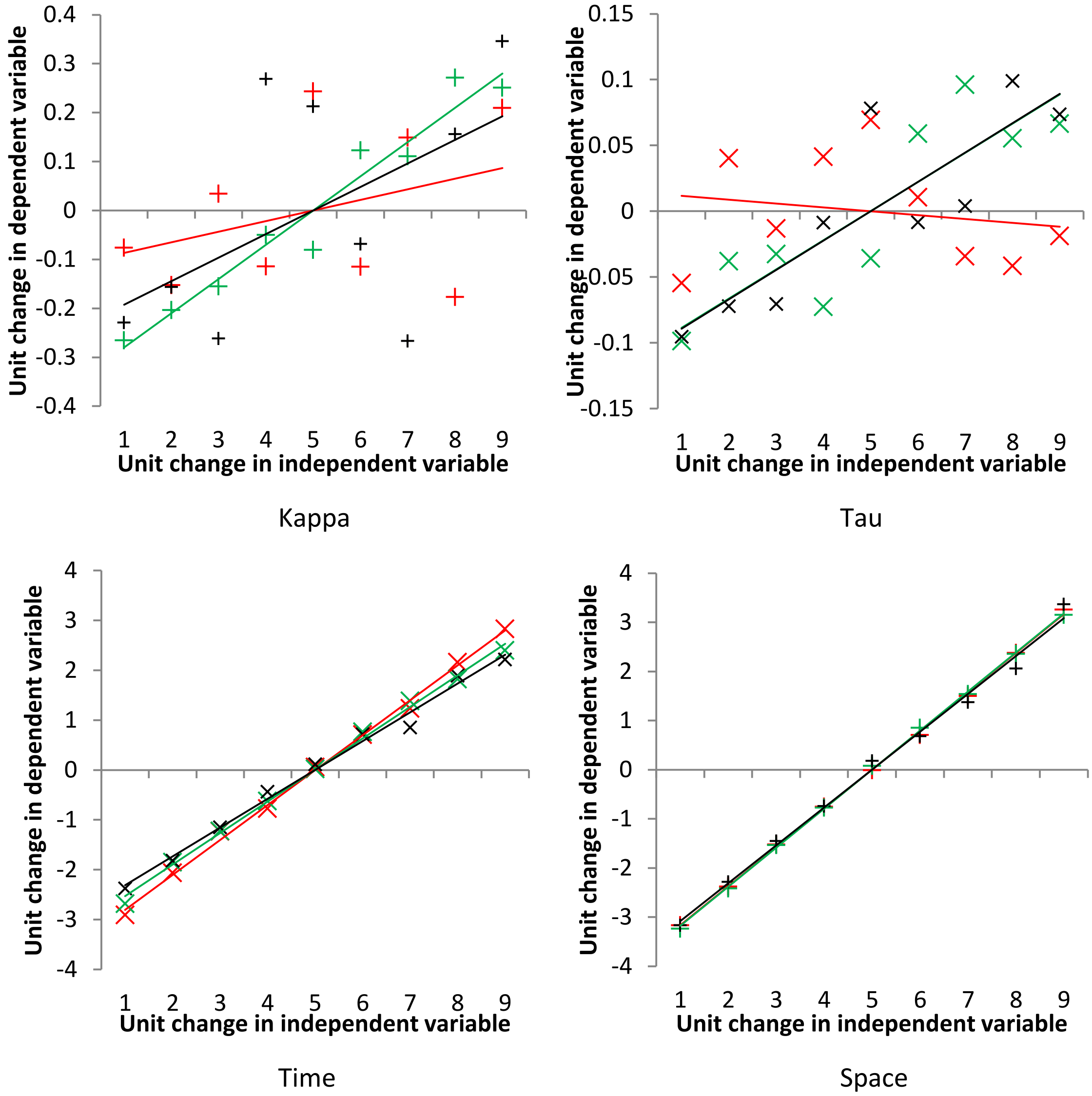

Figure F.1. Standardised values for kappa, tau, space, and time effects, plotted as a function of scan task inclusion order. Black lines show the pre condition, green lines the during condition, and red lines the post condition.

References

Addis, D. R., & Schacter, D. L. (2008). Constructive Episodic Simulation: Temporal Distance and Detail of Past and Future Events Modulate Hippocampal Engagement. *Hippocampus, 18*(2), 227-237. DOI:10.1002/hipo.20405

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the Past and Imagining the Future: Common and Distinct Neural Substrates During Event Construction and Elaboration. *Neuropsychologia, 45*(7), 1363-1377. DOI:10.1016/j.neuropsychologia.2006.10.016

Aflalo, T. N., & Graziano, M. S. (2008). Four-Dimensional Spatial Reasoning in Humans. *J Exp Psychol Hum Percept Perform, 34*(5), 1066-1077. DOI:10.1037/0096-1523.34.5.1066

Agresti, A., & Franklin, C. A. (2007). *Statistics: The Art and Science of Learning from Data*. New Jersey: Pearson Educatoin Inc.

Assmus, A., Marshall, J. C., Noth, J., Zilles, K., & Fink, G. R. (2005). Difficulty of Perceptual Spatiotemporal Integration Modulates the Neural Activity of Left Inferior Parietal Cortex. *Neuroscience, 132*(4), 923-927. DOI:10.1016/j.neuroscience.2005.01.047

Assmus, A., Marshall, J. C., Ritzl, A., Noth, J., Zilles, K., & Fink, G. R. (2003). Left Inferior Parietal Cortex Integrates Time and Space During Collision Judgments. *NeuroImage, 20*, S82-S88. DOI:10.1016/j.neuroimage.2003.09.025

Baddeley, A. (2000). The Episodic Buffer: A New Component of Working Memory? *Trends in Cognitive Sciences, 4*(11), 417-423. DOI:10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. *Nature Reviews Neuroscience, 4*(10), 829-839. DOI:10.1038/Nrn1201

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word Length and Structure of Short-Term-Memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589. DOI:10.1016/s0022-5371(75)80045-4

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed Forgetting in Implicit and Explicit Memory Tests - a Comparison of Methods. *Journal of Experimental Psychology-Learning Memory and Cognition, 19*(3), 603-616. DOI:10.1037//0278-7393.19.3.603

Battellil, L., Walsh, V., Pascual-Leone, A., & Cavanagh, P. (2008). The 'When' Parietal Pathway Explored by Lesion Studies. *Current Opinion in Neurobiology, 18*(2), 120-126. DOI:10.1016/j.conb.2008.08.004

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). *Neuroscience: Exploring the Brain* (3rd ed.). Philadelphia: Lippincott Williams & Wilkins.

Berlingeri, M., Bottini, G., Basilico, S., Silani, G., Zanardi, G., Sberna, M., . . . Paulesu, E. (2008). Anatomy of the Episodic Buffer: A Voxel-Based Morphometry Study in Patients with Dementia. *Behavioural Neurology, 19*(1-2), 29-34.

Beudel, M., Renken, R., Leenders, K. L., & de Jong, B. M. (2009). Cerebral Representations of Space and Time. *NeuroImage, 44*(3), 1032-1040. DOI:10.1016/j.neuroimage.2008.09.028

Bill, J. C., & Teft, L. W. (1972). Space-Time Relations - Effects of Variations in Stimulus and Interstimulus Interval Duration on Perceived Visual Extent. *Acta Psychologica, 36*(5), 358-369.

Binkofski, F., & Block, R. A. (1996). Accelerated Time Experience after Left Frontal Cortex Lesion. *Neurocase, 2*(6), 485-493.

Bisson, N., Tobin, S., & Grondin, S. (2012). Prospective and Retrospective Time Estimates of Children: A Comparison Based on Ecological Tasks. *Plos One, 7*(3). DOI:10.1371/journal.pone.0033049

Block, R. A. (1990). *Cognitive Models of Psychological Time.* Illsdale, NJ: Lawrence Erlbaum Associates, Inc;.

Block, R. A., & Zakay, D. (1997). Prospective and Retrospective Duration Judgments: A Meta-Analytic Review. *Psychonomic Bulletin & Review, 4*(2), 184-197. DOI:10.3758/BF03209393

Boltz, M. G. (1994). Changes in Internal Tempo and Effects on the Learning and Remembering of Event Durations. *Journal of Experimental Psychology-Learning Memory and Cognition, 20*(5), 1154-1171. DOI:10.1037/0278-7393.20.5.1154

Bonato, M., Zorzi, M., & Umilta, C. (2012). When Time Is Space: Evidence for a Mental Time Line. *Neuroscience and Biobehavioral Reviews, 36*(10), 2257-2273. DOI:10.1016/j.neubiorev.2012.08.007

Boroditsky, L. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology, 43*(1), 1-22. DOI:10.1006/cogp.2001.0748

Bottini, R., & Casasanto, D. (2010). Implicit Spatial Length Modulates Time Estimates, but Not Vice Versa. In C. Holscher, T. F. Shipley, M. O. Belardinelli, J. A. Bateman, & N. S. Newcombe (Eds.), *Spatial Cognition Vii* (Vol. 6222, pp. 152-162). Berlin: Springer-Verlag Berlin.

Branas-Garza, P., Espinosa-Fernandez, L., & Serrano-del-Rosal, R. (2007). Effects of Gender and Age on Retrospective Time Judgements. *Time & Society, 16*(1), 99-118. DOI:10.1177/0961463x07074104

Brown, S. W. (1985). Time Perception and Attention - the Effects of Prospective Versus Retrospective Paradigms and Task Demands on Perceived Duration. *Perception & Psychophysics, 38*(2), 115-124. DOI:10.3758/bf03198848

Brown, S. W. (1995). Time, Change, and Motion - the Effects of Stimulus Movement on Temporal Perception. *Perception & Psychophysics, 57*(1), 105-116. DOI:10.3758/BF03211853

Brown, S. W. (1997). Attentional Resources in Timing: Interference Effects in Concurrent Temporal and Nontemporal Working Memory Tasks. *Perception & Psychophysics, 59*(7), 1118-1140. DOI:10.3758/BF03205526

Buckner, R. L., & Carroll, D. C. (2006). Self-Projection and the Brain. *Trends in Cognitive Neuroscience, 11*(2), 49-57. DOI:10.1016/j.tics.2006.11.004

Bueti, D., Bahrami, B., & Walsh, V. (2008). Sensory and Association Cortex in Time Perception. *Journal of Cognitive Neuroscience, 20*(6), 1054-1062. DOI:10.1162/jocn.2008.20060

Bueti, D., & Walsh, V. (2009). The Parietal Cortex and the Representation of Time, Space, Number and Other Magnitudes. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1525), 1831-1840. DOI:10.1098/rstb.2009.0028

Bullock, T. H., Karamursel, S., Achimowicz, J. Z., McClune, M. C., & Basareroglu, C. (1994). Dynamic Properties of Human Visual-Evoked and Omitted Stimulus Potentials. *Electroencephalography and Clinical Neurophysiology, 91*(1), 42-53. DOI:10.1016/0013-4694(94)90017-5

Buonomano, D. V. (2007). The Biology of Time across Different Scales. *Nature Chemical Biology, 3*(10), 594-597. DOI:10.1038/nchembio1007-594

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Mulimodel Inference: A Practical Information-Theoretic Approach* (2 ed.). New York: Springer.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory Dominance over Vision in the Perception of Interval Duration. *Experimental Brain Research, 198*(1), 49-57. DOI:10.1007/s00221-009-1933-z

Cai, Z. G. G., Connell, L., & Holler, J. (2013). Time Does Not Flow without Language: Spatial Distance Affects Temporal Duration Regardless of Movement or Direction. *Psychonomic Bulletin & Review, 20*(5), 973-980. DOI:10.3758/s13423-013-0414-3

Cant, J. S., Large, M. E., McCall, L., & Goodale, M. A. (2008). Independent Processing of Form, Colour, and Texture in Object Perception. *Perception, 37*(1), 57-78. DOI:10.1068/pS727

Casarotti, M., Michielin, M., Zorzi, M., & Umilta, C. (2007). Temporal Order Judgment Reveals How Number Magnitude Affects Visuospatial Attention. *Cognition, 102*(1), 101-117. DOI:10.1016/j.cognition.2006.09.001

Casasanto, D., & Boroditsky, L. (2008). Time in the Mind: Using Space to Think About Time. *Cognition, 106*, 579-593. DOI:10.1016/j.cognition.2007.03.004

Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112*(1), 155-159. DOI:10.1037/0033-2909.112.1.155

Cohen, J., Hansel, C. E. M., & Sylvester, J. D. (1953). A New Phenomenon in Time Judgment. *Nature, 172*(4385), 901-901. DOI:10.1038/172901a0

Cohen, J., Hansel, C. E. M., & Sylvester, J. D. (1955). Interdependence in Judgments of Space, Time and Movement. *Acta Psychologica, 11*(2-4), 360-372. DOI:10.1016/s0001-6918(55)80098-4

Collyer, C. E. (1977). Discrimination of Spatial and Temporal Intervals Defined by 3 Light-Flashes - Effects of Spacing on Temporal Judgments and of Timing on Spatial Judgments. *Perception & Psychophysics, 21*(4), 357-364. DOI:10.3758/bf03199487

Coltheart, M. (1981). The Mrc Psycholinguistic Database. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 33*(NOV), 497-505.

Coslett, H. B., Shenton, J., Dyer, T., & Wiener, M. (2009). Cognitive Timing: Neuropsychology and Anatomic Basis. *Brain Research, 1254*, 38-48. DOI:10.1016/j.brainres.2008.11.015

Coull, J. T., Nazarian, B., & Vidal, F. (2008). Timing, Storage, and Comparison of Stimulus Duration Engage Discrete Anatomical Components of a Perceptual Timing Network. *Journal of Cognitive Neuroscience, 20*(12), 2185-2197. DOI:10.1162/jocn.2008.20153

Coull, J. T., & Nobre, A. C. (1998). Where and When to Pay Attention: The Neural Systems for Directing Attention to Spatial Locations and to Time Intervals as Revealed by Both PET and fMRI. *Journal of Neuroscience, 18*(18), 7426-7435.

Coull, J. T., & Nobre, A. C. (2008). Dissociating Explicit Timing from Temporal Expectation with fMRI. *Current Opinion in Neurobiology, 18*(2), 137-144. DOI:10.1016/j.conb.2008.07.011

Critchley, M. (1953). The Parietal Lobes *The Parietal Lobes*. New York: Hafner.

Danckert, J., Ferber, S., Pun, C., Broderick, C., Striemer, C., Rock, S., & Stewart, D. (2007). Neglected Time: Impaired Temporal Perception of Multisecond Intervals in Unilateral Neglect. *Journal of Cognitive Neuroscience, 19*(10), 1706-1720. DOI:10.1162/jocn.2007.19.10.1706

Debner, J. A., & Jacoby, L. L. (1994). Unconscious Perception - Attention, Awareness, and Control. *Journal of Experimental Psychology-Learning Memory and Cognition, 20*(2), 304-317. DOI:10.1037/0278-7393.20.2.304

Defeyter, M. A., Russo, R., & McPartlin, P. L. (2009). The Picture Superiority Effect in Recognition Memory: A Developmental Study Using the Response Signal Procedure. *Cognitive Development, 24*(3), 265-273. DOI:10.1016/j.cogdev.2009.05.002

Dehaene, S., Bossini, S., & Giraux, P. (1993). The Mental Representation of Parity and Number Magnitude. *Journal of Experimental Psychology-General, 122*(3), 371-396. DOI:10.1037/0096-3445.122.3.371

Delong, A. J. (1981). Phenomenological Space-Time - toward an Experiential Relativity. *Science, 213*(4508), 681-683. DOI:10.1126/science.7256273

Di Bono, M. G., Casarotti, M., Priftis, K., Gaya, L., Umilta, C., & Zorzi, M. (2012). Priming the Mental Time Line. *Journal of Experimental Psychology-Human Perception and Performance, 38*(4), 838-842. DOI:10.1037/a0028346

Donaldson, W. (1992). Measuring Recognition Memory. *Journal of Experimental Psychology-General, 121*(3), 275-277.

Droit-Volet, S. (2010). Speeding up a Master Clock Common to Time, Number and Length? *Behavioural Processes, 85*(2), 126-134. DOI:10.1016/j.beproc.2010.06.017

Durgin, F. H., & Sternberg, S. (2002). The Time of Consciousness and Vice Versa. *Consciousness and Cognition, 11*(2), 284-290. DOI:10.1006/ccog.2002.0566

Dutta, A., & Nairne, J. S. (1993). The Separability of Space and Time - Dimensional Interaction in the Memory Trace. *Memory & Cognition, 21*(4), 440-448. DOI:10.3758/bf03197175

Eisler, H. (1981). Applicability of the Parallel-Clock Model to Duration Discrimination. *Perception & Psychophysics, 29*(3), 225-233. DOI:10.3758/bf03207289

El Haj, M., Moroni, C., Samson, S., Fasotti, L., & Allain, P. (2013). Prospective and Retrospective Time Perception Are Related to Mental Time Travel: Evidence from Alzheimer's Disease. *Brain and Cognition, 83*(1), 45-51. DOI:10.1016/j.bandc.2013.06.008

Field, D. T., & Groeger, J. A. (2004). Temporal Interval Production and Short-Term Memory. *Perception & Psychophysics, 66*(5), 808-819. DOI:10.3758/BF03194975

Firmino, E. A., Bueno, J. L. O., & Bigand, E. (2009). Travelling through Pitch Space Speeds up Musical Time. *Music Perception, 26*(3), 205-209. DOI:10.1525/mp.2009.26.3.205

Fortin, C., & Breton, R. (1995). Temporal Interval Production and Processing in Working-Memory. *Perception & Psychophysics, 57*(2), 203-215. DOI:10.3758/BF03206507

Fougnie, D., & Alvarez, G. A. (2011). Object Features Fail Independently in Visual Working Memory: Evidence for a Probabilistic Feature-Store Model. *Journal of Vision, 11*(12). DOI:10.1167/11.12.3

Frassinetti, F., Magnani, B., & Oliveri, M. (2009). Prismatic Lenses Shift Time Perception. *Psychological Science, 20*(8), 949-954. DOI:10.1111/j.1467-9280.2009.02390.x

Garner, W. R. (1974). *The Processing of Information and Structure*. Potomac. Maryland: Lawrence Erlbaum Associates.

Gibbon, J. (1977). Scalar Expectancy-Theory and Webers Law in Animal Timing. *Psychological Review, 84*(3), 279-325. DOI:10.1037//0033-295x.84.3.279

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar Timing in Memory. *Annals of the New York Academy of Sciences, 423*(May), 52-77. DOI:10.1111/j.1749-6632.1984.tb23417.x

Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997). Toward a Neurobiology of Temporal Cognition: Advances and Challenges. *Current Opinion in Neurobiology, 7*(2), 170-184. DOI:10.1016/S0959-4388(97)80005-0

Gibbons, H., & Stahl, J. (2008). Erp Predictors of Individual Performance on a Prospective Temporal Reproduction Task. *Psychological Research-Psychologische Forschung, 72*(3), 311-320. DOI:10.1007/s00426-006-0105-z

Gilden, D. L., & Marusich, L. R. (2009). Contraction of Time in Attention-Deficit Hyperactivity Disorder. *Neuropsychology, 23*(2), 265-269. DOI:10.1037/A0014553

Gneezy, U., & Rustichini, A. (2000). Pay Enough or Don't Pay at All. *Quarterly Journal of Economics, 115*(3), 791-810. DOI:10.1162/003355300554917

Godden, D. R., & Baddeley, A. D. (1975). Context-Dependent Memory in Two Natural Environments: On Land and Underwater. *British Journal of Psychology, 66*(3), 325-331. DOI:10.1111/j.2044-8295.1975.tb01468.x

Goldreich, D. (2007). A Bayesian Perceptual Model Replicates the Cutaneous Rabbit and Other Tactile Spatiotemporal Illusions. *Plos One, 2*(3). DOI:10.1371/journal.pone.0000333

Goldston.S, & Lhamon, W. T. (1974). Studies of Auditory-Visual Differences in Human Time Judgment .1. Sounds Are Judged Longer Than Lights. *Perceptual and Motor Skills, 39*(1), 63-82.

Gooch, C. M., Wiener, M., Wencil, E. B., & Coslett, H. B. (2010). Interval Timing Disruptions in Subjects with Cerebellar Lesions. *Neuropsychologia, 48*(4), 1022-1031. DOI:10.1016/j.neuropsychologia.2009.11.028

Gorea, A. (2011). Ticks Per Thought or Thoughts Per Tick? A Selective Review of Time Perception with Hints on Future Research. *Journal of Physiology-Paris, 105*(4-6), 153-163. DOI:10.1016/j.jphysparis.2011.09.008

Grier, J. B. (1971). Nonparametric Indexes for Sensitivity and Bias - Computing Formulas. *Psychological Bulletin, 75*(6), 424-&. DOI:10.1037/h0031246

Grondin, S., Bisson, N., Gagnon, C., Gamache, P. L., & Matteau, A. A. (2009). Little to Be Expected from Auditory Training for Improving Visual Temporal Discrimination. *Neuroquantology, 7*(1), 95-102.

Grondin, S., & Plourde, M. (2007). Judging Multi-Minute Intervals Retrospectively. *Quarterly Journal of Experimental Psychology, 60*(9), 1303-1312. DOI:10.1080/17470210600988976

Grossberg, S., & Merrill, J. W. L. (1992). A Neural-Network Model of Adaptively Timed Reinforcement Learning and Hippocampal Dynamics. *Cognitive Brain Research, 1*(1), 3-38. DOI:10.1016/0926-6410(92)90003-A

Gruber, O., Kleinschmidt, A., Binkofski, F., Steinmetz, H., & von Cramon, D. Y. (2000). Cerebral Correlates of Working Memory for Temporal Information. *Neuroreport, 11*(8), 1689-1693. DOI:10.1097/00001756-200006050-00019

Hackley, S. A., & Valle-Inclan, F. (2003). Which Stages of Processing Are Speeded by a Warning Signal? *Biological Psychology, 64*(1-2), 27-45. DOI:10.1016/s0301-0511(03)00101-7

Harrington, D. L., Boyd, L. A., Mayer, A. R., Sheltraw, D. M., Lee, R. R., Huang, M. X., & Rao, S. M. (2004). Neural Representation of Interval Encoding and Decision Making. *Cognitive Brain Research, 21*(2), 193-205. DOI:10.1016/j.cogbrainres.2004.01.010

Hassabis, D., & Maguire, E. A. (2007). Deconstructing Episodic Memory with Construction. *Trends Cogn Sci, 11*(7), 299-306. DOI:10.1016/j.tics.2007.05.001

Hautus, M. J. (1995). Corrections for Extreme Proportions and Their Biasing Effects on Estimated Values of D'. *Behavior Research Methods Instruments & Computers, 27*(1), 46-51. DOI:10.3758/BF03203619

Helson, H., & King, S. M. (1931). The Tau Effect: An Example of Psychological Relativity. *Journal of Experimental Psychology, 14*, 202-217. DOI:10.1037/h0071164

Henry, M. J., & McAuley, J. D. (2009). Evaluation of an Imputed Pitch Velocity Model of the Auditory Kappa Effect. *Journal of Experimental Psychology-Human Perception and Performance, 35*(2), 551-564. DOI:10.1037/0096-1523.35.2.551

Henry, M. J., McAuley, J. D., & Zaleha, M. (2009). Evaluation of an Imputed Pitch Velocity Model of the Auditory Tau Effect. *Attention Perception & Psychophysics, 71*(6), 1399-1413. DOI:10.3758/app.71.6.1399

Heyman, J., & Ariely, D. (2004). Effort for Payment - a Tale of Two Markets. *Psychological Science, 15*(11), 787-793. DOI:10.1111/j.0956-7976.2004.00757.x

Hicks, R. E., Miller, G. W., & Kinsbourne, M. (1976). Prospective and Retrospective Judgments of Time as a Function of Amount of Information Processed. *American Journal of Psychology, 89*(4), 719-730. DOI:10.2307/1421469

Hinton, S. C., Harrington, D. L., Binder, J. R., Durgerian, S., & Rao, S. M. (2004). Neural Systems Supporting Timing and Chronometric Counting: An fMRI Study. *Cognitive Brain Research, 21*(2), 183-192. DOI:10.1016/j.cogbrainres.2004.04.009

Hinton, S. C., & Meck, W. H. (2004). Frontal-Striatal Circuitry Activated by Human Peak-Interval Timing in the Supra-Seconds Range. *Cognitive Brain Research, 21*(2), 171-182. DOI:10.1016/j.cogbrainres.2004.08.005

Hinton, S. C., & Rao, S. M. (2004). "One-Thousand One... One-Thousand Two... ": Chronometric Counting Violates the Scalar Property in Interval Timing. *Psychonomic Bulletin & Review, 11*(1), 24-30. DOI:10.3758/BF03206456

Hirsh, I. J. (1959). Auditory Perception of Temporal Order. *Journal of the Acoustical Society of America, 31*(6), 759-767.

Hirsh, I. J., & Sherrick, C. E. (1961). Perceived Order in Different Sense Modalities. *Journal of Experimental Psychology, 62*(5), 423-&. DOI:10.1037/h0045283

Hopfield, J. J., & Brody, C. D. (2001). What Is a Moment? Transient Synchrony as a Collective Mechanism for Spatiotemporal Integration. *Proceedings of the National Academy of Sciences of the United States of America, 98*(3), 1282-1287. DOI:10.1073/pnas.031567098

Hourihan, K. L., Goldberg, S., & Taylor, T. L. (2007). The Role of Spatial Location in Remembering and Forgetting Peripheral Words. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 61*(2), 91-101. DOI:10.1037/cjep2007010

Intonspeterson, M. J. (1983). Imagery Paradigms - How Vulnerable Are They to Experimenters Expectations. *Journal of Experimental Psychology-Human Perception and Performance, 9*(3), 394-412. DOI:10.1037/0096-1523.9.3.394

Intonspeterson, M. J., & White, A. R. (1981). Experimenter Naivete and Imaginal Judgments. *Journal of Experimental Psychology-Human Perception and Performance, 7*(4), 833-843. DOI:10.1037//0096-1523.7.4.833

Ivry, R. B., & Hazeltine, R. E. (1992). Models of Timing-with-a-Timer. In F. Macar, P. V, & F. W. J (Eds.), *Time, Action, and Cognition*. Dordecht: Kluwer Academic Publishers.

Ivry, R. B., & Spencer, R. M. C. (2004). The Neural Representation of Time. *Current Opinion in Neurobiology, 14*(2), 225-232. DOI:10.1016/j.conb.2004.03.013

James, W. (1890). *The Principles of Psychology* (Vol. 1). New York: Henry Holt and Co Inc.

Johnson, V. E. (2013). Revised Standards for Statistical Evidence. *Proceedings of the National Academy of Sciences, 110*(48), 201313476. DOI:10.1073/pnas.1313476110

Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially Localized Distortions of Event Time. *Current Biology, 16*(5), 472-479. DOI:10.1016/j.cub.2006.01.032

Jolicoeur, P., & Kosslyn, S. M. (1985). Is Time to Scan Visual Images Due to Demand Characteristics. *Memory & Cognition, 13*(4), 320-332. DOI:10.3758/BF03202500

Jones, B., & Huang, Y. L. (1982). Space-Time Dependencies in Psychophysical Judgment of Extent and Duration - Algebraic Models of the Tau and Kappa Effects. *Psychological Bulletin, 91*(1), 128-142. DOI:10.1037/0033-2909.91.1.128

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the Absence of Clocks: Encoding Time in Neural Network States. *Neuron, 53*(3), 427-438. DOI:10.1016/j.neuron.2007.01.006

Kawabe, T., Miura, K., & Yamada, Y. (2008). Audiovisual Tau Effect. *Acta Psychologica, 128*(2), 249-254. DOI:10.1016/j.actpsy.2008.01.004

Koch, G., Costa, A., Brusa, L., Peppe, A., Gatto, I., Torriero, S., . . . Caltagirone, C. (2008). Impaired Reproduction of Second but Not Millisecond Time Intervals in Parkinson's Disease. *Neuropsychologia, 46*(5), 1305-1313. DOI:10.1016/j.neuropsychologia.2007.12.005

Koch, G., Oliveri, M., Carlesimo, G. A., & Caltagirone, C. (2002). Selective Deficit of Time Perception in A. Patient with Right Prefrontal Cortex Lesion. *Neurology, 59*(10), 1658-1659.

Koch, G., Oliveri, M., Torriero, S., & Caltagirone, C. (2003). Underestimation of Time Perception after Repetitive Transcranial Magnetic Stimulation. *Neurology, 60*(11), 1844-1846.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual Images Preserve Metric Spatial Information: Evidence from Studies of Image Scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4*(1), 47-60. DOI:10.1037//0096-1523.4.1.47

Kosslyn, S. M., Margolis, J. A., Barrett, A. M., Goldknopf, E. J., & Daly, P. F. (1990). Age-Differences in Imagery Abilities. *Child Development, 61*(4), 995-1010. DOI:10.1111/j.1467-8624.1990.tb02837.x

Kosslyn, S. M., Reiser, B. J., Farah, M. J., & Fliegel, S. L. (1983). Generating Visual Images - Units and Relations. *Journal of Experimental Psychology-General, 112*(2), 278-303. DOI:10.1037//0096-3445.112.2.278

Kristofferson, A. B. (1980). A Quantal Step Function in Duration Discrimination. *Perception & Psychophysics, 27*(4), 300-306. DOI:10.3758/BF03206118

Lebensfield, P., & Wapner, S. (1968). Configuration and Space-Time Interdependence. *American Journal of Psychology, 81*(1), 106-&. DOI:10.2307/1420814

Lejeune, H., & Wearden, J. H. (2009). Vierordt's the Experimental Study of the Time Sense (1868) and Its Legacy. *European Journal of Cognitive Psychology, 21*(6), 941-960. DOI:10.1080/09541440802453006

Lewis, P. A., & Miall, R. C. (2003a). Brain Activation Patterns During Measurement of Sub- and Supra-Second Intervals. *Neuropsychologia, 41*(12), 1583-1592. DOI:10.1016/S0028-3932(03)00118-0

Lewis, P. A., & Miall, R. C. (2003b). Distinct Systems for Automatic and Cognitively Controlled Time Measurement: Evidence from Neuroimaging. *Current Opinion in Neurobiology, 13*(2), 250-255. DOI:10.1016/s0959-4388(03)00036-9

Lewis, P. A., & Miall, R. C. (2006). A Right Hemispheric Prefrontal System for Cognitive Time Measurement. *Behavioural Processes, 71*(2-3), 226-234. DOI:10.1016/j.beproc.2005.12.009

Liberman, N., & Trope, Y. (2008). The Psychology of Transcending the Here and Now. *Science, 322*(5905), 1201-1205. DOI:10.1126/science.1161958

Loftus, E. F. (2005). Planting Misinformation in the Human Mind: A 30-Year Investigation of the Malleability of Memory. *Learning & Memory, 12*(4), 361-366. DOI:10.1101/lm.94705

Logothetis, N. K., Eschenko, O., Murayama, Y., Augath, M., Steudel, T., Evrard, H. C., . . . Oeltermann, A. (2012). Hippocampal-Cortical Interaction During Periods of Subcortical Silence. *Nature, 491*(7425), 547-553. DOI:10.1038/nature11618

Lourenco, S. F., & Longo, M. R. (2010). General Magnitude Representation in Human Infants. *Psychological Science, 21*(6), 873-881. DOI:10.1177/0956797610370158

Lustig, C., Matell, M. S., & Meck, W. H. (2005). Not "Just" a Coincidence: Frontal-Striatal Interactions in Working Memory and Interval Timing. *Memory, 13*(3-4), 441-448. DOI:10.1080/09658210344000404

Macar, F., & Vidal, F. (2004). Event-Related Potentials as Indices of Time Processing: A Review. *Journal of Psychophysiology, 18*(2-3), 89-104. DOI:10.1027/0269-8803.18.2-3.89

MacMillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2 ed.). New Jersey: Lawrence Erlbaum Associates Inc.

Malapani, C., Rakitin, B., Levy, R., Meck, W. H., Deweer, B., Dubois, B., & Gibbon, J. (1998). Coupled Temporal Memories in Parkinson's Disease: A Dopamine-Related Dysfunction. *Journal of Cognitive Neuroscience, 10*(3), 316-331. DOI:10.1162/089892998562762

Mangels, J. A., Ivry, R. B., & Shimizu, N. (1998). Dissociable Contributions of the Prefrontal and Neocerebellar Cortex to Time Perception. *Cognitive Brain Research, 7*(1), 15-39. DOI:10.1016/S0926-6410(98)00005-6

Masson, M. E. J., & Loftus, G. R. (2003). Using Confidence Intervals for Graphically Based Data Interpretation. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 57*(3), 203-220. DOI:10.1037/h0087426

Masuda, T., Kimura, A., Dan, I., & Wada, Y. (2011). Effects of Environmental Context on Temporal Perception Bias in Apparent Motion. *Vision Research, 51*(15), 1728-1740. DOI:10.1016/j.visres.2011.05.016

McFarland, C. P., & Glisky, E. L. (2009). Frontal Lobe Involvement in a Task of Time-Based Prospective Memory. *Neuropsychologia, 47*(7), 1660-1669. DOI:10.1016/j.neuropsychologia.2009.02.023

Meck, W. H. (1996). Neuropharmacology of Timing and Time Perception. *Cognitive Brain Research, 3*(3-4), 227-242. DOI:10.1016/0926-6410(96)00009-2

Melgire, M., Ragot, R., Samson, S., Penney, T. B., Meck, W. H., & Pouthas, V. (2005). Auditory/Visual Duration Bisection in Patients with Left or Right Medial-Temporal Lobe Resection. *Brain and Cognition, 58*(1), 119-124. DOI:10.1016/j.bandc.2004.09.013

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review, 63*(2), 81-97. DOI:10.1037/h0043158

Miller, R. (1996). *Axonal Conduction Time and Human Cerebral Laterality. A Psychobiological Theory*. Amsterdam: Harwood Academic Publishers.

Mills, J. N., Minors, D. S., & Waterhouse, J. M. (1974). Circadian-Rhythms of Human Subjects without Timepieces or Indication of Alternation of Day and Night. *Journal of Physiology-London, 240*(3), 567-594.

Mioni, G., Stablum, F., McClintock, S., & Grondin, S. (2014). Different Methods for Reproducing Time, Different Results. *Attention, Perception, & Psychophysics*, 1-7. DOI:10.3758/s13414-014-0625-3

Moore, K. E. (2006). Space-to-Time Mappings and Temporal Concepts. *Cognitive Linguistics, 17*(2), 199-244. DOI:10.1515/Cog.2006.005

Nicholls, M. E. (1996). Temporal Processing Asymmetries between the Cerebral Hemispheres: Evidence and Implications. *Laterality, 1*(2), 97-137. DOI:10.1080/713754234

Ogden, R. S., Wearden, J. H., Gallagher, D. T., & Montgomery, C. (2011). The Effect of Alcohol Administration on Human Timing: A Comparison of Prospective Timing, Retrospective Timing and Passage of Time Judgements. *Acta Psychologica, 138*(1), 254-262. DOI:10.1016/j.actpsy.2011.07.002

Oldfield, R. C. (1971). The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsyhcologia, 9*, 97-113.

Oliveri, M., Koch, G., Salerno, S., Torriero, S., Lo Gerfo, E., & Caltagirone, C. (2009). Representation of Time Intervals in the Right Posterior Parietal Cortex: Implications for a Mental Time Line. *NeuroImage, 46*(4), 1173-1179. DOI:10.1016/j.neuroimage.2009.03.042

Ornstein, R. E. (1969). *On the Experience of Time*. Harmondsworth: Penguin.

Paulsen, J. S., Zimbelman, J. L., Hinton, S. C., Langbehn, D. R., Leveroni, C. L., Benjamin, M. L., . . . Rao, S. M. (2004). fMRI Biomarker of Early Neuronal Dysfunction in Presymptomatic Huntington's Disease. *American Journal of Neuroradiology, 25*(10), 1715-1721.

Perbal, S., Deweer, B., Pillon, B., Vidailhet, M., Dubois, B., & Pouthas, V. (2005). Effects of Internal Clock and Memory Disorders on Duration Reproductions and Duration Productions in Patients with Parkinson's Disease. *Brain and Cognition, 58*(1), 35-48. DOI:10.1016/j.bandc.2005.02.003

Piaget, J. (1970). *The Child's Perception of Time*. New York: Basic Books.

Pollack, I., & Norman, D. A. (1964). A Non-Parametric Analysis of Recognition Experiments. *Psychonomic Science, 1*(5), 125-126.

Poppel, E. (1997). A Hierarchical Model of Temporal Perception. *Trends in Cognitive Sciences, 1*(2), 56-61. DOI:10.1016/S1364-6613(97)01008-5

Predebon, J. (1996). The Effects of Active and Passive Processing of Interval Events on Prospective and Retrospective Time Estimates. *Acta Psychologica, 94*(1), 41-58. DOI:10.1016/0001-6918(95)00044-5

Pricewilliams, D. R. (1954). The Kappa-Effect. *Nature, 173*(4399), 363-364. DOI:10.1038/173363a0

Pylyshyn, Z. W. (1973). What Minds Eye Tells Minds Brain - Critique of Mental Imagery. *Psychological Bulletin, 80*(1), 1-24. DOI:10.1037/h0034650

Pylyshyn, Z. W. (1981). The Imagery Debate - Analog Media Versus Tacit Knowledge. *Psychological Review, 88*(1), 16-45. DOI:10.1037//0033-295X.88.1.16

Pylyshyn, Z. W. (2003). Mental Imagery: In Search of a Theory. *Behavioral and Brain Sciences, 25*(2), 157-+. DOI:10.1017/s0140525x02000043

Richards, W. (1973). Time Reproductions by Hm. *Acta Psychologica, 37*(4), 279-282. DOI:10.1016/0001-6918(73)90020-6

Richman, C. L., Mitchell, D. B., & Reznick, J. S. (1979). Mental Travel - Some Reservations. *Journal of Experimental Psychology-Human Perception and Performance, 5*(1), 13-18. DOI:10.1037/0096-1523.5.1.13

Riemer, M., Trojan, J., Kleinbohl, D., & Holz, R. (2012). A "View from Nowhen" on Time Perception Experiments. *Journal of Experimental Psychology-Human Perception and Performance, 38*(5), 1118-1124. DOI:10.1037/a0027073

Rotello, C. M., Masson, M. E. J., & Verde, M. F. (2008). Type I Error Rates and Power Analyses for Single-Point Sensitivity Measures. *Perception & Psychophysics, 70*(2), 389-401. DOI:10.3758/Pp.70.2.389

Rudner, M., Fransson, P., Ingvar, M., Nyberg, L., & Ronnberg, J. (2007). Neural Representation of Binding Lexical Signs and Words in the Episodic Buffer of Working Memory. *Neuropsychologia, 45*(10), 2258-2276. DOI:10.1016/j.neuropsychologia.2007.02.017

Saj, A., Fuhrman, O., Vuilleumier, P., & Boroditsky, L. (2014). Patients with Left Spatial Neglect Also Neglect the “Left Side” of Time. *Psychological Science, 25*(1), 207-214. DOI:10.1177/0956797612475222

Sarrazin, J. C., Giraudo, M. D., Pailhous, J., & Bootsma, R. J. (2004). Dynamics of Balancing Space and Time in Memory: Tau and Kappa Effects Revisited. *Journal of Experimental Psychology-Human Perception and Performance, 30*(3), 411-430. DOI:10.1037/0096-1523.30.3.411
Schlerf, J. E., Spencer, R. M. C., Zelaznik, H. N., & Ivry, R. B. (2007). Timing of Rhythmic Movements in Patients with Cerebellar Degeneration. *Cerebellum, 6*(3), 221-231. DOI:10.1080/14734220701370643
Schmitter-Edgecombe, M., & Rueda, A. D. (2008). Time Estimation and Episodic Memory Following Traumatic Brain Injury. *Journal of Clinical and Experimental Neuropsychology, 30*(2), 212-223. DOI:10.1080/13803390701363803
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of Measuring Recognition Memory - Applications to Dementia and Amnesia. *Journal of Experimental Psychology-General, 117*(1), 34-50. DOI:10.1037//0096-3445.117.1.34
Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory Prior Entry. *Journal of Experimental Psychology-General, 130*(4), 799-832. DOI:10.1037//0096-3445.130.4.799
Spencer, R. M. C., Ivry, R. B., & Zelaznik, H. N. (2005). Role of the Cerebellum in Movements: Control of Timing or Movement Transitions? *Experimental Brain Research, 161*(3), 383-396. DOI:10.1007/s00221-004-2088-6
Spencer, R. M. C., Zelaznik, H. N., Diedrichsen, J., & Ivry, R. B. (2003). Disrupted Timing of Discontinuous but Not Continuous Movements by Cerebellar Lesions. *Science, 300*(5624), 1437-1439. DOI:10.1126/science.1083661
Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-Analysis. *Journal of Cognitive Neuroscience, 21*(3), 489-510. DOI:10.1162/jocn.2008.21029
Srinivasan, M., & Carey, S. (2010). The Long and the Short of It: On the Nature and Origin of Functional Overlap between Representations of Space and Time. *Cognition, 116*(2), 217-241. DOI:10.1016/j.cognition.2010.05.005
Stevens, S. S. (1957). On the Psychophysical Law. *Psychological Review, 64*(3), 153-181. DOI:10.1037/h0046162
Suddendorf, T., & Busby, J. (2005). Making Decisions with the Future in Mind: Developmental and Comparative Identification of Mental Time Travel. *Learning and Motivation, 36*(2), 110-125. DOI:10.1016/j.lmot.2005.02.010
Tanner, W. P. (1954). A Decision-Making Theory of Visual Detection. *Psychological Review, 61*(6), 401-409. DOI:10.1037/h0058700
Tarantino, V., Ehlis, A. C., Baehne, C., Boreatti-Huemmer, A., Jacob, C., Bisiacchi, P., & Fallgatter, A. J. (2010). The Time Course of Temporal Discrimination: An Erp Study. *Clinical Neurophysiology, 121*(1), 43-52. DOI:10.1016/j.clinph.2009.09.014
Treisman, M., Cook, N., Naish, P. L., & MacCrone, J. K. (1994). The Internal Clock: Electroencephalographic Evidence for Oscillatory Processes Underlying Time Perception. *Q J Exp Psychol A, 47*(2), 241-289.
Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annu Rev Psychol, 53*, 1-25. DOI:10.1146/annurev.psych.53.100901.135114
Tversky, B. (2003). Structures of Mental Spaces: How People Think About Space. *Environment and Behavior, 35*(1), 66-80. DOI:10.1177/0013916502238865
Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-Cultural and Developmental-Trends in Graphic Productions. *Cognitive Psychology, 23*(4), 515-557. DOI:10.1016/0010-0285(91)90005-9
Ulbrich, P., Churan, J., Fink, M., & Wittmann, M. (2007). Temporal Reproduction: Further Evidence for Two Processes. *Acta Psychologica, 125*(1), 51-65. DOI:10.1016/j.actpsy.2006.06.004
Ulbrich, P., Churan, J., Fink, M., & Wittmann, M. (2009). Perception of Temporal Order: The Effects of Age, Sex, and Cognitive Factors. *Aging Neuropsychology and Cognition, 16*(2), 183-202. DOI:10.1080/13825580802411758

Ulrich, R., & Maienborn, C. (2010). Left-Right Coding of Past and Future in Language: The Mental Timeline During Sentence Processing. *Cognition, 117*(2), 126-138. DOI:DOI 10.1016/j.cognition.2010.08.001

Unrugneervoort, A., Kaiser, J., Coenen, A., & Vanluijtelaar, G. (1991). Prospective and Retrospective Time-Estimation. *Perceptual and Motor Skills, 73*(3), 1157-1158. DOI:10.2466/pms.73.8.1157-1158

Vallesi, A., Binns, M. A., & Shallice, T. (2008). An Effect of Spatial-Temporal Association of Response Codes: Understanding the Cognitive Representations of Time. *Cognition, 107*(2), 501-527. DOI:10.1016/j.cognition.2007.10.011

van Rijn, H., & Taatgen, N. A. (2008). Timing of Multiple Overlapping Intervals: How Many Clocks Do We Have? *Acta Psychologica, 129*(3), 365-375. DOI:10.1016/j.actpsy.2008.09.002

VanRullen, R., Pascual-Leone, A., & Battelli, L. (2008). The Continuous Wagon Wheel Illusion and the 'When' Pathway of the Right Parietal Lobe: A Repetitive Transcranial Magnetic Stimulation Study. *Perception, 37*, 83-83. DOI:10.1371/journal.pone.0002911

Verde, M. E., MacMillan, N. A., & Rotello, C. M. (2006). Measures of Sensitivity Based on a Single Hit Rate and False Alarm Rate: The Accuracy, Precision, and Robustness of D', Az, and A'. *Percept Psychophys, 68*(4), 643-654. DOI:10.3758/BF03208765

Vicario, C. M., Pecoraro, P., Turriziani, P., Koch, G., Caltagirone, C., & Oliveri, M. (2008). Relativistic Compression and Expansion of Experiential Time in the Left and Right Space. *Plos One, 3*(3). DOI:10.1371/journal.pone.0001716

Vidalaki, V. N., Ho, M. Y., Bradshaw, C. M., & Szabadi, E. (1999). Interval Timing Performance in Temporal Lobe Epilepsy: Differences between Patients with Left and Right Hemisphere Foci. *Neuropsychologia, 37*(9), 1061-1070. DOI:10.1016/S0028-3932(98)00155-9

Waldum, E. R., & Sahakyan, L. (2013). A Role for Memory in Prospective Timing Informs Timing in Prospective Memory. *Journal of Experimental Psychology-General, 142*(3), 809-826. DOI:10.1037/a0030113

Walsh, V. (2003). A Theory of Magnitude: Common Cortical Metrics of Time, Space and Quantity. *Trends in Cognitive Sciences, 7*(11), 483-488. DOI:10.1016/j.tics.2003.09.002

Wearden, J. H. (2008). Slowing Down an Internal Clock: Implications for Accounts of Performance on Four Timing Tasks. *Quarterly Journal of Experimental Psychology, 61*(2), 263-274. DOI:10.1080/17470210601154610

Wearden, J. H., & Bray, S. (2001). Scalar Timing without Reference Memory? Episodic Temporal Generalization and Bisection in Humans. *Quarterly Journal of Experimental Psychology Section B-Comparative and Physiological Psychology, 54*(4), 289-309. DOI:10.1080/02724990042000173

Wearden, J. H., & Ferrara, A. (1995). Stimulus Spacing Effects in Temporal Bisection by Humans. *Q J Exp Psychol B, 48*(4), 289-310.

Wearden, J. H., Philpott, K., & Win, T. (1999). Speeding up and (... Relatively ...) Slowing Down an Internal Clock in Humans. *Behavioural Processes, 46*(1), 63-73. DOI:10.1016/S0376-6357(99)00004-2

Weger, U. W., & Pratt, J. (2008). Time Flies Like an Arrow: Space-Time Compatibility Effects Suggest the Use of a Mental Timeline. *Psychonomic Bulletin & Review, 15*(2), 426-430. DOI:10.3758/Pbr.15.2.426

Wilson, M. D. (1988). The Mrc Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers, 20*(1), 6-11. DOI:10.3758/BF03202594

Wittmann, M., & Paulus, M. P. (2008). Decision Making, Impulsivity and Time Perception. *Trends in Cognitive Sciences, 12*(1), 7-12. DOI:10.1016/j.tics.2007.10.004

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger Stimuli Are Judged to Last Longer. *J Vis, 7*(10), 21-25. DOI:10.1167/7.10.2

Yates, M. J., Loetscher, T., & Nicholls, M. E. (2012). A Generalized Magnitude System for Space, Time, and Quantity? A Cautionary Note. *J Vis, 12*(7), 1-7. DOI:10.1167/12.7.9

Zach, P., & Brugger, P. (2008). Subjective Time in near and Far Representational Space. *Cognitive and Behavioral Neurology, 21*(1), 8-13.

Zakay, D., & Block, R. A. (1997). Temporal Cognition. *Current Directions in Psychological Science, 6*(1), 12-16. DOI:10.1111/1467-8721.ep11512604