DOCTOR OF PHILOSOPHY

Individual differences in reaction time variability : a combined psychometric and electroencephalographic approach

Saville, Christopher W.N.

Award date:
2010

Awarding institution:
Bangor University

Individual differences in reaction time variability: a combined psychometric and electroencephalographic approach

Christopher W.N. Saville

September 2010.

This thesis is submitted in partial fulfillment for the degree of Doctor of Philosophy,

Work completed in the School of Psychology, Bangor University.

This thesis is dedicated to the memory of my brother Alexander Saville

(1st December 1986 – 3rd January 2007)

Acknowledgements

The work described in this thesis is the result of four years of collaboration with many people, who I would like to recognise here. First and foremost, I would like to thank Christoph Klein for being such a helpful, knowledgeable, and agreeable supervisor. I consider myself extraordinarily lucky to have had the opportunity to learn from him, and I hope to be able to continue to collaborate with him for years to come.

I am very grateful to Dave Daley for being my second supervisor. Dave has been a source of timely, helpful, and pragmatic advice throughout my PhD, and played a vital role in keeping the thesis on track. I am sure Bangor that will greatly miss him now he has moved on to Nottingham.

I was also lucky enough to have my undergraduate supervisor, James Intriligator, as a chair of committee. I am told that the chair is often a purely administrative role but James was an endless source of support and advice, and acted as a third supervisor throughout the PhD.

In addition to my supervisors, I am very grateful to Bernd Feige and Stephan Boehm, who collaborated with us on the EEG papers. They provided all sorts of helpful suggestion for my data analysis and I learned a lot from them both.

Through my PhD, I worked with many highly competent MSc students. I would particularly like to acknowledge the help of those who supported the projects described in this thesis: Ralph Pawling and Mark Trullinger (Chapter Six); Reem Dean (Chapter Seven); Sailee Shikare and Sarayu Iyengar (Chapter Nine).

I am indebted to too many other people to describe their help in detail, but I would like to mention Jospeh Dien, Nick Peatfield, Nick, Oosterhof, Bruce Keefe, John Taylor, Corinna Haenschel, Nick Davis, Helen Morgan, Helma de Moree, Janet Bultitude, Ingmar Gutberlet, Chris Whitaker, Kevin MacKenzie, Julia Müller, Caroline Junge, Jan-Rouke Kuipers, David Hill,

Amy Hulson-Jones, and David Ingledew. I have benefitted from working in such a friendly and community-spirited department, and am very grateful. I would also like to thank my parents, Tim and Jenny Saville, for their support and indeed their proof-reading.

Finally I would like to express my gratitude to all of my participants for their patience and concentration.

CONTENTS

Abstract

Reaction times (RTs) have been of interest to empirical psychology for as long as the discipline has existed. However, most studies summarise RTs using measures of central tendency, ignoring high levels of intra-subject variability (ISV). ISV has recently become an important topic in neuroscience and differential psychology in its own right, showing strong associations with intelligence, cognitive ageing, and several psychiatric and neurological disorders. The present thesis describes a programme of research into the measurement of individual differences in ISV, using a mixture of psychometric and psychophysiological techniques.

Study One explores measurement issues, comparing various metrics of ISV in terms of their reliability and statistical redundancy. Study Two compares the single-trial event-related potentials of participants with high and low levels of ISV on a working memory oddball task. Study Three looks at the cross-task and cross-modal structure of ISV, using a latent variable approach. Study Four integrates approaches from the previous three chapters, deriving supra-task latent variables for ISV and several single-trial P3b parameters, and exploring the latent variable correlations between these constructs.

These studies suggest that a latent variable framework may be a promising framework for studying ISV, not just in behavioural studies but in also psychophysiological research. Possible implications for future research are discussed, based on the empirical and theoretical portions of the thesis.

Chapter One

An Introduction to Reaction Times

One of the principal aims of psychology has long been to gain insight into the black box of human cognition. In recent years psychologists and neuroscientists have made huge strides by using neurophysiological techniques to measure the brain correlates of mental operations. However these approaches, fascinating as they are, have not, and probably cannot hope to, replace measurements of unambiguously observable behaviour. These direct measurements of behaviour form the foundations of empirical psychology, and are evidence that other measures we take are functionally relevant and not mere epiphenomena of cognition.

One of the oldest and most important of these measures of cognition is reaction time (RT). An RT is a measurement of the latency between a cue to respond, generally the onset of a stimulus, and a participant's response. Assuming that the participant attempts to respond promptly, an RT represents a measurement of the speed of a cognitive operation. Franciscus Cornelius Donders, one of the fathers of empirical psychology, is generally credited with introducing the use of RTs to the field in 1868 but the idea of RTs is older still. The concept is clearly present in the writing of the Islamic polymath Abu Rayhan Biruni at the turn of the second millennium AD:

> *"Not only is every sensation attended by a corresponding change localized in the sense-organ, which demands a certain time, but also, between the stimulation of the organ and consciousness of the perception an interval of time must elapse, corresponding to the transmission of stimulus for some distance along the nerves."* (Biruni, as cited in Iqbal, 1930.)

While the latency Biruni described was between sensory stimulation and conscious perception, and thus was not an RT in the strictest sense of the word, Biruni captured most aspects of the modern conceptualisation of an RT. While Biruni explicitly attributed this latency to nerve conduction, the transformation he described between the stimulation of a sensory organ to perceptual consciousness suggested cognition. It is this aspect of an RT that most interests

empirical psychology. Measuring the speed at which a participant can carry out a cognitive operation is generally taken as a measurement of the complexity of this operation, and of the processing speed of the participant. The early work on RTs found that when participants had to respond differentially to two types of stimuli, implicitly requiring stimuli to be distinguished from one another and classified, they took longer than when they merely had to respond to the presence of a stimulus (Donders, 1868). This approach laid the bedrock of empirical psychology for years to come; by manipulating the demands of a task and recording the changes in RT, inferences could be drawn regarding the complexity of the cognitive operations underpinning a task (see Posner, 1976). While simple, this paradigm proved remarkably effective and can probably be credited with no small amount of the progress made on research into cognition over the last century.

A standard experiment does not collect just one measurement of RT, rather the task is repeated a number of times. The logic here, according to classical test theory (Novick, 1966), is that individual RTs are unreliable and noisy, but that by aggregating a number of RTs a reliable measurement can be made (Spearman, 1910; Brown, 1910). The most obvious way to aggregate several RTs is a measure of central tendency such as taking the mean or median RT. Such a measure gives an indication of how long the average RT was and thus gives a robust measure of the speed of a cognitive process. An assumption implicit in this approach is that the variation in the individual RTs is residual noise and measurement error, which must be cleaned away from the data before it can be safely interpreted. This assumption, however, hides a hugely important question.

Consider a participant in the simplest kind of RT task: the participant sits watching a computer monitor with their finger on a response key. A light appears on the screen periodically and the participant's task is to wait until they sees the light and then to press the response key as fast as possible. The situation is identical in every trial of the task, the stimulus is identical every

time, and the response demands are fixed. Nevertheless, RTs vary from trial to trial and these variations are substantial. The standard deviation of an individual's raw RTs on such a task might easily be 60ms. This sounds like a trivial period of time but with a median RT for such a task in the region of 250ms[1], this level of variability is notable. Such a level of variability is certainly beyond what might be expected as a result of instrumental error. Additionally mean condition effects have been shown to explain only around 10% of intraindividual variance in RTs in a number of simple decision tasks (Gilden, 1997). The scale of intra-subject variability in RTs (ISV), relative to average RT and experimental effects on RT, suggests that dismissing such variability as meaningless noise is unconvincing.

To summarise, the use of RTs has provided a window onto cognition for over a century. The use of measures of central tendency alone, however, hides a great deal of variability, which will be the subject of this thesis. Chapter 2 will discuss measurement issues regarding ISV, Chapter 3 will cover literature on the relevance of ISV to psychology and neuroscience, Chapter 4 will review evidence on the neural bases of ISV, and Chapter 5 will introduce electroencephalography, a key approach used in this thesis. Chapters 6 to 9 will then describe the empirical work undertaken over the project, before Chapter 10 discusses the thesis as a whole.

---

[1] These figures are based on data reported in Chapter Eight of this thesis.

Chapter Two

Taxonomic and Measurement Issues in the Study of Intra-Subject Variability of Reaction Times

Before reviewing the empirical literature, it is important first to identify the different types of ISV, and how they can be quantified. This chapter will first cover some taxonomic assumptions of the field, before moving on to describe the issues surrounding the measurement of ISV.

Taxonomy

It is possible to conceive of more than one form of ISV, and taxonomic distinctions between these different kinds of variability are crucial to research. One strand of ISV research, stemming mainly from the developmental field, categorises variability as diversity, dispersion, and inconsistency (Hultsch, MacDonald, & Dixon, 2002). Diversity is variability on the group level; dispersion is variability in an individual's level of performance on different tasks; while inconsistency is variability on the same task on multiple occasions or trials. Inconsistency is thus the type of variability with which this thesis will be concerned. It is, however, interesting to note that certain groups that show increased ISV, such as older adults (Hultsch *et al.*, 2002) and ADHD patients (Klein, Wendling, Huettner, Ruder, & Peper, 2006), also show increased diversity and dispersion.

Fiske and Rice (1955) suggested the earliest taxonomic system of ISV. They classified Type One variability, or spontaneous variability, as where two observations of behaviour, taken in objectively identical circumstances, varied. The authors explicitly state that the order of the two trials must be immaterial. Type Two variability, on the other hand, is variability where observations show some sort of temporal pattern. Fiske and Rice claim that such variability is usually reactive. A final form of variability, Type Three, contains all instances of variability where the situation varies in a dimension other than time, and is thus comparable to Hultsch *et al*'s (2002) concept of dispersion.

It is probably fair to say that modern taxonomic assumptions of the field have quietly dropped the idea that ISV with some form of temporal structure is non-spontaneous. As will be discussed in Chapter Four, neural activity shows apparently endogenous oscillations. It is however still worth making the distinction between periodic and aperiodic conceptions of ISV because they differ markedly in how they are measured. Variability which is assumed to be aperiodic is generally measured using distributional approaches, while periodic variability is measured using time series methods.

Distributional Approaches to Reaction Times

As mentioned in the previous chapter, many RTs are recorded in the course of experimental tasks, and so these RTs can be understood as a statistical distribution. The most often used distribution is the *Gaussian* or *normal* distribution. Gaussian distributions can be summarised perfectly using only two independent parameters: *mu*, representing the central tendency of the distribution (its position on the *x*-axis), and *sigma*, representing the degree of dispersal around this point. A distribution with a low sigma value shows little variance around mu while one with a large sigma value shows great variation around mu. In empirically observed distributions, however, mu and sigma are not known and instead must be estimated. The most common way of doing so is to take either the mean or median value as a proxy for mu, and to use the standard deviation of the values as a proxy for sigma. Accordingly, the most common approach in ISV research has been to employ standard deviations of reaction times (SDRT) as a measure of variance.

While the use of standard deviations has proven largely successful, it is not without problems. One key issue is that the median and standard deviation of RT distributions tend to be substantially positively correlated: participants who produce high median RT (MnRT) also tend

to produce high SDRT (Wagenmakers & Brown, 2007). This raises the possibility that measurements of ISV are implicitly also measurements of average RT. Some researchers attempt to remove the variance shared with the mean using statistical corrections. This can be accomplished in a number of ways. One possibility is to divide the standard deviation by the mean to produce a coefficient of variation: the amount of variance that exists for each unit of the mean (CVRT; *e.g.* Flehmig, Steinborn, Langner, Scholz, & Westoff, 2007). Alternative approaches involve regressing or partialling out the variance associated with the mean, or variables which may affect the mean such as age, (*e.g.* Bunce *et al.*, 2007). Other researchers argue that attributing shared variance to the mean is merely another example of the implicit primacy given to measures of the mean that is so common in empirical psychology. In contrast, Jensen (1992) makes a case for the primacy of the variance. RT distributions, he notes, are effectively capped at the fast tail by basic physiological limits – nerve conduction velocity and movement time – but uncapped at the slow tail. Thus a mean RT is often determined to a great extent by the slower RTs in a distribution (*cf.* Larson & Alderton, 1990), namely the variance.

The perspective of this thesis on the matter is that attributing all variance shared between median RT and SDRT to the median is largely theoretically unjustified, and is primarily done for historical reasons: the perceived primacy of the median over SDRT. However, while it is possible to argue that ISV is a more fundamental measure than median RT (*eg.* Jensen, 1992), such disputes are largely unproductive. Recognising the importance of the higher moments of the RT distribution need not de-emphasise the importance of the first moment.

As stated above, the mu and sigma of a Gaussian distribution are formally independent. The mu should tell one nothing about the sigma. The mean and standard deviation of RT distributions, however, are highly correlated. This raises an important point: RT distributions are in fact not Gaussian. RT distributions instead exhibit positive skew: a tendency for the 'centre of mass' of the distribution to be above the modal value.

This poses an issue for using standard deviations for the measurement of ISV. Standard deviations treat deviation from the mean in both directions identically, a property which may be inappropriate for skewed RT distributions. There are three main approaches to this problem: alternative distributions, statistical transformations, and what could be described as the 'agnostic' approach.

Alternative distribution approaches address the non-Gaussianicity of RT distributions by fitting distributions other than the Gaussian to RT data. A number of distributions have been used but the most popular is the ex-Gaussian. An ex-Gaussian is the convolution of a Gaussian distribution, summarised by the parameters of mu and sigma, and an exponential distribution, summarised by the parameter *tau*. The distribution obtained by convolving these components has the characteristic rightward skew shown by RTs, and appears to be a good fit for RT data. What is less clear is whether the three parameters of the ex-Gaussian represent physiologically distinct properties of RTs, or whether the ex-Gaussian is merely superficially similar to an RT distribution. This remains an open question.

Statistical transformations represent an alternative approach to handling RT distributions. These involve mathematical operations on groups of raw RTs that render the distribution Gaussian, while preserving other information. Of course transforming data merely so that it fits a normal distribution is questionable and any such transformation is only as worthwhile as its rationale is plausible. One successful approach to RTs, coming from the eye movement literature, is the reciprocal transformation. Carpenter (1981) argues that viewing RTs as latencies is somewhat erroneous and that they are better understood as representing the rate at which a process occurs. Rates are conventionally measured using reciprocals and, when reciprocally transformed, RT distributions are indeed approximately Gaussian.

It is worth noting here that the ex-Gaussian and reciprocal approaches both provide good fits to RT data, but imply different underlying processes. Carpenter's reciprocal approach

suggests a single, essentially Gaussian, process underlying RT while the ex-Gaussian approach implies that two separate processes, one Gaussian and one exponentially distributed, underpin RTs. Thus while obtaining a good fit to the data is clearly a necessary criterion for any distributional model of ISV, showing superficial similarities between an empirical and a theoretical distribution does not prove that the two distributions are underpinned by similar dynamics.

The 'agnostic' approach to non-Gaussianicity rests on the argument that, while means and standard deviations have many interesting and statistically desirable qualities when the underlying distribution is Gaussian, this does not necessarily invalidate their use in summarising non-Gaussian distributions. The use of a standard deviation, in short, does not necessarily rely upon Gaussianicity. Indeed a standard deviation will quantify the variance in a sample of data regardless of the distribution, which is not necessarily the case for other ISV parameters. While arguably less analytically appropriate, this stance has a number of advantages. The type of distribution from which RTs are drawn remains unknown. Indeed it is reasonable to speculate that different types of task may produce different distributions of RTs. Furthermore, as noted above, mutually exclusive models of RT distributions can both provide adequate fits to the same data. This being the case, when simply seeking to quantify variance, it may be better to use a robust, well understood parameter such as the standard deviation, rather than using a more theory-laden measure such as tau. This agnostic approach has been taken by most ISV researchers – albeit generally more implicitly than explicitly.

Time Series Approaches to Reaction Times

Distributional approaches to RTs assume that each RT is drawn at random from a distribution, an idea that assumes that the order of RTs is essentially immaterial. RTs can also be

understood as a time series, with systematic trends in RTs over time. The distributional approaches described above assume no temporal structure, and so different analysis methods must be used in order to identify such trends.

The simplest kind of time series in RT data is a linear trend. As a participant becomes more used to a task, they may become faster at responding. Alternatively, they may become tired and bored, leading to slower RTs. Such time-on-task effects, for example due to practice or fatigue, are clearly reactive, in the language of Fiske and Rice (1956), and as such do not reflect the kind of spontaneous variability that ISV is thought to represent. Indeed such linear trends are often removed from data before the data are analysed further, using regression-based approaches (*e.g.* Hultsch, MacDonald, & Dixon, 2002).

Other methods can identify periodicities in time series data that are more complex than linear trends. One such approach is the autocorrelation. Given n observations of two variables ($x1....xn$ and $y1....yn$), the extent to which the two variables co-vary can be obtained by computing a correlation coefficient, a standardised measure of covariance. This will give an indication of whether the two variables tend to vary together across observations, and can help to indicate whether a relationship exists between x and y. It is also possible to explore the extent to which the variable x is related to previous observations of itself, by correlating ($x1$, $x2$), ($x2$, $x3$),..., (xn-1, xn). Such a correlation is termed the lag one autocorrelation. Autocorrelations can be computed for any lag greater than one (a zero lag autocorrelation will of course always give $r = 1$), although Chatfield (1980) suggests that autocorrelations where lag > $n/4$ are not generally useful.

Autocorrelations are a useful way of assessing the periodic structure of a time series. A genuinely random time series (often called *white noise*) will show an autocorrelation of $r = \sim 0$ at all lags, while the presence of temporal structure is suggested by non-zero autocorrelation coefficients. RT distributions typically show at least short term autocorrelation, indicative of

autoregressive structure (Kelly, Heathcote, Heath, & Longstaff, 2001). There are however a number of issues to consider when using autocorrelations, raised by Chatfield (1980). Firstly, autocorrelations, like any other use of correlation, can give spuriously high values by chance. The use of a series of autocorrelations at various lags across a number of participants increases the risk of finding high correlation coefficients by chance. Secondly, correlations are vulnerable to extreme outliers and autocorrelations are, if anything, particularly vulnerable. Outliers will generally depress the strength of association between two time series, but the existence of more than one extreme value can lead to a spurious autocorrelational peak or trough at the lag that 'pairs' the two extreme values. Thirdly, the existence of a non-stationary trend, such as the linear trends described above, can inflate the strength of autocorrelations as an observation that deviates from the mean in one direction will tend to be followed by other values that deviate in the same direction. This renders the apparent autocorrelational structure misleading.

An alternative approach to identify periodicities in a time series is by use of Fourier analysis. The Fourier transform, based upon Jean-Baptiste Joseph Fourier's pioneering 1822 work into the conduction of heat, decomposes a time series into sets of sine and cosine waves of various frequencies. Convolution of a time series with a sine wave of a certain frequency will only yield non-zero values if the time series contains periodicities at the frequency in question. Thus convolving a time series with sine waves of many different frequencies will identify which frequencies are present in a time series, and how much of the variance in the time series each accounts for. Modern Fourier algorithms, such as the Fast Fourier Transform (*e.g.* Cooley and Tukey, 1965), provide more efficient ways of decomposing the signal, sparing researchers from manually convolving a signal with many different sine waves.

Fourier analytic approaches to RT data have begun to appear in the literature in recent years (*e.g.* Johnson, *et al.*, 2007), and may be a helpful way of identifying the frequencies that

contribute most to overall ISV. It is worth stating that autocorrelation and power spectra contain the same information, and simply differ in the way they express it (Chatfield, 1980).

Two major assumptions of Fourier analysis are that the signal should be infinite in duration, and that it should be stationary. The former is clearly an impossible criterion for any real signal to meet, as thus is routinely violated, but the duration of the signal does place a limit on the frequency resolution. The assumption of stationarity is again possible to violate, but can lead to misleading results, as the spectra produced by a Fourier transform does not give any indication of how a signal is evolving over time. These two related assumptions are the two sides of the trade-off between resolution in the time and frequency domains – the result of a kind of *uncertainty principle* in Fourier mathematics. These assumptions do not invalidate the use of Fourier approaches with RTs, but the possibility of the RT series being too short or non-stationary should be assessed before using these approaches.

Reliability of Reaction Time Parameters

Another important issue is that of psychometric reliability – the proportion of variance in an observed variable that represents the *true score* of the underlying latent variable rather than error variance (Observed score = True score + Error; see Spearman, 1910; Brown, 1910; & Novick, 1966). Reliability is often tested by taking two measurements of the same trait and seeing to what extent the two measures differ. Assuming the true score remains the same, any difference will be due to error variance, so the extent to which the measurements vary is an index of the relative proportions of true score variance and error.

Reliability is especially important to ISV as the latter was long regarded as mere noise. Stable individual differences in ISV are important in establishing ISV as a *bona fide* cognitive trait construct. Reliability is also important beyond establishing construct validity for ISV, it is

also a key issue in measurement. The reliability of a variable determines the strength of association which can be drawn between it and other variables. Thus besides satisfying the criterion of construct validity, a metric of ISV must also exhibit acceptable reliability.

A factor influencing reliability is the number of homogeneous trials a metric is derived from. This is also the case with means and medians and is predominantly a matter of data aggregation: basing a summary statistic on a greater number of observations of the same process or phenomenon leads to more robust measurement (Spearman, 1910; Brown, 1910). However measures of variability, by definition, require multiple measurements; if it were possible to record a perfectly accurate RT with one trial (observed RT = true RT + error), this would suffice to measure speed of reaction but clearly would not suffice to measure ISV. Estimating higher moments of distributions requires more observations than relatively simple measures of central tendency, and the number of trials used to measure MnRT may not be appropriate for measuring SDRT.

Another sampling issue, primarily for time series approaches, is the Nyquist-Shannon sampling criterion (Nyquist, 1928; Shannon, 1949). Oscillations can of course occur on an essentially infinite number of scales and to identify an oscillation it is necessary to sample with sufficient frequency, and to record for a sufficient duration. Specifically, the sampling rate must be greater than twice the bandwidth of the highest frequency in the signal. These requirements are formalised by Shannon's proof (Shannon, 1949). To apply the same logic to ISV, certain oscillations in attention will be too high in frequency to be captured by the rate of stimulus presentation that is feasible in an RT experiment. Likewise short experiments will fail to capture slow oscillations. Therefore, even where the level of data aggregation provides reasonable reliability, a task may sample too sparsely or be too short to capture certain important dynamics. While this issue is predominantly an issue for Fourier approaches, it is also worth considering as a theoretical issue for distributional approaches. The ISV dynamics captured by a twenty trial

task may be qualitatively different to those captured by a five hundred trial task. Even if a task has adequate reliability in a psychometric sense, it may not be measuring the phenomena that underlie low frequency RT oscillations.

Conclusion

The reliability and equivalence of different metrics of ISV will be taken up empirically in Chapter Six. However, to summarise: a raw RT distribution is non-Gaussian, and a number of approaches exist for dealing with this. These approaches, except the agnostic approach, carry mutually exclusive assumptions about the process underlying RTs, but it is difficult to make a conclusive *a priori* case for which is most appropriate, as such a case essentially comes down to construct validity. RTs can also be understood as a time series, autocorrelational and Fourier analytic approaches being the most common approaches used to accomplish this. An important characteristic for ISV is psychometric reliability, without which, from a trait-theoretic perspective, such metrics are theoretically irrelevant and practically useless. Psychometric reliability demands a certain level of data aggregation, but further demands are placed on the sampling rate and duration of RT tasks by Nyquist's theorem.

Chapter Three

The Relevance of Intra-Subject Variability in Reaction Times

Chapter One addressed reasons why ISV is an interesting area for research, namely the high variance of RTs relative to both median RTs and experimental effect sizes, as well as the more fundamental question of why RTs vary at all. Chapter Two focused on measurement, the summary statistics that best characterise the RT distribution. There is, however, a difference between identifying a *prima facie* reason why ISV might be philosophically interesting and showing that ISV can be mathematically well characterised on the one hand, and on the other hand demonstrating that ISV is relevant to empirical psychology at large. This chapter will review three areas of research where ISV has become a key topic: intelligence, psychiatry, and gerontology. A notable omission here is the brain lesion literature, which will be covered in the following chapter on the neural bases of ISV.

ISV and Intelligence

Intelligence was probably the earliest psychometric trait to be linked to ISV. Baumeister and Kelas (1968) compared a group of healthy student volunteers to a group of men from a 'residential institution for the retarded' on a warned RT task and found significantly increased MnRT and SDRT in the intellectually disabled group. The authors drew a distinction between a participant's optimal level of performance and their ability to maintain performance at that level, and suggested that ISV may be at least as important as median RT in understanding impaired performance.

This link between inconsistent performance and intellectual ability was substantiated and extended by later work. Jensen (1982, 1992) reported that not only were correlation coefficients between ISV and IQ substantial, they were in many cases larger than those between MnRT and IQ. This was in spite of the generally poorer psychometric reliability of ISV as compared to MnRT.

Larson and Alderton (1990) took a different approach by dividing RTs into 16 quantile bands, averaging the RTs in these bands, and correlating all 16 scores with the participants' IQ scores. The correlation coefficient between the fastest band, measuring optimal performance, and IQ was modest (rho=.20) but coefficients increased steadily with successively slow bands, reaching their maximum for the final band (rho=.37). Thus when participants were grouped by IQ, RTs for the fastest bands were similar but diverged at slower bands. These results imply that the correlation between IQ and ISV reported by other authors was primarily driven by the slower RTs, supporting speculation by Baumeister and Kellas (1968) that participants with low IQ may not be impaired in their optimal level of performance so much as their inability to maintain performance at this level. Larson and Alderton's findings were dubbed the worst performance rule. This pattern of results has been replicated by several groups, using various tasks, with IQ showing a generally stronger relationship with a participant's slowest RTs than with their fastest (Kranzler, 1992; Diascro & Brody, 1993; Coyle, 2001; but see Salthouse, 1998 for a counterexample. See Coyle, 2003 for a review).

Schmiedek, Oberauer, Wilhelm, Süβ, and Wittmann (2007) approached the relationship between general mental ability and separable features of the RT distribution in another way, but arrived at similar conclusions. They fitted an ex-Gaussian distribution to RT data, and found that it was tau, and thus the slow tail, that best correlated with differences in general mental abilities.

The picture emerging from the IQ literature suggests that individuals with low IQ and individuals with high IQ may be fairly similar in their optimal level of speed of processing, but that those with low IQ have difficulty maintaining this level of performance. Thus measurements that are sensitive to failures in performance maintenance are also sensitive to IQ differences. The study of ISV may have important implications for how we conceptualise and measure IQ.

ISV and Biological Psychiatry

The second area where ISV has become a key topic is biological psychiatry. At present psychiatric diagnoses are made on the basis of clinician-appraised symptoms. These symptoms are, according to a widely held opinion in biological psychiatry, necessarily somewhat removed from the biological underpinnings of the disorders. The hope, therefore, is that by focussing on neurocognitive and physiological differences between probands and healthy controls, it will be possible to identify characteristic features of psychiatric disorders on a scale closer to their neural bases. These features could then serve both to improve understanding of the pathophysiology of these disorders and as biological markers of a disorder's presence and its severity. One such putative marker of several psychiatric conditions is increased ISV. A broadening of the RT distribution has been found in schizophrenia (Birkett *et al.*, 2007), chronic fatigue syndrome (Fuentes, Hunter, Strauss, & Hultsch, 2001), bipolar disorder (Brotman, Rooney, Skup, Pine, & Leibenluft, 2009), and attention deficit hyperactivity disorder (ADHD; Klein, Wendling, Huettner, Ruder, & Peper, 2006), relative to healthy controls. The majority of research in this area, however, has been on schizophrenia and ADHD, and so this is where this section will focus.

ISV has long been known to be elevated in patients with schizophrenia (Shakow, 1977), a finding which appears to be specifically linked to the syndrome of schizophrenia, rather than psychotic symptomology (Schwartz *et al.*, 1989). Interestingly there is evidence that while schizophrenic patients and their first-degree relatives both show increased mean RT, only patients show an increase in ISV (Birkett *et al.*, 2007). From the analysis by Birkett *et al.* of the distribution of RTs, it appears that the difference between patients and non-patients is strongest in the slowest RTs, in a manner reminiscent of the worst performance rule described above. It is worth noting, however, that this effect does not appear to be driven by intelligence, as Rentrop *et al.* (2010) found differences in ex-Gaussian tau between high functioning schizophrenic patients

and controls in the absence of differences in intelligence. This effect is also not due to mean RT, as Birkett *et al.* (2007) found differences between probands and controls in CVRT.

In addition to differentiating between those with schizophrenia and those without, ISV appears to respond to psychopharmacological treatment. Cleghorn, Kaplan, Szechtman, Szechtman, and Brown (1990) found reduced ISV (the metric used was unspecified) in patients treated with neuroleptic medication, relative to untreated patients. Both groups of patients, however, still showed higher ISV than controls.

Likewise, ADHD has been linked to elevated ISV by a number of studies. Klein *et al.* (2006) found that ISV, measured by SDRT and coefficient of variation, better discriminated between patients and controls than a number of other measures, including commission errors, omission errors, and mean RTs. Again, similarly to schizophrenia, this effect appears to be driven largely by the slower RTs in the distribution, with a number of studies (Leth-Steensen, King Elbaz, & Douglas, 2000; Hervey *et al.*, 2006; Buzy, Medoff, & Schweitzer, 2009) showing increases in ex-Gaussian tau. Williams, Strauss, Hultsch, Hunter, and Tannock (2007) argue that increased variability is present in both the fast and slow end of ADHD patients' RT distributions, but their findings were based on a very small number of trials and it is not clear if these findings would apply to larger datasets.

There is also strong evidence for at least a partial normalisation of ISV following treatment with psychostimulants. Spencer *et al.* (2009) showed a reduction in skewness of RT distributions following treatment with methylphenidate. Castellanos *et al. (*2005) found that ISV, measured both as standard deviation of reaction times and as power in a frequency band centred on .05Hz, was attenuated by treatment with methylphenidate, relative to treatment with placebo. ISV (metric unspecified) appears to be sensitive to therapeutic approaches other than pharmacotherapy, showing reduction following neurofeedback training (Egner & Gruzelier, 2004). Interestingly however, despite ISV showing a response to treatment, there is evidence that

high ISV is predictive of poor treatment response to methylphenidate, Lee *et al.* (2009) found that high SDRT at baseline was predictive of non-response to treatment, despite showing that ISV was reduced by treatment in both responders and non-responders.

ISV and Ageing

Another area of research where ISV has proved to be important is the study of cognitive ageing. Both cross-sectional (Hultsch, MacDonald, & Dixon, 2002) and longitudinal studies (Lövdén, Li, Shing, & Lindenberger, 2007) have found increasing CVRT with advancing age, while Burton, Strauss, Hultsch, Moll, and Hunter (2006) found a relationship between the severity of cognitive decline and ISV, suggesting that ISV may be a measure of cognitive ageing. What makes ISV particularly intriguing as a measure of cognitive ageing is that ISV increases appear to precede and predict subsequent declines in other measures of neurocognitive performance (Lövdén et al., 2007), suggesting that ISV may be a useful 'canary down the mineshaft'. More dramatically, increases in SDRT appear to predict impending death in longitudinal designs (MacDonald, Hultsch, & Dixon, 2008) suggesting that ISV may be an early indicator of terminal decline, although it must be stated that high ISV appears to be predictive of mortality in young as well as older adults, suggesting that this predictive characteristic may not purely be an index of cognitive decline (Shipley, Der, Taylor, & Deary, 2006).

The relationship between ISV and age is also not a linear one. Williams, Hultsch, Strauss, Hunter, and Tannock (2005) found a quadratic relationship between ISV and age, with young children and older adults exhibiting higher ISV than young adults. It remains unclear, however, whether the causes for the high ISV at both extremes of age are the same.

Overview

The emerging picture from the intelligence, psychiatric, and gerontological literature is that ISV appears to be sensitive to a number of important conditions, and herein lies the problem. The lack of specificity of high ISV renders it unlikely to be useful as a diagnostically useful marker. There is a possibility that emerging approaches to ISV, such as model fitting and spectral approaches, may reveal differences between the increased ISV seen in different conditions, but these differences, if they indeed exist, have yet to be identified. On a brighter note, the study of ISV may be helpful in identifying neurophysiolgical characteristics which these different conditions share. Indeed the specificity criterion may be unreasonable if nosologically distinct disorders overlap etiologically.

What ISV lacks in specificity, it makes up for in sensitivity. ISV appears to be a highly sensitive early indicator for cognitive decline (Lövdén et al., 2007; MacDonald, Hultsch, & Dixon, 2008), with increases in ISV preceding more traditional indicators of cognitive decline. ISV appears to be sensitive to treatment (Cleghorn *et al.*, 1990; Spencer *et al.,* 2009; Egner & Gruzelier, 2004) and symptom severity (Burton *et al.*, 2006), which suggest a possible role for ISV in measuring the time and course of a disorder.

One intriguing possibility is that ISV is an index of some general neurocompuational function of the brain, which is aberrant in a number of conditions. Put another way, ISV could reveal commonalities between different psychiatric and neurological states. Writers in the ageing (Li, Lindenberger, & Frensch, 1999) and the schizophrenia literature (Winterer *et al.*, 2006) have posited disordered dopaminergic neuromodulation as a possible neural substrate of ISV, which would fit with certain models of pathophysiology in these conditions. The subject of the biological bases of ISV will be taken up in the next chapter.

To conclude, ISV research has moved beyond an abstract effort to understand neurobehavioural determinism and has become a topic of great interest in a number of clinical and applied psychometric research domains. ISV appears to err on the side of sensitivity, as opposed to specificity, being a marker for a number of psychiatric conditions, cognitive ageing, and low IQ. This may point to generalities between these conditions, and the study of ISV may help to illuminate the physiological underpinnings of more general computational properties of the brain.

Chapter Four

Neural Bases of Intra-Subject Variability of Reaction Times

In the previous chapter it was claimed that ISV represents a promising line of enquiry as to the neural underpinnings of a number of psychiatric conditions, cognitive aging, and intelligence. Such promise, however, can only be realised if it is clear what the neural underpinnings of ISV itself might be. There is a growing neuroscientific literature on the topic, which this chapter will review. The relationship between neural structure and ISV will be addressed, before findings on ISV and haemodynamic, electrophysiological, psychopharmacological, and computational neural functioning are reviewed. The major theoretical models of ISV will then be covered in the final section of the chapter.

Empirical Findings

Structural neuroanatomy

Probably the most well documented structural finding on ISV is the relationship between ISV and brain lesions. Stuss and colleagues (1989) found that traumatic brain injury patients showed higher SDRT than matched controls on a number of tasks. There is evidence that increased ISV is specifically associated with injuries to the frontal lobes. Stuss, Murphy, Binns, and Alexander (2003) compared patients with frontal lesions to those with more posterior injuries and found that abnormal RTSD and CVRT were only present in frontal lobe patients. This finding is supported by Bunce *et al.* (2007), who find that white matter hyperintensities are associated with increased mean-independent variation (mean absolute variation with mean RT regressed out), but only when they are located in frontal regions.

The second major structural finding is an apparent link between white matter volume and ISV. Walhovd and Fjell (2007) measured white matter volume using magnetic resonance imaging (MRI) in a sample of 71 healthy participants. White matter volume was negatively correlated with SDRT, independently of mean RT. Mean RT itself showed a trend level correlation with

grey matter, suggesting that mean RT and SDRT may have partially separable neural bases. Anstey *et al.* (2007) found a link between the volume of the corpus callosum, a large white matter structure, and mean independent variation in participants with mild cognitive impairment. The same relationship was not present in healthy controls.

Functional neuroanatomy

Haemodynamic imaging

Moving to functional neuroimaging research, the haemoimaging literature on ISV fits well with the lesion literature. Two studies used functional MRI to study the relationship between haemodynamic measures and ISV. Bellgrove, Hester, and Garavan (2004) found a positive correlation between CVRT and activation in medial-frontal, thalamic, and parietal regions. Simmonds and colleagues (2007) found a similar positive relationship between CVRT and prefrontal and caudate areas. Both sets of authors interpret these results as a greater reliance on prefrontal executive and inhibitory function to maintain adequate performance in participants with high ISV than those with more stable RTs. It should of course be noted that the go-no go tasks employed by these studies primarily tap inhibitory and cognitive control, and as such the relationship between ISV and inhibition suggested by these studies may not be specific as a shallow reading of this literature would suggest.

Electrophysiology

A number of studies have related ISV to electrophysiological measures of cognition. Segalowitz, Dywan, and Unsal (1997) found a negative correlation between SDRT (mean RT regressed out) and the amplitudes of the P3 and contingent negative variation components of the event-related potential (ERP) in brain injured patients, but not in healthy controls. Di Russo and

Spinelli similarly found reduced P3 amplitude and increased CVRT in boxers (putatively due to their exposure to chronic head trauma) relative to non-athletic and fencer control groups.

Other electrophysiological studies have examined ISV using non-ERP methodologies. McIntosh, Kovacevic, and Itier (2008) examined the relationship between behavioural variability and neural variability (measured using multiscale entropy) and found, somewhat paradoxically, that they were negatively correlated. They argue that this increase in neural variability does not represent noise so much as complexity, and that this complexity provides a level of behavioural metastability. Gerson, Parra, and Sajda (2005) approached ISV from a different angle by investigating evoked components which discriminated between trials where a target was present and trials where it was absent on individual trials of a rapid serial visual presentation task. They identified a component, which they suggested was similar to a P3, which covaried in latency with RT.

Psychopharmacology

Several studies have linked ISV to the catecholaminergic neuromodulatory systems. MacDonald, Cervenka, Farde, Nyberg, and Bäckman (2009) measured dopamine D2 receptor binding using positron emission tomography and found that SDRT was negatively associated with extrastriatal D2 binding. Stefanis *et al.* (2005) employed a gene association approach to examine the influence of catecholamines on ISV, and found that an allele associated with a higher rate of catecholamine metabolism (the val allele of the COMT Val^{158}Met polymorphism) led to higher levels of SDRT in its carriers.

Invasive electrophysiology studies also observe an association between ISV and the firing mode of the locus coeruleus, the origin of the norepinephrine system. High tonic activity of the locus coeruleus is associated with reduced phasic firing in response to stimuli and an increase in RT distribution width, while moderate tonic activation is associated with a more marked phasic

response and reduced RT distribution width (Usher, Cohen, Servan-Schreiber, Rajkowski, & Aston-Jones, 1999).

As stated in previous chapters, increased ISV is also a hallmark of a number of neuropsychiatric conditions, several of which (ADHD, schizophrenia) are strongly associated with catecholaminergic abnormalities. Indeed Spencer *et al.* (2009) found that treatment using methylphenidate, a catecholamine agonist, normalised increased RT skewness in ADHD patients.

Computational

The computational literature strongly compliments the psychopharmacological findings described above. A number of influential models of catecholamine function find a relationship between catecholaminergic neuromodulation and ISV.

Li, Lindenberger, and Frensch (2000) modelled catecholamine activity as a gain parameter that amplified a system's sensitivity to sensory input, relative to background noise. Lower levels of 'catecholaminergic' gain were associated with increased ISV, while higher levels of gain were associated with reduced ISV.

Usher *et al.*'s (1999) model of locus coeruleus activity is in close agreement with invasive electrophysiological work on monkeys; showing increased ISV when tonic activity is high and reduced ISV with a more phasic mode of noradrenergic neuromodulation. This relationship appears to be mediated by changes in the electrotonic coupling of locus coeruleus neurons, with high coupling being related to high phasic activity and low ISV, and low coupling being associated with increased tonic activity and high ISV. Taken together with the invasive electrophysiology, these findings suggest a more nuanced relationship between catecholaminergic neuromodulation and ISV than a simple negative correlation between ISV and catecholamine levels.

These results are especially compelling when one considers that these models were not expressly devised to explain ISV, rather the changes in ISV resulting from changes to the catecholaminergic parameters of the model emerged naturally.

Theoretical Perspectives

The major theoretical stances as to the neural origins of ISV can be broadly divided into those citing neural noise as the cause of ISV, those that view ISV as representing an oscillatory signal, and those that view ISV as the consequence of attentional lapses.

The neural noise model

One possible conceptualisation of ISV is as representing neural noise. Hendrickson (1982), writing on the neural correlates of intelligence, first suggested that ISV could represent failures in neural transmission, leading to delays in the delivery of information and thus delays in the mental operations that underpin RTs. Jensen (1992) suggested that these transmission failures could be related to faulty myelination of axonal tracts, a theory which ties in well with literature suggesting a relationship between volumetric white matter and ISV (*e.g.* Waldhovd & Fjell, 2007).

More recent incarnations of this model have tended to relate ISV to neurocomputational deficits due to dysregulaton of the catecholaminergic system. Aston-Jones and Cohen (2005) posit a role for norepinephrine in tuning behavioural flexibility. With high tonic noradrenergic activation, an organism is sensitive to all environmental stimuli. This creates a noisy neural decision-making environment and leads to increased ISV and greater distractibility. Moderate tonic activity allows greater phasic tuning of attention and allows attention to be concentrated on particular stimuli, with low ISV at the cost of behavioural flexibility. Finally low tonic activity

leads to low responsiveness to all stimuli and is conducive to sleep. Thus ISV could represent an indicator of the noradrenergically mediated attentional state of an organism.

Neural noise can of course stem from a variety of sources in the brain. Synaptic failures, faulty myelination, and disordered catecholaminergic neuromodulation could all result in increased behavioural output. ISV may in fact be sensitive to neural noise, regardless of its specific physiological underpinnings, making measures of ISV more of a functionalist measure than one associated with a certain biological property.

The oscillatory model

An alternative class of models views ISV as having a temporal structure, rather than representing white noise (see Chapter 2). RTs, by this model, show oscillations at certain frequencies, which are indicative of underlying neural oscillations. Probably the earliest version of this model was Surwillo's (1975) EEG gating-signal hypothesis which suggested that ISV was moderated by two factors: the frequency of cortical oscillations and neural recovery periods. If a signal arrived at the unresponsive phase of a neuron's oscillatory cycle, it would have to wait until the neuron was responsive again before the impulse could proceed. Thus natural periodicities would occur in RT distributions.

These periodicities would be relatively high frequency, but other work has found oscillations in performance on a number of different scales, reflecting the fractal nature of the brain's oscillatory dynamics. Aue, Arruda, Kass, and Stanny (2009) found periodicities with wavelengths of 1.5 minutes in longer attention tasks. This low frequency spectral activity may be an important feature of RTs – van Ordern, Holden, and Turvey (2003) found the $1/f$ frequency distribution that is common to biological systems in RTs, and suggested that is indicative of self-organisational dynamics thought to underpin goal oriented behaviour. Interestingly, Gilden and

Hancock (2007) showed that participants high in ADHD symptoms demonstrate a deviation from the $1/f$ distribution noise found in those low in such symptoms.

One recent oscillatory model of ISV and sustained attention is the default mode interference hypothesis (Sonuga-Barke & Castellanos, 2007). This model cites the default network, a self-organising group of brain regions first described by Raichle *et al.* (2001), as the cause of periodic attentional failures. The default network is hypothesised to have two wings that exist in a state of dynamic tension. One wing directs attention externally to the outside world, while one promotes an introspective mode of attention. The two wings show substantial anticorrelation, with the network oscillating between a default mode and a task-positive mode, and this system has been hailed as a means of explaining directed attention without resort to an executive homunculus. Kelly, Uddin, Biswal, Castellanos, and Milham (2008) found that the strength of the anticorrelation was negatively related to ISV, suggesting that the dynamic tension of the network was important in regulating attention.

The oscillatory class of models are exciting because they appear to offer a glimpse of the elusive self-organisation of neural behaviour that is increasingly believed to underpin cognition. That variance in RTs appears to show the $1/f$ spectral properties common to a number of biological signals, including EEG, also supports the idea that ISV is a meaningful index of functioning. However it is important to note that RTs showing temporal structure is not necessarily indicative of oscillatory self-organisation. It is possible to view certain oscillatory neural signals as interfering with unrelated processes, rather than having an organisational role *per se*.

The lapse model

The final conceptualisation of ISV is that slow RTs represent attentional lapses. Unsworth Redick, Lakey, and Young (2009) found that measures of executive control were negatively correlated with RT parameters sensitive to the slowest RTs of a distribution, suggesting that failures in fronto-executive systems may lead to these lapses. The prefrontal cortices are of course already implicated in ISV by the literature mentioned above. Weissman, Roberts, Visscher, and Woldorff (2006) explored the neural bases of these lapses and found that reduced prefrontal and cingulate activity tended to precede lapses, while increased activity in regions thought to be part of the default network was seen during these lapses. Schmiedek , Oberauer, Wilhelm, Süβ, and Wittmann (2007) however examined the relationship between IQ and various RT distributional parameters, and found that lapses were not necessary to explain the relationship.

Lapse models have an undeniable intuitive appeal, but lack strong explanatory power on a neural level - it is not clear what the neural underpinnings of a lapse might be. A number of models that make reference to lapses do so within the framework of catecholaminergic neuromodulation or default network-based oscillatory behaviour leading one to question what exactly the concept of lapses are contributing to our understanding. Lapses may be at best the subjective correlates of neural delays that are more fundamentally explained by noise or oscillation-based dynamics.

Synthesising different conceptualisations of ISV

It is worth considering the implications of this last idea. These three classes of explanation are by no means mutually exclusive. For example, periodic fluctuations in levels of catecholaminergic neurotransmitter activity could be understood in terms of neural noise or oscillatory activity, and indeed an explanation synthesising both of these ideas may lead to a

fuller understanding of the phenomenon. However it is also possible that these models may represent the same phenomenon on different neural scales, in such a way that one explanation is clearly more fundamental. As mentioned, the increased responsiveness to irrelevant stimuli brought about by increased tonic levels of norepinephrine could be understood as a lapse of attention, but the phenomenon is clearly noradrenergic at its core.

Conclusion

To summarise, an increasing body of work has begun to elucidate both the possible neural mechanisms by which ISV may arise, and the biological substrates of individual differences in ISV. The empirical literature points mainly to fronto-executive, axonal, and catecholaminergic determinants of ISV. The findings have been interpreted in a number of ways but it may be helpful to loosely categorise them as neural noise, oscillatory, and lapse models. These models are largely compatible, and it may be that some fusion of these concepts is necessary to explain ISV.

Chapter Five

An Introduction to Electroencephalographic Data Analysis

Discovered by Richard Caton in 1875[2], and first applied to humans by Hans Berger in 1929, electroencephalography (EEG) is the oldest non-invasive functional neurophysiological measurement approach used in neuroscience. Despite its age, EEG has enjoyed largely continuous use since its discovery, and advances in computing and signal processing methods have lead to something of a renaissance in the approach in recent years. The empirical work described in this thesis makes extensive use of EEG and this chapter aims to provide a short primer in the technique. EEG methods, however, are highly varied, so this chapter will focus on the approaches used in later chapters. Specific description of the analysis techniques used in this thesis will follow in Chapters Seven and Nine, and some important aspects of these analyses will be evaluated in Chapter Ten. The present chapter is intended to give a more general overview of EEG analysis as a technique.

Recording Principles

EEG is the recording of the electrical activity of the brain using electrodes on the scalp. The neural signal is tiny compared to the electrical noise present in the recording environment, and thus common to all electrodes, so at least three scalp electrodes are needed to record EEG: a recording electrode (E_{Rec}), a reference electrode (E_{Ref}), and a ground electrode (E_{Gnd}). EEG is the signal given by: $(E_{Rec} - E_{Gnd}) - (E_{Ref} - E_{Gnd})$, which is amplified many thousands of times before being digitised and stored electronically. This process of differential amplification allows the signal of interest to be disentangled from environmental noise.

---

[2] It should be noted that Richard Caton recorded electrical activity from the cortices of animals, rather than the scalp, so this was thus strictly speaking electrocorticography rather than electroencephalography. Hans Berger was the first to use the technique non-invasively, and the first to apply it in humans.

Most modern research EEG uses a number of recording electrodes, referred to ground and reference electrodes, in order to capture activity from across the scalp, but the basic recording principles are the same. With a multi-channel recording there is the option of moving away from using a single electrode as a reference channel (called a common reference), to using the average of all electrodes (called a common average reference) or a system where electrodes are referenced to mathematical functions of groups of nearby electrodes (a class of approaches called Laplacian referencing). For a thorough treatment of referencing issues, the reader is directed to Nunez and Srinivasan (1981; 2006).

Biological Underpinnings

When EEG was discovered it was initially unclear exactly what aspect of neural activity it was sensitive to. Neurons are induced to fire by a depolarisation of a post-synaptic membrane, effected by the flow of ions through gated channels; and it is possible to measure the change in voltage using an extracellular microelectrode. This change in voltage can be modelled as a *dipole*, a paired current source and current sink. In EEG however, macroelectrodes are used at the scalp. Each EEG electrode measures the summed activity of up to a billion neurons; surely such a number of tiny voltage changes would average to zero? Fortunately, cortical neurons do not operate independently; they are embedded in densely interconnected columns. The manner of this interconnection leads neurons to fall into synchrony, firing together rhythmically (see Singer, 1993). Voltage changes in human tissue at the frequencies relevant to EEG (conventionally <100Hz) obey Ohm's law, and thus voltages from smaller circuit elements, such as individual neurons, can be linearly summed (Nunez & Srinivasan, 1981). This means that rather than modelling brain activity on the level of individual *microsources* generated by individual synaptic membranes, it is possible to summarise this low level activity using *mesosources*: aggregations of

a number of coherent sources that can be modelled as a single dipole. Moving up in scale, individual columns fall into synchrony with neighbouring columns until widespread areas of cortex are synchronous. It is this property of synchronisation that allows the recording of scalp EEG. In fact, approximately $6cm^2$, some sixty million neurons, of cortex must be synchronised in order to be measurable in raw EEG (Cooper, Winter, Crow, & Walter, 1965), although signal processing approaches, such as averaging, can reduce this. The generators of EEG are therefore, formally speaking, better characterised by a dipole layer than a point dipole. Once again though, it is possible to abstract dipole layers into *single equivalent dipoles*.

The amplitude of the EEG signal is thus a measure of short range coherence between groups of neurons. This is one reason why EEG is an index of membrane potentials rather than of action potentials – action potentials are simply too brief for large areas of cortex to produce them in synchrony. The longer membrane potentials allow enough margin of error for the EEG to identify synchrony. The same reason explains the $1/f$ power distribution observed in EEG data (Singer, 1993). It is easier for large areas of cortex to synchronise at low frequencies than it is at high frequencies, and so there tends to be an inverse relationship between frequency and power (Schaul, 1998).

Another reason such a large number of synchronous neurons are needed to generate measureable scalp EEG has to do with the properties of the skull. The skull acts as a spatial low pass filter, limiting the ability of EEG to resolve high spatial frequencies and register small areas of coherent cortex. It also acts as a temporal low pass filter, strongly attenuating the amplitude of the EEG signal – especially at high frequencies. A final factor affecting measurement of EEG at the scalp is the orientation of electrical fields. Areas of cortex on the crests of gyri, produce *radial dipoles* which are best oriented for their electrical fields to project through the scalp. Sources on gyral walls instead produce *tangential dipoles* which are relatively de-emphasised by

scalp EEG. In practice however, the area of cortical tissue required to produce measurable EEG is likely to minimise the importance of this concern.

Event-Related Potentials

The most common use of EEG in cognitive neuroscience is to compute event-related potentials (ERPs). This involves cutting an EEG recording, taken while a participant performs repeated trials of the same task, into segments that are time-locked to an observed event, such as the participant making a response or a stimulus appearing onscreen. These segments can then be averaged together time point by time point. The rationale is that ongoing EEG contains activity that is relevant to the task, as well as activity which is incidental. As EEG can be positive or negative, and is centred on zero, activity that is not consistently related to the task will average to zero, given sufficient trials. Activity which occurs consistently at the same time on every trial, however, will remain in the average ERP. This approach has proven highly successful and has identified a number of features of ERPs that appear to occur predictably under certain circumstances. These *ERP components* have been widely studied, and ever since the first ERP component, the contingent negative variation, was discovered by Grey Walter (Walter, Cooper, Aldridge, McCallum, & Winter, 1964)), their functional roles have been matters of wide debate in neuroscience. As Gaillard (1988) points out, these controversies are exacerbated by issues of component classification and consistency across paradigms.

Despite much to recommend them, ERPs have their drawbacks. ERPs are highly sensitive to time-locked and phase-locked EEG, generally referred to as *evoked* activity. If a component occurs in every trial, but with varying latency (*i.e.* shows poor time-locking to the event used for segmentation), the ERP will smear the activity out in time, reducing its amplitude (Spencer,

2005). Worse still, if a component is oscillatory but its phase does not line up perfectly across trials, the component will self-cancel in the average.

ERPs are also unusual in that the signal is averaged before summary statistics such as amplitudes and peak latencies are taken. This contrasts with measures such as RTs, where measurements are taken on a trial-by-trial basis and then averaged, an issue that will be taken up below.

Source Localisation Approaches

Another issue with standard ERP approaches is that the data remain on the electrode level. While EEG is recorded at the scalp, these measurements are, in effect, proxy measurements for the intra-cerebral sources of this activity: the dipole layers mentioned above. Some applications of EEG use the scalp data to quantitatively estimate the location and strength of the underlying dipoles, an approach which is called *source localisation*. Such an application has much intuitive appeal, the locus of the neural generators of a certain EEG component could be of theoretical importance. There are, however, sizeable technical issues to confront before such ambitions can be realised.

Imagine the opposite problem. A dipole of known location is active in the cortex and we want to predict its scalp projections. This situation is called the *forward problem*, and, using basic physical principles, it is possible to solve this and calculate the projections of this dipole to the scalp. The *inverse solution*, the localisation of a dipole given its scalp projections, is however mathematically intractable. Unless an infinite number of sites on the scalp are measured, a clearly impossible criterion, it is possible for an infinite number of dipole configurations to lead to the observed scalp projections (Plonsey, 1963).

This is less of an impediment than it might appear. Clearly some dipole configurations are implausible: those which involve dipoles inside ventricles, within the skull, or outside the head, for example, can be eliminated. A wide range of source localisation algorithms now exist (see Koles, 1998 for a review) and, using a variety of criteria, have shown promising levels of agreement with other neuroscientific methods, such as magnetic resonance imaging, which do not have equivalent localisation problems (see Linden, 2005 for a review). The intractability of the inverse solution, however, remains a fundamental problem for EEG source localisation.

Factor-Based Approaches

A way of moving beyond electrode-level analyses, without having to confront the issues that apply to source localisation, is the growing area of factor[3]-based approaches. EEG data can be viewed as a linear mixture of different neural sources, with each source measured in different proportions at different electrodes (Parra, Spence, Gerson, & Sajda, 2005). Standard EEG approaches analyse individual electrodes separately, relying on choosing the electrodes that are in the closest proximity to the source of interest to preferentially measure the source they are interested in. This approach is somewhat crude as the activity recorded at any single electrode still represents a mixture of different sources. Furthermore remaining at the sensor level means that any physical noise specific to a single electrode is also present. Factor-based approaches instead take advantage of the statistics of the observed data to create linearly weighted sums of the activity of all electrodes. If the correct weighting of electrodes can be identified then it is possible to measure a signal that is closer to the underlying sources of interest than any one

[3] Traditionally the term *component* is used to refer to constructs representing linearly weighted sums of observed variables, while *factor* is used to refer to constructs estimated from observed variables using a least-squares or equivalent approach. This thesis will however use the term *factor*, in order to avoid confusion with ERP components.

electrode. Unlike source localisation however, these methods do not attempt to solve the inverse problem, and so make far fewer neurophysical assumptions.

Consider two adjacent electrodes E_1 and E_2. Both measure activity from a single EEG source S as well as electrode-specific noise sources N_1 and N_2. So $E_1 = S + N_1$ and $E_2 = S + N_2$. As Ohm's law applies in neural tissue, we can linearly sum E_1 and E_2, to give factor $F = (E_1 = S + N_1) + (E_1 = S + N_1)$, or stated differently $F = 2S + N_1 + N_2$). Assuming Gaussian noise, this results in an improvement in signal-to-noise ratio compared to analysing each electrode separately. One problem with this is that giving electrodes equal weighting will often not be the optimal approach. If E_1 were directly over S, while E_2 were further away, giving E_1 a greater weighting than E_2 may yield a better estimate of S. But how can one tell whether E_1 or E_2 better measures S? And even if we know this, how can one derive specific weights for E_1 and E_2? Will a 60:40 or a 90:10 ratio better represent S? These issues become all the more pressing when one considers that EEG montages generally contain far more than two electrodes. With electrodes covering the whole scalp, giving each electrode equal weighting for a source that appears to be maximal at the vertex makes little sense, as EEG sources behave like dipoles, some electrodes record the current source while others measure the current sink, and summing these electrodes using equal weightings could amount to modelling F as $F = S - S + N_1 + N_2$, namely measuring only, admittedly attenuated, noise[4].

Fortunately, a number of statistical approaches exist to identify an optimal set of electrode weightings; with each approach optimising weightings using a different set of criteria (see Parra, Spence, Gerson, & Sajda, 2005). One possibility is the use of regression to identify the weighting

---

[4] An example whereby only a single source exists, and its projections to the recording electrodes entirely cancel may seem somewhat fanciful, but were it possible to record with infinite electrodes, granting perfect spherical coverage of the brain, an unweighted sum of these electrodes would cancel due to current conservation.

of electrodes that optimally predicts another variable, either continuous, such as RT, or categorical, such as whether data come from one trial type or another. Such an approach is comparable to the difference waves computed in some traditional ERP applications, but with the aforementioned advantages for signal-to-noise associated with data aggregation.

A second possibility is the use of principal components analysis (PCA; see Dien, 2009) or independent components analysis (ICA; see Makeig, Bell, Jung, & Sejnowski, 1996) to identify weightings using the inter-electrode covariance matrix, or factor correlation matrix respectively. PCA is a multivariate statistical approach which reduced a number of variables into a smaller set of axes, called *principal components*, which summarise the variance. These principal components are sequentially subtracted from the data until a predetermined number of factors have been extracted, with any residual variance being discarded. PCA is closely related to the similar technique of factor analysis; the major difference being that while factor analysis only works on shared variance, in order to estimate a latent variable, PCA works on all variance, directly calculating a linear combination of observed variables. While the relative advantages and disadvantages of factor analysis and PCA are beyond the scope of the present work, it is worth noting that almost all variance present in EEG data is shared, and so both approaches will generally arrive at similar solutions. PCA is primarily chosen above factor analysis for reasons of computational efficiency – a key concern when working with the enormous datasets common in EEG research.

The Infomax rotation algorithm (Bell & Sejnowski, 1995), is the most popular for ICA in EEG analysis, and has proven particularly effective at identifying sets of weightings that correspond to functional networks (Dien, Khoe, & Mangun, 2007). Infomax rotation works by creating factors which load on each electrode and iteratively adjusting these weightings in order to minimise both the factor correlation coefficients and the Gaussianicity of each factor. The logic of the latter criterion is that, according to the central limit theorem, genuinely separate

factors will be less Gaussian than combinations of factors. These two criteria work to slowly separate the factors until they show maximum statistical independence.

Single-Trial Approaches

Pham, Möcks, Köhler, and Gasser (1987) formalise the *standard model* assumed by average ERPs, $x_j(t) = s(t) + e_j(t)$, where $x_j(t)$ is the EEG recorded at time t for the jth trial of an ERP task, $s(t)$ is the EEG activity reliably evoked by the stimulus, and $e_j(t)$ is the ongoing EEG activity which is not related to the task. Single trials have poor signal-to-noise ratio (*s*:*e*), preventing reliable estimates of $s(t)$. However, as this model assumes $e_j(t)$ to be Gaussian, taking the average of all trials will attenuate $e_j(t)$ while preserving $s(t)$, thus improving the signal-to-noise ratio.

While this model may approximate early sensory components, there is evidence that later, more cognitive components, such as the P3b, show substantial variability in latency (Kutas, McCarthy, & Donchin, 1977). Pham *et al.* thus modify this model by adding the parameter τ_j which represents trial-to-trial changes in the latency of $s(t)$, (latency jitter) to give the model $x_j(t) = s(t + \tau_j) + e_j(t)$. According to this second model, the cross-trial average of $x_j(t)$ will *not* give an unbiased estimate of $s(t)$. Instead such latency jitter leads to a smearing of the ERP peak in the time domain, with a corresponding fall in peak amplitude (see Spencer, 2005). Such distortions can make averaged ERPs misleading estimates of the amplitude, latency, and morphology of an ERP component.

An alternative to averaging ERPs is to calculate parameters from the EEG of single trials, thus allowing trial-by trial estimates of EEG activity, and preventing distortions that can arise from averaging. Without averaging, however, the signal-to-noise ratio remains low, and it can be difficult to differentiate peaks from the ongoing EEG.

Fortunately, alternatives to averaging exist for improving signal-to-noise ratio. As stated above, factor-based approaches improve signal-to-noise ratio by integrating information across electrodes, rather than across trials. Another option is to identify the frequency band in which the EEG phenomenon of interest is supported, and isolating it. This can be accomplished using time-frequency transformations, or by employing band-pass filtering to attenuate power in other frequencies.

Once the signal has been *denoised* it is necessary to identify component amplitudes and latencies in single trials. There are four main approaches to do this: visual inspection, peak picking, template matching, and maximum-likelihood estimation. Visual inspection is as it sounds: each trial is inspected by an experienced observer and latency and amplitude are estimated in single trials. While a trained eye is not to be underestimated, this approach clearly lacks transparent criteria, is time consuming, and may introduce unintended biases into the data. Peak picking identifies the timepoint with the highest or lowest amplitude within a certain latency range. This approach relies upon stringent low-pass filtering (~3.4Hz) in order to avoid picking spurious peaks (Smulders, Kenemans, & Kok, 1994). Template matching involves passing a template, generally a sinusoidal half-wave or the averaged ERP, across each trial and identifying the point of maximum cross-correlation or covariance. The most well known application of this technique is the Woody filter (Woody, 1967), an iterative approach which starts with the average ERP as a template, before using the new average, derived from the picked timepoints, and repeating the process. There is evidence that using cross-covariance gives better results than cross-correlation (Smulders, Kenemans, & Kok, 1994), presumably as this allows amplitude information to be used. Finally, Pham, Möcks, Köhler, and Gasser (1987) describe a method for identifying peaks using maximum likelihood estimation. The model $x_j(t) = s(t + \tau_j) + e_j(t)$ is Fourier transformed and τ_j can be estimated using iterative Fisher scoring. The advantage of this

approach is that rather than attempting to identify the latency of s in spite of $e_j(t)$, maximum likelihood estimation is able to take advantage of the noise to help estimate s.

Jáskowski and Verleger (2000) tested Pham *et al*'s method against template matching and peak picking, and found some advantage for maximum likelihood estimation where levels of jitter were low, but that all approaches performed equally with levels of jitter that would be realistic in a P3b task. Smulders *et al.* (1994) compared the reliability of peak picking and template-matching, and found an advantage for peak picking with low-pass filtering. Both Jáskowski and Verleger (2000) and Smulders *et al.* (1994) found that these methods were all highly contingent on the signal-to-noise ratio, so it is worth reiterating that successful single-trial analysis is at least as reliant upon employing an effective denoising technique as it is on the choice of peak identification approach.

Conclusions

To summarise, while EEG is principally similar to the 19th century science practised by Catton and Berger, huge changes in computer technology, biophysics, and signal processing have led to continued innovation in the field. While the ERP approaches pioneered by researchers such as Grey Walter continue to serve neuroscientists well, technical advances and the shortcomings of ERPs under certain conditions have driven many neuroscientists to experiment with new approaches to EEG data. These approaches include attempts to identify and estimate the dipolar sources that underlie EEG, using source localisation algorithms, as well as approaches for finding linear combinations of electrodes that better characterise the underlying sources, but stop short of attempting to locate them in the brain. Future analytic developments will hopefully allow EEG's recent renaissance to continue, including the growing area of single-trial analysis, where alternative processing approaches to the averaging used by ERPs are needed.

Research programme

The preceding five chapters described the current literature on ISV. Controversies exist on every level of the field, from measurement issues to putative aetiology, but some consistent themes exist. ISV appears to be a common correlate of several psychiatric and neurological conditions, particularly those which putatively affect catecholaminergic neuromodulation, cerebral white matter, or the frontal cortices. This has led to neural noise and 'lapse' models of ISV, while the conceptualisation of ISV as representing low frequency neural oscillations stems from time series approaches to RTs.

The position of this thesis is that ISV represents a general and global construct in two ways. Firstly, ISV is thought to represent a global construct, which affects performance on a supra-task level. This makes it an ideal candidate for a latent variable approach, allowing task-specific variance (which may largely represent measurement error) to be removed from estimates of true scores. This should lead to more accurate and reliable estimations of 'trait ISV'. Secondly, as ISV has been empirically linked to a number of aetiological factors, it may be that ISV represents a functional neurocomputational property of the brain, rather than any aetiological factor in particular. This second assumption is not explicitly tested by this thesis, but is discussed in greater detail during Chapter 10.

The following four chapters describe the empirical work conducted as part of this thesis. These studies use a variety of methodologies, including psychometrics, electroencephalography, and latent variable modelling. In doing so, they show the development of a distinct approach to the study of ISV, starting with measurement issues and the measurement of ISV on the observed variable level, and ending by using a latent variable approach to explore true score correlations between ISV and electroencephalographic variables.

Study One focuses on the measurement of ISV. A number of metrics are used to quantify ISV and it is not clear to what extent they provide different information from one another. It is also not clear how these measures compare in terms of their reliability, an important issue for measurement. This study assesses the test-retest and odd-even reliability, and redundancy, of a number of RT parameters. The differential impact of trial number on measures of ISV and measures of central tendency is also explored.

In Study Two high and low ISV participants, recruited from Study One, carry out a working memory oddball task while having an EEG recording. A principal components analysis based approach is used to compose the data, allowing P3b latency and amplitude to be identified in single trials, and latency jitter-free parameters to be computed for both groups. This allows the two groups to be compared on measures of latency jitter and on jitter-free measurements of P3b amplitude, unlike previous ERP work in the area.

Study Three examines the extent to which individual differences co-vary across tasks and modalities, in order to test the assumption, implicit in the literature, that ISV represents a unitary and global trait. This is accomplished using latent variable modelling to identify the psychometric dimensionality of ISV across two tasks, each performed in two modalities.

The final empirical chapter, Study Four combines approaches from the four preceding chapters. By employing a latent variable approach, supra-task parameters are derived for a number of RT and P3b parameters. and latent variable correlations between these parameters are estimated.

Lastly, Chapter Ten discusses the methodological and theoretical implications of this thesis, drawing on the empirical data and the reviewed literature to suggest future directions for the field.

Chapter Six

On the Stability of Instability:

Optimising the Reliability of Intra-Subject Variability of Reaction Times

Abstract

While reaction times have been traditionally aggregated using measures of central tendency, interest in higher moments of the reaction time distribution, particularly intra-subject variability (ISV) has grown in recent years. However it is unclear to what extent individual differences in these higher moments are stable across time, reflecting trait-like features. The present study compares the reliability of a number of metrics for higher moments of the reaction time distribution on a battery of speeded tasks. The reliability of ISV is shown to be dependent on both the metric used and the number of trials used to calculate them. However, when using sufficient trials and appropriate metrics, ISV shows good test-retest reliability. This study has important practical implications for the design and analysis of studies into ISV, as well as theoretical importance for the trait concept of ISV.

Introduction

Experimental psychology has a long tradition of using speed of processing, measured using reaction times (RTs), in order to identify the mechanisms that underpin mental operations. In order to obtain reliable measurements, most cognitive experiments involve many trials so require summary statistics in order to make sense of the data. Generally a measure of central tendency, such as the mean or median, is used to characterise the RT distribution. This provides an easily interpretable, and generally robust, measure of a participant's RT. However, RTs come from multi-parameter distributions and the use of means and medians to the exclusion of measures of intra-subject variability distract researchers from the observation that variation around the mean may be more than mere noise, and instead may be another facet of the signal.

While the study of higher moments of the RT distribution has yet to find widespread favour in mainstream experimental psychology, the topic is rapidly gaining importance within differential psychology and neuroscience. Intra-subject variability (ISV) of reaction times, reflecting the intra-individual variation of response times around the mean and thus the second moment of the RT distribution, has been shown to exhibit strong relationships with a number of neurological and neuropsychiatric conditions, including attention deficit hyperactivity disorder (ADHD, Klein, Wendling, Huettner, Ruder, & Peper, 2006), schizophrenia (Birkett *et al.,* 2007), Alzheimer's (Burton, Strauss, Hultsch, Moll, & Hunter, 2006) as well as old age (Hultsch, MacDonald, & Dixon, 2002) and terminal decline (MacDonald, Hultsch, & Dixon, 2008). It is also related to psychometric intelligence, with correlations between ISV and intelligence surpassing those between mean RT (MRT) and intelligence (Jensen, 1992). Investigating the higher moments of RT distributions is, however, not without complications. A great many measures of ISV exist, and there are few studies

comparing them. There is also the issue of reliability, which is of cardinal importance in differential psychology

Choice of parameter

Firstly there is the choice of parameter. The majority of ISV research focuses on measures of variance, such as the standard deviation (SD) or the mean absolute deviation (MAD). These parameters measure the spread of observations around the mean, irrespective of the direction of the deviation. RT distributions, however, are not Gaussian and show significant rightward skew. It is thus questionable whether such measures of ISV are appropriate, or whether they are making an implicit but erroneous assumption of Gaussianicity.

Indeed there is evidence to suggest that this asymmetry, characterised by an extended tail at the slow end of the RT distribution, may be important. Larson and Alderton (1990) coined the *worst performance rule* by showing that RTs in the slow tail of the RT distribution are more highly correlated with IQ than those in the fast tail. These right-tail RTs also seem to discriminate between children with and without ADHD (Leth-Steensen, King Elbaz, & Douglas, 2000), and thus may be of special interest to differential psychologists.

In order to address this problem, a number of parameters have been suggested to quantify the right tail of the distribution. Skewness is the third moment of a distribution, and captures the distribution's asymmetry, with positive values indicating a longer or denser tail at the high, right-handed, end of the distribution and negative values indicating that the left-handed end is more densely populated. As a higher moment however, it may require a large number of observations to be reliably estimated, and so most researchers use other approaches. The *tau* parameter of the ex-Gaussian distribution is another important measure of skewness. The ex-Gaussian is a convolution of a Gaussian and an exponential distribution, with the parameters *mu* and *sigma* respectively representing the mean and standard deviation of the

Gaussian component, and tau representing the mean and standard deviation of the exponential component, which gives the ex-Gaussian its positive skew. For these parameters to be derived from the RT data, a model must be fitted to the data that is reliant on not only having enough observations (at least 40; Heathcote, Brown, & Cousineau, 2004), but also on the data approximating an ex-Gaussian distribution.

While ex-Gaussian parameters represent one conceptual framework for characterising RT distributions, namely by treating the skewness as a meaningful parameter to be measured in its own right, other approaches exist. Carpenter (1981) makes an interesting case for carrying out analyses on the reciprocals of RTs, rather than on raw values. Standard RT distributions, Carpenter argues, are non-Gaussian not because the underlying process is non-Gaussian, but because we misunderstand what an RT represents. By taking the reciprocal of an RT, we no longer treat it as a measurement of latency, but of a rate. When plotted, reciprocal RTs appear Gaussian, and do not show the skewness that characterises raw RT plots (Carpenter, 1981). The first two moments of this distribution can be measured by taking the mean and standard deviations of the reciprocals of the RTs, or by more explicitly fitting Carpenter's Linear Approach to Threshold with Ergodic Rate (LATER) model, which gives the mean and standard deviation of the best-fitting Gaussian distribution of the reciprocal RTs. This model, however, has as yet largely been used in saccadic RT research only.

The question of whether these different parameters are measuring separate aspects of ISV, or whether they render one another largely redundant has not been widely addressed in the literature (but see Schmiedek, Oberauer, Wilhelm, Süβ, & Wittmann, 2007). Our lack of knowledge in how interchangeable these metrics are raises problems when making comparisons across studies using different measures of ISV. On one hand, we may be glossing over genuine differences in what various parameters measure; on the other, we may be overemphasising the differences between metrics that are essentially measuring the same thing.

Reliability of ISV

A second key issue in the measurement of ISV for research into individual differences is that of reliability. All measurements, according to test theory (Novick, 1966) represent a true score, plus additional error variance. The reliability of a measurement is the extent to which it measures the true score, rather than error. ISV was long thought to represent error variance, rather than a trait construct, and it is in part the apparent reliability of ISV (Flehmig, Steinborn, Langner, Scholz, & Westoff, 2007; Kuntsi , Stevenson, Oosterlaan, & Sonuga-Barke, 2001) which led to its recognition as a variable of interest. ISV does, however, appear to have poorer reliability than measures of central tendency (Jensen, 1992) raising the concern that reliability may be something of an issue for ISV. This is of particular concern as there is hope that ISV may represent a marker for ADHD, schizophrenia, cognitive aging, intelligence, and terminal decline. ISV can clearly only be of use as a marker if it exhibits adequate reliability.

Although there is work suggesting that ISV may be reliable, albeit less so than measures of central tendency, this work is restricted to a small number of metrics. It remains unknown how other measures, such as ex-Gaussian and LATER parameters, compare to more common measures like standard deviations and coefficients of variation.

One key factor that determines the reliability of ISV is the number of trials it is based upon. According to the Spearman-Brown formula (Spearman, 1910; Brown, 1910), the Gaussian distribution of error variance leads to this error cancelling itself out with increasing aggregation of data, leaving the true score intact.

Such a conceptualisation of data aggregation is however strongly tied to the idea of a true score being analogous to a measure of central tendency of the distribution. Estimating the variance of a distribution may show a less transparent relationship between data aggregation and reliability. Measures of ISV could need greater numbers of RTs to 'settle' than first

moments such as means and medians, and be more vulnerable to outlier observations. Furthermore, ISV, by definition, must be estimated from a number of RTs, whereas average RT could, assuming perfect reliability, be measured from a single RT. This implies that the relationship between number of trials and ISV may go beyond simple data aggregation. Despite this, at present many researchers use similar numbers of trials for estimating ISV as they would for estimating measures of central tendency, a practice which may yield unreliable measures of ISV.

The aims of the present study are therefore threefold. Firstly, we set out to derive a large set of different ISV metrics from standard neuropsychological tasks that are popular in clinical research and explore the relationships between these measures. Do different measures of ISV seem redundant with regard to each other, or do they offer different information? Secondly, we determined the instrumental (odd-even) and test-retest reliabilities of these measures to assess their appropriateness as trait correlates in differential research. Which measures of ISV are most reliable? Finally, we conducted a Monte Carlo analysis of the test-retest reliability of the mean and standard deviation of RTs with different numbers of trials. Is the increase in reliability that the Spearman-Brown formula predicts with increasing trials slower for ISV than mean RT?

Methodology

Ethical approval was obtained from the School of Psychology ethics and research governance committee at Bangor University.

Participants

Ethical approval was obtained Eighty-seven student volunteers from Bangor University participated in exchange for printer credits and course credit. Twelve participants' data were excluded due to non-completion of the second testing session and a further ten participants'

data were excluded due to responding correctly on fewer than 30% of trials on any condition on any of the six tasks (list-wise exclusion). This left N=65 participants which all reported analyses are based upon.

Apparatus

Participants were tested on Pentium IV personal computers running Windows. Stimuli were presented on liquid crystal display monitors measuring 382x300 mm and refreshing at 60 Hz, and participants responded using Qwerty keyboards.

Design

The study used a within-subject test-retest correlational design.

Procedure

Participants were tested on two occasions exactly one week apart. On both occasions they carried out the six tasks, described below, used by Klein *et al.* (2006). Each session was held in the same location, at the same time of day, and was overseen by the same experimenter in order to maximise consistency between testing sessions.

In the Zero-Back Task (0BT) participants watched a series of letters (100 trials) appear on the screen. They were instructed to press the F key every time a letter appeared, unless it was the letter E (occurring in 20% of trials), in which case they were to press the J key. The One-Back Task (1BT) was identical to the 0BT, except that instead of pressing J when an E appeared, they pressed J whenever a letter matched the preceding letter. The Two-Back Task (2BT) was identical to the above tasks, but the J key was to be pressed when a letter matched the last but one letter. Henceforth, trials requiring an F response will be termed *non-events*

and those requiring a J response will be termed *events*. These tasks will also be referred to collectively as *n-back tasks* (NBTs)

In the Continuous Performance Task (CPT), participants watched a series of letters (300 trials) appear on the screen and pressed the space key whenever an X appeared (46 trials). To all other letters they were instructed to withhold any response. The Go-No Go Task (GNG) was the opposite of the CPT; participants were asked to respond to every letter, only withholding responses to the letter X (thus there were 254 trials requiring a response).

In the Stop Signal Task (SST), participants watched a series (300 trials) of Ls and Rs appear on the screen, and were asked to press the F key in response to Ls and the J key in response to Rs. In 75 trials, however, a red stop sign appeared. Participants were instructed to withhold responses on trials when this happened. The stop sign was presented 150ms post-stimulus on the first stop trial, and this latency was increased by 25ms every time the participant successfully withheld their response and was decreased by 25ms every time they responded.

In all tasks the participants were asked to respond as quickly and accurately as possible to every letter.

Data analysis

In order to assess the reliability of ISV, a number of parameters were computed. Unless stated, all parameters were computed separately for each participant on each task, session, and, where appropriate, condition.

Firstly, an accuracy measure (correct responses/total number of trials) was computed. After this, trials where responses were either incorrect or were made within 120ms (and thus were presumably pre-emptive) were removed from the data before RT parameters (see below) were computed.

The parameters fell into four main groups, corresponding to the first four moments of a statistical distribution. The first group consisted of measures of central tendency: *Mean RT* (MRT), *median RT* (MnRT), the *mu* parameter of the ex-Gaussian distribution, the mean reciprocal RT (RecipMRT), and the *mu* parameter of the LATER model.

The second group was made up of measures of variance: *Standard deviation* (SDRT) is probably the most familiar measure of variability. Due to its squaring of deviations in order to obtain uniformly positive values, it gives a somewhat disproportionately high weighting to extreme scores. *Coefficient of variation* (CVRT), defined as SDRT/MRT, is a measure of how much variability exists per unit of MRT. *Mean Absolute Deviation* (MAD) is the mean deviation from the MRT, rectified so that opposite signed deviations do not cancel. As MAD uses absolute rather than squared values it should be less influenced by extreme values than SDRT. *Range* is simply the difference between the fastest and slowest RT. *Interquartile range* (IQR) is the difference between the first and third quartile of the RT distribution. *Sigma* is the estimated standard deviation of the Gaussian component of the ex-Gaussian distribution. Finally the *standard deviation of reciprocal RTs* (RecipSDRT) and *Sigma* parameter of the LATER model were used.

The third group consisted of measures of asymmetry and of the size of the right-handed tail of the distribution: *Skewness* is a measure of a distribution's asymmetry. RT distributions are generally positively skewed, possibly because RTs are effectively capped at the fast end, but largely uncapped in terms of how slow a response can be. Thus a high positive skew is indicative of these very slow responses. *Tau* is the estimated mean and standard deviation of the exponential component of the ex-Gaussian distribution.

The final measure was *Kurtosis,* a measure of peakedness of a distribution, which indexes the extent to which the variance results from infrequent outliers, as opposed to more frequent smaller deviations from the average.

The three ex-Gaussian parameters were computed using Heathcote, Brown, and Cousineau's (2004) quantile maximum probability estimation (QMPE) software, available freely from the University of Newcastle Software Repository (http://www.newcl.org/). The authors suggest that meaningful parameters can be calculated from as few as 40 observations. The NBTs only had 20 event trials however, and so ex-Gaussian parameters are not calculated for these trials.

The LATER model was fitted using SPIC, a freely available software package (Carpenter, 1994; http://www.cudos.ac.uk/spic.htm). As it is inadvisable to fit the LATER model using fewer than 100 trials (Carpenter, personal communication), the two LATER parameters were only derived for the GNG.

Pearson's *r* coefficients were computed between session one and two for each parameter in order to estimate test-retest reliability. Pearson's *r* coefficients were also derived for odd and even trials, for session one only, in order to measure instrumental reliability. Due to insufficient trials, odd-even reliability coefficients were not computed for NBT event trials on any parameter, or for the NBTs or CPT on the ex-Gaussian and LATER parameters.

In order to assess the extent to which different ISV parameters provided different information, correlation matrices were computed for the parameters from the GNG (the task with the most trials) RTs across each occasion. The correlation coefficients were then averaged across sessions to obtain a single matrix.

While comparing the test-retest correlations for tasks with different numbers of trials will give some indication of the relationship between reliability and the number of trials used to compute a metric, this is a rather indirect measurement of the relationship between number of trials and reliability. Monte Carlo approaches represent a far more rigorous method to explore this question. We sampled random (with replacement) subsets of trials from each participant's performance on the GNG, separately for sessions one and two, derived ISV parameters from these trials and computed test-retest correlation coefficients. This was

carried out 100 times for each number of trials, which ranged from 10 to 300 in 10 trial increments.

Results

Test-retest reliability coefficients for the various tasks and parameters can be seen in Table 1. The data show that the event trials of the NBTs exhibited the lowest levels of reliability, particularly for the 0BT. This is likely due to the relatively small number of these trials (20). In contrast the GNG, which required 254 responses, shows good reliability across most measures of ISV.

Odd-even reliability coefficients are presented in Table 2. Most parameters show very good reliability, again especially on the GNG – the task with the most trials – where reliability coefficients are mainly at .90 or above.

Table 3 shows the correlation matrix of the thirteen parameters for the GNG task, averaged across sessions. There are substantial correlations between most of the parameters, even between parameters that ostensibly measure different moments of the distribution. For example, ex-Gaussian tau, generally thought of as a measure of skewness, is extremely highly correlated (r=.96) with SDRT, a measure of variance. It appears that some of the most reliable ISV parameters (SDRT, MAD, range) are so highly correlated that they are essentially redundant with regard to one another.

Figure 1 shows reliability on the Y-axis graphed against number of trials on the X-axis for MRT and SDRT. While MRT begins to reach a plateau after as few as 50 trials, SDRT does not reach a comparable plateau until as many as 200 trials, showing that SDRT needs a much greater level of data aggregation to achieve adequate reliability than MRT.

	MRT	SDRT	CVRT	MAD	Skew	Kurt	MnRT	Range	IQR	RecipMRT	RecipSDRT	ex-G Mu	ex-G Sigma	ex-G Tau	Later Mu	Later Sigma
0BTE	0.614	0.213	0.125	0.241	0.140	0.075	0.592	0.197	0.134	0.704	0.374	-	-	-	-	-
0BTNE	0.645	0.565	0.391	0.681	0.034	-0.075	0.641	0.289	0.741	0.750	0.497	0.631	0.608	0.486	-	-
1BTE	0.796	0.567	0.543	0.495	0.390	0.273	0.794	0.604	0.377	0.830	0.594	-	-	-	-	-
1BTNE	0.811	0.603	0.471	0.639	0.162	0.233	0.836	0.482	0.614	0.849	0.511	0.849	0.518	0.551	-	-
2BTE	0.851	0.604	0.269	0.620	-0.205	-0.198	0.798	0.518	0.527	0.743	0.076	-	-	-	-	-
2BTNE	0.832	0.812	0.506	0.821	0.487	0.360	0.800	0.637	0.766	0.835	0.544	0.681	0.521	0.118	-	-
CPT	0.665	0.609	0.431	0.645	0.268	0.178	0.653	0.480	0.519	0.783	0.389	0.569	0.310	0.384	-	-
GNG	0.836	0.842	0.862	0.804	0.250	0.120	0.837	0.727	0.663	0.808	0.803	0.912	0.771	0.753	0.630	0.920
SST	0.831	0.499	0.300	0.454	0.464	0.305	0.802	0.477	0.370	0.868	0.499	0.773	0.348	0.404	-	-

Table 1. Test retest correlation coefficients for tasks and RT parameters

	MRT	SDRT	CVRT	MAD	Skew	Kurt	MnRT	Range	IQR	RecipMRT	RecipSDRT	Ex-G Mu	Ex-G Sigma	Ex-G Tau	LATER Mu	LATER Sigma
0BT	0.986	0.805	0.698	0.872	0.123	-0.057	0.980	0.534	0.816	0.977	0.783					
1BT	0.982	0.503	0.344	0.672	0.151	-0.028	0.983	0.308	0.746	0.982	0.723					
2BT	0.977	0.840	0.696	0.866	0.506	0.264	0.963	0.645	0.810	0.975	0.741					
CPT	0.965	0.724	0.569	0.711	0.244	0.507	0.964	0.692	0.391	0.952	0.411					
GNG	0.995	0.911	0.866	0.950	0.340	0.239	0.995	0.735	0.902	0.994	0.907	0.987	0.857	0.903	0.986	0.897
SST	0.996	0.946	0.833	0.965	0.599	0.270	0.995	0.781	0.948	0.992	0.919	0.980	0.936	0.747		

Table 2. Odd-Even reliabilities for tasks and RT parameters.

	MRT	SDRT	CVRT	MAD	Skew	Kurt	MnRT	Range	IQR	Mu	Sigma	Raw Tau	RecipMRT	RecipSDRT	LATER Mu
SDRT	.713														
CVRT	.383	.915													
MAD	.779	.985	.852												
Skew	-.057	.301	.469	.166											
Kurt	-.086	.149	.288	.025	.934										
MnRT	.989	.612	.259	.689	-.122	-.116									
Range	.607	.917	.885	.865	.572	.472	.515								
IQR	.835	.908	.714	.961	.022	-.075	.771	.775							
Mu	.881	.319	-.059	.405	-.236	-.163	.933	.2485	.507						
Sigma	.792	.542	.290	.603	-.213	-.147	.810	.4735	.656	.772					
Tau	.643	.962	.892	.960	.270	.086	.537	.858	.901	.218	.386				
RecipMRT	-.931	-.659	-.359	-.720	.008	.061	-.914	-.579	-.775	-.799	-.628	-.637			
RecipSDRT	-.001	.639	.857	.573	.315	.176	-.110	.619	.434	-.382	.172	.615	.070		
LATER Mu	-.836	-.706	-.472	-.744	.002	.065	-.799	-.587	-.752	-.656	-.605	-.655	.872	-.118	
LATER Sigma	.003	.586	.778	.518	.236	.100	-.104	.500	.366	-.339	.110	.550	.088	.881	-.192

Table 3. Intercorrelations for RT parameters on the GNG task.

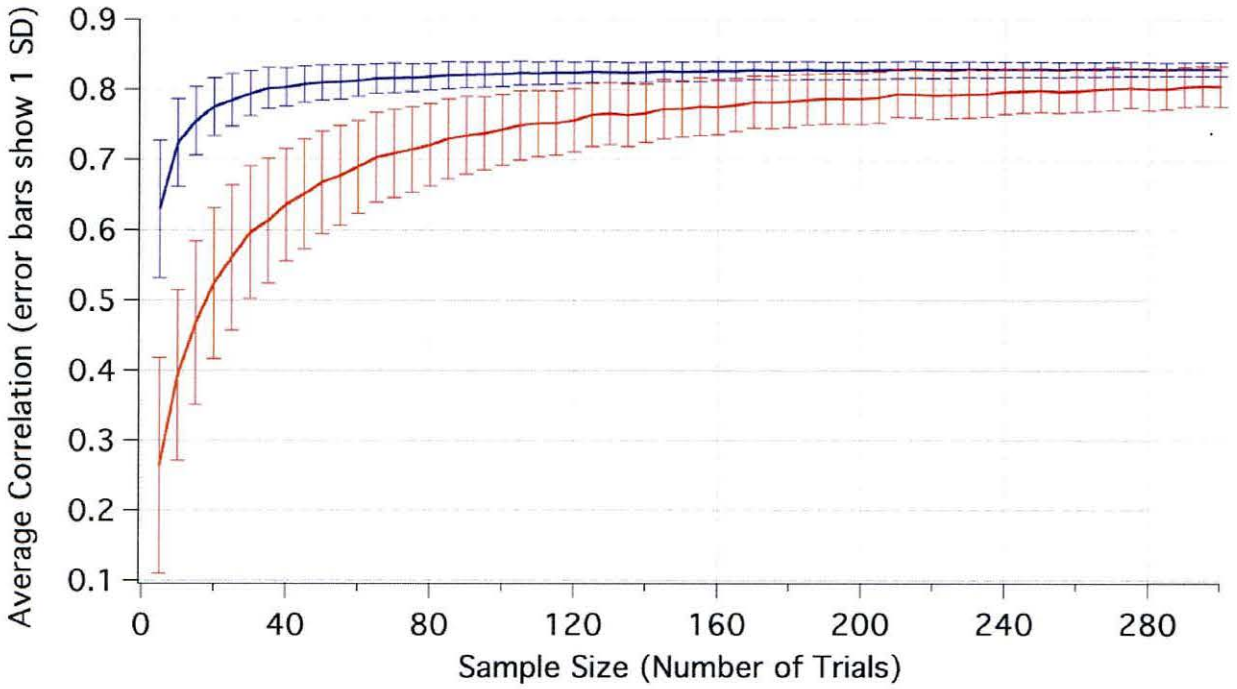

Fig 1. Summary of results for Monte Carlo analysis of test-retest reliability. Blue bar represents MRT and red bar represents SDRT.

Discussion

The present psychometric study on different metrics of ISV, assessed for six established neuropsychological tests with a sample of 65 undergraduate students, yielded the following main results. First, given sufficient levels of data aggregation (*i.e.* based on the GNG task), most measures of variability exhibited good to excellent instrumental reliabilities, with odd-even correlations reaching as high .95 for MAD, and around the .90 mark for many other parameters. Measures of the higher moments of the distribution, namely skewness and kurtosis, however exhibited poor levels of reliability. Second, these metrics also showed good test-retest reliabilities, reaching as high as .92 for LATER sigma, with most measures surpassing the .80 mark. Thirdly, most of these metrics showed substantial inter-correlations, reaching the point where a number of metrics could be argued to be redundant with one another. This was true for metrics that might be expected to be largely statistically interdependent (*e.g.* SDRT and MAD), but, more surprisingly, also true for SDRT and tau, metrics which theoretically measure different moments of the RT distribution.

The key finding of the present study is that metrics of ISV show good test-retest and odd-even reliability, and are thus suitable for individual differences research. These metrics of ISV did vary somewhat in reliability however, and our results suggest that SDRT, CVRT, MAD, the ex-Gaussian parameters, and the reciprocal and LATER parameters were the best choices in terms of reliability. These data do however draw attention to the importance of the number of trials used to compute such metrics. Our Monte Carlo simulation shows a striking difference between MRT and SDRT in the number of trials needed to achieve equivalent reliability. We thus encourage researchers utilising measures of ISV to consider using a greater number of trials than they would normally use in an RT experiment, as these parameters require a greater number of trials than the measures of central tendency which psychologists are used to.

Another important upshot of the high test-retest reliability of ISV metrics observed in the present study is that it provides further evidence that ISV shows trait-like characteristics. This is important for the idea of ISV as a suitable topic for differential psychology, but also as a possible endophenotype for psychiatric disorders – a key topic in ISV research. These data also suggest that ISV is a relatively stable trait in healthy participants, rather than only being a feature of pathological states such as psychosis or terminal decline.

Our data also suggest that the majority of ISV metrics are largely redundant with regard to one another. Indeed, even measures that theoretically measure different moments of the RT distribution, *e.g.* SDRT and tau, are highly intercorrelated. This has two important implications. Firstly, it may support the idea that the right tail of the distribution is of great theoretical importance. Tail RTs may be the source of much of the variance usually attributed to the second moment of a distribution, a notion supported by the stronger correlation between SDRT and tau than SDRT and sigma. Secondly, however, a pragmatic case may be made that if the previous point is correct, the SDRT or MAD may still be the best choice of parameter for measuring ISV. In the light of the high intercorrelation between these measures, a researcher could do worse than to choose their parameter based on psychometric reliability, and many of the most reliable measures, like SDRT, MAD, and CVRT, were relatively simple variance measures.

It is however important to address some of the limitations of the present study. It is likely that the reliability of RT parameters are somewhat task dependent, and while the battery of tasks used includes some of the more common RT paradigms, it is far from exhaustive. The disparity in number of trials between the different tasks also makes it difficult to compare the different tasks in terms of their reliability. It is also unclear to what extent these data can be extrapolated beyond the group of healthy students investigated here. Further work is necessary to explore whether comparable levels of reliability can be found in clinical and developmental populations.

In conclusion, it appears that ISV can be reliably measured across occasions. It also appears that, due to the high redundancy between many of these measures, psychometric reliability should be a key criterion when deciding how to operationalise ISV. These data also make a strong claim for the importance of using sufficient trials when studying ISV, as trial numbers which have proved adequate in research using first moment parameters may prove inadequate when studying the higher moments of the RT distribution.

Chapter Seven

Electrocortical Correlates of Intra-Subject Variability in Reaction Times:

Average and Single-Trial Analyses

Abstract

Intra-subject variability of reaction times (ISV), long passed over in favour of measures of central tendency, has become increasingly important for cognitive neuroscience. Event-related potentials (ERPs) have identified an inverse relationship between ISV and P3b amplitude; however ERP methods' reliance on averaged waveforms may be unsuitable for studying ISV due to potentially distorting effects of latency jitter. The present study investigates the neural bases of ISV through single-trial analysis of P3bs in groups with low and high ISV, using Infomax-rotated principal components analysis of ERP data. Results indicate that while latency jitter contributes to the reduced P3b amplitude seen in average ERPs of high ISV participants, amplitude differences exist using a single-trial approach that was robust to such artefacts. A decoupling of P3b and RT was also seen in the high ISV group. The results are discussed in the context of the P3b's dimensionality, and its possible catecholaminergic underpinnings.

Introduction

Recent years have witnessed a steep increase in research into intra-subject variability (ISV), that is, the moment-to-moment fluctuations of behavioural performance, typically assessed with the intra-individual standard deviation of reaction times (SDRT). What had been dismissed as meaningless noise under the primacy of information theory (Shannon & Weaver 1949) in the cognitive sciences, has proved a robust predictor of individual differences in general mental abilities (Larson & Alderton, 1990; Schmiedek, Oberauer, Wilhelm, Suess, & Wittmann, 2007); psychiatric (Klein, Wendling, Huettner, Ruder, Peper, 2006; Vinogradov, Poole, Willis-Shore, Ober, & Shenault, 1998) and neurological health (Burton, Strauss, Hultsch, Moll, & Hunter, 2006); as well as cognitive aging (Hultsch, MacDonald, & Dixon 2002) and terminal decline (MacDonald, Hultsch, & Dixon, 2008).

However, controversy about ISV's neural underpinnings was apparent even in early neuroscientific accounts of individual differences in intelligence. While Eysenck (1982; also see Jensen, 1992) thought of SDRT as an indicator of neural noise: poor fidelity neural transmission, possibly related to myelination or synaptic communication failures, Jensen considered SDRT to be a reflection of an "inherent periodicity in the nervous system", related to the refractory period of neurons (Jensen 1982), and, as such, a proper signal. Despite this fundamental difference in the conceptualisation of ISV, Eysenck and Jensen's theories have two key commonalities. First, in relating SDRT to the highly stable trait of general intelligence, both theories assume that ISV is itself a trait. Second, in relating ISV to neural noise or neural periodicities, both theories conceptualise ISV as reflecting a global property of the nervous system.

Regarding the *trait-like character of SDRT*, investigations of individual differences in this measure have yielded equivocal results. Despite finding reliability coefficients in the range of .90 for median RT (MnRT), Jensen (1992) reports mediocre reliability for ISV, with coefficients in

the range of .30-.40. More recent work, however, has found rather more encouraging stability of ISV. Flehmig, Steinborn, Langner, Scholz, & Westhoff (2007) tested a large sample of healthy participants on a battery of multi-modal speeded tasks on two occasions one week apart. They found that, although reliability estimates for ISV still trailed those of mean RT (MRT), they achieved reliability coefficients in the range of .80. Kuntsi, Stevenson, Oosterlan, and Sonuga-Barke (2001) also report SDRT reliability data for a stop-signal task. In this study, reliability coefficients for SDRT (.74) exceeded those for MRT (.66). Johnson *et al.* (2008) also find test-retest correlation coefficients of .75 for ISV on a task measuring sustained attention in healthy children. Our own data (see Chapter Six) also suggest promisingly high test-retest coefficients for a number of metrics of ISV across a variety of tasks, with reliability estimates as high as .92. Interestingly, Rabbitt, Osman, Moore, and Stollery's (2001) findings suggest that some of the unreliability of these measures is caused by the variability itself: higher ISV participants showed poorer test-retest reliability than low ISV participants.

The limited but growing body of literature into the neural bases of ISV can be grouped into anatomical, physiological, and pharmacological evidence (see Kuntsi & Klein, in revision). Anatomically, increased ISV has been found following lesions of the prefrontal cortex, but not following more posterior lesions (Stuss *et al.*, 2003). A negative correlation has also been identified between ISV and white matter volume (Waldhovd & Fjell, 2007), a finding that is compatible with Eysenck's conception of ISV as representing noise in neural transmission.

Physiologically, Bellgrove and colleagues (2004) found that activation in left pre-central, right inferior frontal, bilateral middle frontal, right inferior parietal, and thalamic regions was positively related to ISV during successful response inhibition in a go no-go task. Simmonds *et al.* (2007) employed a similar approach in 8-12 year-old children, finding that inhibition-related activation of the post-central gyrus, anterior supplementary motor area, anterior cerebellum, and inferior parietal lobule correlated negatively, while prefrontal cortex and caudate activity

correlated positively with ISV. These findings sit well with Stuss *et al.*'s (2003) anatomical findings, suggesting that prefrontal areas are important in maintaining behavioural consistency.

Pharmacologically, ISV appears to by modulated by the catecholaminergic system. Increased ISV has been linked to the Val allele of the Catechol-*O*-methyltransferase gene (Stefanis *et al.*, 2005), a polymorphism that leads to faster metabolising of catecholamines and thus weaker catecholamine neuromodulation. Reduced RT variability under methylphenidate is another frequently reported finding (Spencer *et al.*, 2009, Heiser *et al.,* 2004; Teicher, Lowen, Polcari, Foley, & McGreenery, 2004) suggesting catecholamine involvement in determining ISV. ISV has also been linked to dopaminergic activity by MacDonald, Cervenka, Farde, Nyberg, and Bäckman (2009) who used PET in 16 healthy adults. ISV was inversely related to D2 receptor binding in the anterior cingulate, hippocampus and orbitofrontal cortex, but not the striatum, suggesting the involvement of extra-striatal dopaminergic neurotransmission in the modulation of ISV.

Computational models of catecholamine function (Li, Lindenberger, & Frensch, 2000; Usher, Cohen, Servan-Schreiber, Rajkowski, & Aston-Jones, 1999), as well as invasive electrophysiology (Aston-Jones, Rajkowski, Kubiak, & Alexinsky, 1994), suggest that one role of catecholamines is to modulate signal-to-noise ratio, a finding that links back to the idea of ISV as neural noise. Interestingly, said studies also find that decreased phasic catecholamine action can lead to increased output variability.

Not all pharmacological effects on ISV directly involve catecholamines however. Pouget *et al.* (2009) injected GABAergic agonists and antagonists into the dorsolateral prefrontal cortex (DLPFC) of monkeys performing a pro-saccade task. While the agonist, muscimol, had no effect on ISV or the auto-correlational structure of the RT series, the antagonist, bicuculline, resulted in the increased occurrence of slow RTs, reducing RT auto-correlations up to lag 6. These effects appeared to be specific to the DLPFC and could not be replicated by injecting bicuculline into V1

or subcortical structures.

While the previously cited studies identify brain systems, such as the prefrontal cortices and catecholaminergic system, that appear to have a role in regulating behavioural variability, none of these studies established direct and functionally relevant intra-individual brain-behaviour relationships in humans. Given that reaction times only vary in the range of a few hundred milliseconds, only neurophysiological techniques that have an equivalent temporal resolution can identify such direct brain-behaviour relationships (Gerson, Parra, & Sadja , 2005). While we are not aware of any magneto-encephalographic study on ISV, there are, to our knowledge, three studies that exploited the high temporal resolution of electroencephalography (EEG) in the investigation of ISV. Segalowitz, Dywan, and Unsal (1997) analyzed ERP components in head injury patients who performed two tasks: a two-stimulus auditory oddball and an S1-S2 cuing task. They found that P3 and contingent negative variation (CNV) amplitude (in the oddball task and cuing task respectively) explained 83% of the ISV variance in their sample of head injury patients, with reduced amplitudes associated with high ISV. However, maybe due to the restriction of the range of variability in the control sample, no such relationship was found in the healthy control group. Gerson *et al.* (2005) employed a rapid serial visual presentation paradigm and found that a component in the P3 time-range best predicted RT. This study, however, did not focus on individual differences in ISV, focussing more on the functional significance of the P3. Finally, Di Russo and Spinelli (2010) investigated the neuropsychological deficits arising from chronic brain trauma in boxers, as compared to non-athletic and fencing control groups. They found reduced P3 amplitude, P3 latency delay, and increased ISV in the boxer group, relative to both control groups.

These findings suggest that the P3b is an important ERP component for ISV research. The P3b, a late parietal component, has been suggested to reflect a primarily 'strategic' context updating process (Donchin, 1981), but this view is controversial, with Verleger, Jaśkowsi, and

Wascher (2005) suggesting a more 'tactical' role for the P3b in reflecting response selection. Somewhat intermediate views suggest that the P3b is made up of separate subcomponents representing stimulus evaluation and response selection (Pritchard, Houlihan, & Robinson, 1999), although this position is itself disputed (Dien, Spencer, & Donchin, 2004).

More recent models have taken a different approach to explaining the P3b, with Nieuwenhuis, Aston-Jones, and Cohen (2005) suggesting that the P3b reflects the proliferation of a decision signal originating in the locus coeruleus, the origin of the noradrenergic system.This explicit link between the P3b and a major catecholaminergic system is particularly interesting in this context, due to the apparent influence of catecholamines on ISV and signal-to-noise ratio in neural communication.

While the above findings suggest that the P3b provides a promising area for the study of ISV, there are features of the P3b that raise practical problems for such an approach. The P3b is an unusual component in that it appears, to similar degrees (Verleger, Jaśkowsi, & Wascher, 2005), in both stimulus- and response-locked average ERPs presumably because the P3b represents a process intermediate between stimulus processing and response planning, and is dependent on some but not all of the factors that determine RT (Kutas, McCarthy, & Donchin, 1977). As the P3b shows significant response-locking, this suggests that participants with increased ISV may show greater variability of P3b latency. Such increased latency variability, or *latency jitter*, presents problems for standard average ERP approaches, as latency jitter can distort not only measures of latency, but also of amplitude and morphology (see Spencer, 2005). This raises the possibility that the apparent relationship between P3b amplitude and ISV could be an artefact of a relationship between ISV and latency jitter. For this reason, this study will employ a single-trial approach to ERP analysis.

Based on these considerations, the present study aims at investigating the neurophysiological basis of ISV in a two-fold manner. *First*, considering ISV as a behavioural

trait we investigate the electro-cortical correlates of this trait, hypothesising that highly behaviourally variable as opposed to relatively stable participants will exhibit reduced P3b amplitudes that will be robust to control of the increased P3b peak latency jitter that we expect to see in high ISV participants, relative to that of the low scorers. *Second*, considering behavioural variability as an emergent feature of dynamic brain systems, potentially involving catecholaminergic fronto-striatal systems and linked to neural noise, we investigate these intra-individual brain-behaviour relationships by linking the catecholaminergically sensitive single-trial P3bs to RT on a trial-by-trial basis.

Methodology

Ethical approval was obtained from the School of Psychology ethics and research governance committee at Bangor University.

Participants

Participant recruitment was accomplished on the basis of the procedure described in Chapter Six, in which a sample of 87 students performed a battery of neuropsychological tasks on two occasions one week apart. Participants were recruited for the present study on the basis of having particularly high or low ISV, quantified using standard deviations of reaction times (SDRT), for a visual 1-Back Task (1BT; see below) on both sessions of behavioural screening. Participants were ranked in order of SDRT and were invited on the basis of this list. Those with the highest SDRT formed the high-scorer group and those with the lowest SDRT formed the low-scorer group. Participants were excluded if they had any psychiatric or neurological diagnoses (1 with multiple sclerosis, 1 with epilepsy, both were high scorers), if their SDRT in the present task fell into the opposite group's range (1 high scorer, 2 low scorers), or if their EEG data contained

too many artefacts to compute averages (1 high scorer, 1 low scorer). This left 13 high scorers (mean age: 22.9 ± 3.5 years, 5 males, 1 left-handed) and 13 low scorers (mean age: 20.8 ± 2.2 years, 6 males, 3 left-handed). Table 1 summarises each group's performance on the screening tasks. All participants gave written informed consent to participate after being fully briefed on the study.

Group	RTSD Pretest 1				RTSD Pretest 2			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Low Scorers	75.86	13.98	50.06	100.02	62.49	13.92	41.71	86.55
High Scorers	167.88	51.35	116.32	274.12	170.59	52.51	112.39	331.57

Table 1. Reaction time standard deviations for both groups on screening phase.

Apparatus and Materials

Stimuli were presented on a 17'' monitor (with a copper-shielded power unit), connected to a stimulation PC running E-Prime V. 1.2 (Psychology Software Tools, USA). EEG activity was recorded with a sampling rate of 500Hz and a low pass filter of 250Hz, using 63 Ag/AgCl electrodes (Falk Minow, Munich) of the international 10-10 system (American Electroencephalographic Society, 1991) attached over both hemispheres, with Cz as the recording reference, AFz as ground, as well as two infra-orbital electrodes for a total of 65 electrodes. Impendences were reduced to < 5kΩ and electrodes were connected to two BrainAmp DC amplifiers (0.1μv resolution, Brain Products Germany), which were connected to a recording PC running Brain Vision Recorder (Brain Products, Germany). Data were collected inside a sound-attenuated Faraday cage.

Stimuli and Procedure

Participants performed a 1-back working memory task. Here, participants watched a series of letters appear on the screen, responding to each with a left-handed key press unless a letter matched the preceding letter, in which case they responded with a right-handed key press. Non-repeated letters (80% of trials) will be called "non-events" and repeated letters (20% of trials) will be called "events". Participants were asked to respond as quickly and as accurately as possible.

Stimuli were white letters in Arial typeface (visual angle approx. 3°), presented on a black background. Stimulus duration was 500ms and stimulus-onset asynchrony was 2,000ms. The experiment was divided into ten blocks, each containing 150 pseudo-randomly fixed-ordered trials, for a total of 1,500 trials, plus 20 practice trials given before the first block. Participants were given a ten-minute break after the fifth block.

Data Analysis

EEG analysis was primarily carried out using Brain Vision Analyzer (Version 2.0, Brain Products, Germany) while Dien's ERP-PCA toolkit (Dien, 2009) was used for the principal components analysis. Due to the complexity and number of processing steps, the analysis is summarised in Table 2 below.

Data were average referenced and 0.1-50.0Hz filtered with symmetrical 24dB per octave slopes. Any data contaminated by direct current offset corrections made during recording were rejected before an Infomax independent component analysis, trained on 200s of data starting from 100s into each dataset, was applied to the data. Components representing eye activity, heartbeats, muscle activity, movement artefacts, or bad channels were removed before back-projection (Jung *et al.*, 2000).

	Process	Details
1	Compute average ERPs	Raw data preprocessed and stimulus-locked and response-locked average ERPs derived.
2	Infomax-PCA	PCA with Infomax rotation applied to concatenated stimulus-locked averages, to identify electrode weightings for P3b factor.
3	Single-trial analyses	Single-trial P3bs identified for P3b factor. Medians and SDs, of amp. & lat., and correlation coefficients with RT for amp. and lat., calculated for each participant. This process conducted separately for stimulus-locked and response-locked P3bs.

Table 2. Summary of analysis steps.

Any sections of data where amplitude still ranged by more than 100μv, or less than 0.5μv, within 200ms were marked as bad. The data were then filtered 0.5-4Hz, and data for event trials were initially segmented into stimulus-locked epochs of 1300ms pre-stimulus to 1650ms post-stimulus, with the period between -600 and -400ms in each epoch used as a baseline. Epochs containing any data marked as bad, those based around an incorrect response, or those where the response was faster than 120ms or slower than 1400ms were discarded. Separate stimulus-locked and response-locked segments were then cut from the initial segments, so that the baseline would be the same for stimulus and response-locked segments. The stimulus-locked segments ranged from -600 to 1400ms, relative to stimulus-onset, while the response-locked segments ranged from

-1400 to 600ms, relative to response. Trials that featured only as stimulus or response-locked segments, but not both, were removed so that analyses were carried out on the same trials.

Conventional average ERPs were first derived by separately averaging stimulus-locked and response-locked trials. The individual average P3b peaks were defined as the maximum positive voltage between 250 and 750ms post-stimulus, while average P3b peaks were defined as the maximum positive voltage between 250ms pre-response and 250ms post-reponse. Latency and amplitude (measured as the mean amplitude of a 20ms window centred on the peak) measurements were obtained for each participant at channels Cz, CPz, Pz, and POz. Separate latency and amplitude measurements were made for stimulus and response-locked averages (Step 1 in Table 2).

The data were then prepared for single-trial analysis. Due to volume conduction, recorded EEG represents a mixture of signals originating from all over the brain. This activity includes not only the evoked activity that is the subject of most ERP research, but also spontaneous EEG, muscle activity, eye movements, and all manner of miscellaneous electrical noise. This means that the signal-to-noise ratio of single-trial ERPs is generally very poor. While this problem has been traditionally overcome by averaging activity from single electrodes that is time-locked to a certain event, this approach is obviously unsuitable for single-trial analysis. Instead of aggregating information across trials, we sought to integrate information across electrodes, and measure activity that is closer to the underlying EEG sources.

A statistical signal processing approach that has shown promising results for isolating independent sources of activity in EEG data is the Infomax algorithm (Bell & Sejnowski, 1995).The Infomax linearly transforms recorded channels into maximally independent factors[5] of

---

[5] While the term *components* is generally used to refer to the latent variables derived from Infomax rotation, we will follow Dien *et al*'s (2007) lead and use the term factors, to avoid confusion with the standard meaning of the

the data. This essentially operates as a principal components analysis (PCA) rotation, although it is not necessary to carry out a PCA before applying the algorithm. Unlike more conventional PCA rotations however, the Infomax does not optimise the amount of variance explained, but rather optimises the statistical independence of the data.

While it is perfectly legitimate to apply the Infomax to raw data, we decided to use PCA as a pre-processing step (Dien, Khoe, & Mangun, 2007). This decision was made for several reasons. Firstly, as Eichele *et al.* (2008) point out, the Infomax is not best suited for deriving general factors across participants, but rather is generally used on a participant-by-participant basis to identify 'bespoke' sets of factors. Employing a PCA beforehand allows the Infomax to operate on the data in factor space, rather than electrode space, allowing the same factors to be identified for the whole group. Secondly, where more 'traditional' PCA rotations, such as the Promax, tends to err on the side of parsimony, the Infomax will tend to separate factors wherever it can. Thus where the Promax runs the risk of conflating genuinely separate factors, the Infomax runs the risk of 'over-splitting' factors (Dien *et al.*, 2007). Särelä and Vigário (2000) suggest PCA as a possible pre-processing step for dimensionality reduction where over-splitting may occur, and we also used PCA for this reason.

Stimulus-locked participant averages were concatenated before a spatial PCA was run on the data using Dien's ERP-PCA toolkit (Dien, 2009). Based on a parallel Scree test (Horn, 1965), seven factors were extracted and rotated using the Infomax algorithm. Figure 1 shows factor topographies, with the depth of colour representing the strength of the weighting of each electrode given by the factor pattern matrix. Only factor 1 shows a clear P3b topography and so all further analyses are based upon this factor (Step 2 in Table 2).

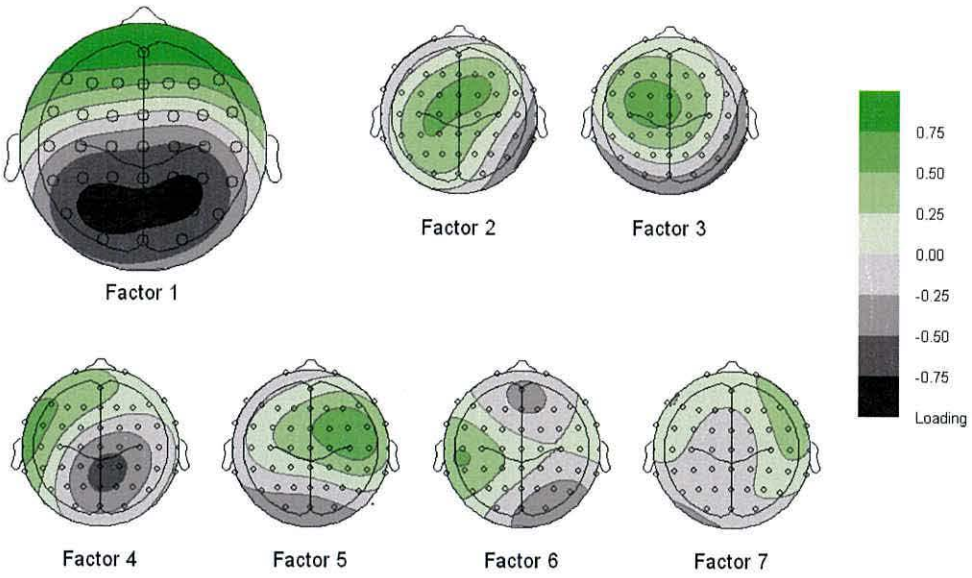

Fig. 1. Topographic maps of Infomax rotated factors. Depth of colour represents strength of weighting at each electrode with positive weightings in green and negative weightings in black (Lambda =0.0001).

The factor pattern matrix for factor 1 was then applied to the single-trial data. This transforms the data from the electrode space to factor space, with the amplitude at each timepoint now representing a linear sum of the voltage of all 65 electrodes, weighted by the loadings of the factor matrix. P3b peaks were then picked for each trial, defined as the time-point with the maximum voltage (the Infomax rotation renders polarity arbitrary, so the decision of whether to pick positive or negative voltages was based upon the direction of the P3b in the average ERP for factor 1) between 250 and 750ms post-stimulus. Latency and amplitude information for each trial were recorded and used to create 'jitter-free' amplitude and latency measurements (median single-trial amplitude and latency), as well as measurements of amplitude and latency variability (intra-participant standard deviation of amplitude and latency). The relationship between single-trial ERP parameters and RT was measured by computing within-participant Pearson's correlation coefficients between RT and peak latency and amplitude.

In order to assess whether any potential group differences in latency jitter could be explained purely by response-locking of the P3b, the factor pattern matrix was also applied to the response-locked segments, and peaks were picked from 250ms pre-response to 250ms post-response. Analogous measures to the stimulus-lock single-trial analysis were derived, with latency in this case being relative to response rather than stimulus onset (Step 3 on Table 2).

Inferential statistics

SPSS version 16.0 (SPSS Inc., USA) was used to conduct inferential statistics on the data. A repeated measures ANOVA was carried out for the behavioural variables, GROUP was the between-participants factor, TRIAL TYPE was the within-participants variable, and MRT, MnRT, SDRT, coefficient of variation in reaction times (CVRT = SDRT/MRT) and accuracy (correct trials/total trials) were the dependent variables. This analysis was then rerun as an ANCOVA, with SDRT and CVRT as dependent variables and MnRT as a covariate, in order to confirm that the differences in ISV were not dependent on MnRT.

A repeated measures ANOVA was run on the traditional averaged ERPs, with GROUP as the between-participants factor, LOCKING (stimulus and response) ELECTRODE (Cz, CPz, Pz, and POz) as within-participants factors, and P3 amplitude and latency as dependent variables. Greenhouse-Geisser correction was used where appropriate.

A MANOVA was run on the single-trial ERP data, with GROUP as a between-participants factor and medians and standard deviations of latency and amplitude, intra-participant Pearson's correlation coefficients between RT and latency, and RT and amplitude; and finally the number of trials used in each STA as dependent variables. Separate versions of each variable were included for stimulus and response-locked data.

In order to avoid an increased risk of Type I Errors stemming from multiple comparisons, the Holm-Bonferroni procedure (Holm, 1979) was used to modify the significance criterion.

Results

Behavioural data

	Low		High	
	Event	Non-event	Event	Non-event
Accuracy	0.73 ± .125	0.97 ± .007	0.74 ± .153	0.97 ± .009
MRT	413 ± 29	341 ± 29	555 ± 105	463 ± 134
SDRT	69 ± 17	71 ± 14	121 ± 31	124 ± 22
CVRT	.16 ± .04	.21 ± .04	.22 ± .04	.27 ± .04

Table 3. Descriptive statistics for behavioural data.

As Table 3 shows, high-scorers responded significantly slower (MRT: $F_{(1,24)}$=15.576, p=.001, h_p^2=.394; MnRT: $F_{(1,24)}$=10.526, p=.003, h_p^2=.305) and more variably (SDRT: $F_{(1,24)}$=53.923, p<.001, h_p^2=.692; CVRT: $F_{(1,24)}$=20.357, p<.001, h_p^2=.459) than low-scorers, confirming that the groups were distinct in terms of ISV. There were no significant differences between groups on accuracy ($F_{(1,24)}$= .033, p = .857, h_p^2=.001). Both groups were significantly slower ($F_{(1,24)}$=100.469, p<.001, h_p^2=.807) and less accurate ($F_{(1,24)}$=72.845, p<.001, h_p^2=.752) for events than non-events. TRIAL TYPE had no effect on SDRT (F<1), but CVRT was significantly higher for non-events than events ($F_{(1,24)}$=24.600, p<.001, h_p^2=.506). No behavioural variables showed significant GROUP x TRIAL TYPE interactions.

Group differences remained highly significant for both variables after controlling for MnRT with the ANCOVA (SDRT: $F_{(1,23)}$=30.298, p<.001, h_p^2=.568; CVRT: $F_{(1,23)}$=27.361, p<.001, h_p^2=.543), confirming that there were genuine differences in ISV between groups that were not attributable to MnRT.

Electroencephalographic data

Figures 2 and 3 respectively show the traditional average stimulus-locked and response-locked waveforms for each group at all four electrodes, as well as topography maps. P3b amplitudes were greater for low scorers and maximal at Pz for both sets of averages. Table 4, displaying descriptive statistics for the average P3bs, confirms these findings. The ANOVA for the average ERPs found a significant effect of GROUP on amplitude ($F_{(1,24)}$=11.103, p=.003, h_p^2=.316), but not latency ($F_{(1,24)}$=1.320, p=.262, h_p^2=.052).

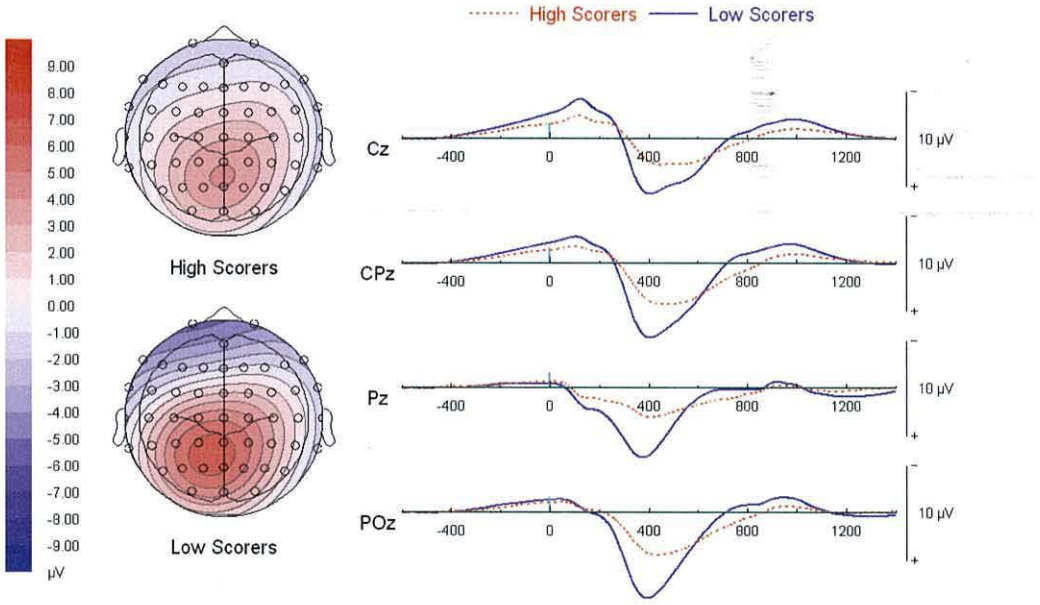

Fig. 2. Standard stimulus-locked P3bs for low and high scorers.

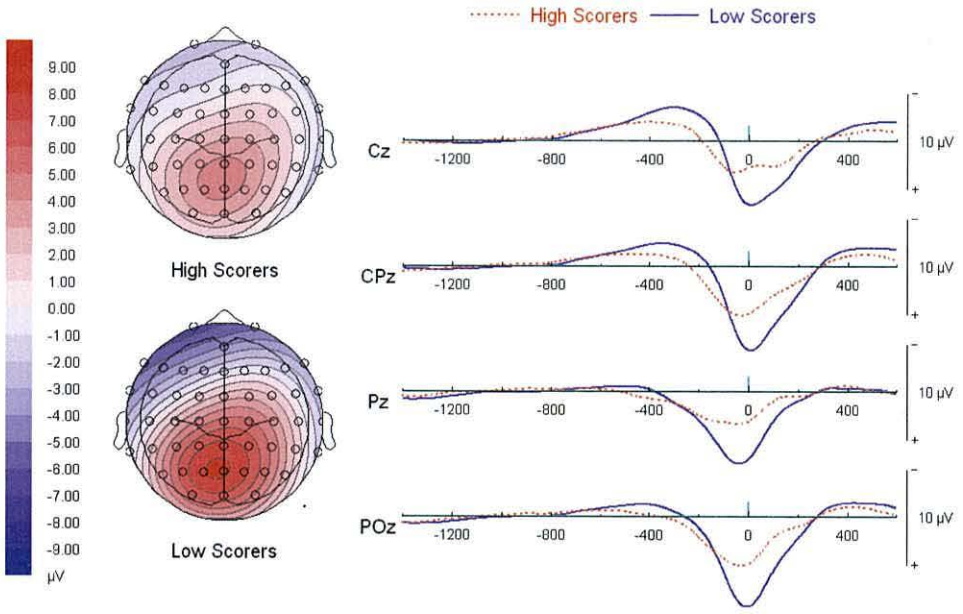

Fig. 3. Standard response-locked P3bs for low and high scorers

			Low	High
Cz	Stimulus-locked	Latency (ms)	396 ± 35	437 ± 125
		Voltage (μv)	5.94 ± 2.78	2.58 ± 2.48
	Response-locked	Latency (ms)	-9 ± 29	-54 ± 89
		Voltage (μv)	6.95 ± 3.01	2.80 ± 3.28
CPz	Stimulus-locked	Latency (ms)	396 ± 35	437 ± 125
		Voltage (μv)	8.03 ± 3.18	4.62 ± 3.35
	Response-locked	Latency (ms)	-9 ± 29	-54 ± 89
		Voltage (μv)	8.96 ± 3.62	5.01 ± 3.83
Pz	Stimulus-locked	Latency (ms)	396 ± 35	437 ± 125
		Voltage (μv)	9.23 ± 3.62	5.12 ± 3.15
	Response-locked	Latency (ms)	-9 ± 29	-54 ± 89
		Voltage (μv)	9.67 ± 3.91	5.36 ± 3.63
POz	Stimulus-locked	Latency (ms)	396 ± 35	437 ± 125
		Voltage (μv)	7.36 ± 3.86	3.71 ± 3.05
	Response-locked	Latency (ms)	-9 ± 29	-54 ± 89
		Voltage (μv)	7.35 ± 3.86	3.67 ± 3.36

Table 4. Descriptive statistics for traditional averaged P3bs.

Figure 4 shows the grand average waveforms for each group on factor 1. The top waveform presents a standard stimulus-locked average, the middle waveform presents a latency-adjusted average – where trials were time-locked to peak amplitude before averaging, and the

bottom waveform represents a response-locked average. The stimulus-locked and response-locked averages (top and bottom respectively) show what appears to be a group amplitude difference, but the apparently increased P3b duration in high scorers could be indicative of distortion caused by latency jitter. The latency-adjusted average (middle) suggests, however, that this amplitude difference is not just an artefact of latency jitter.

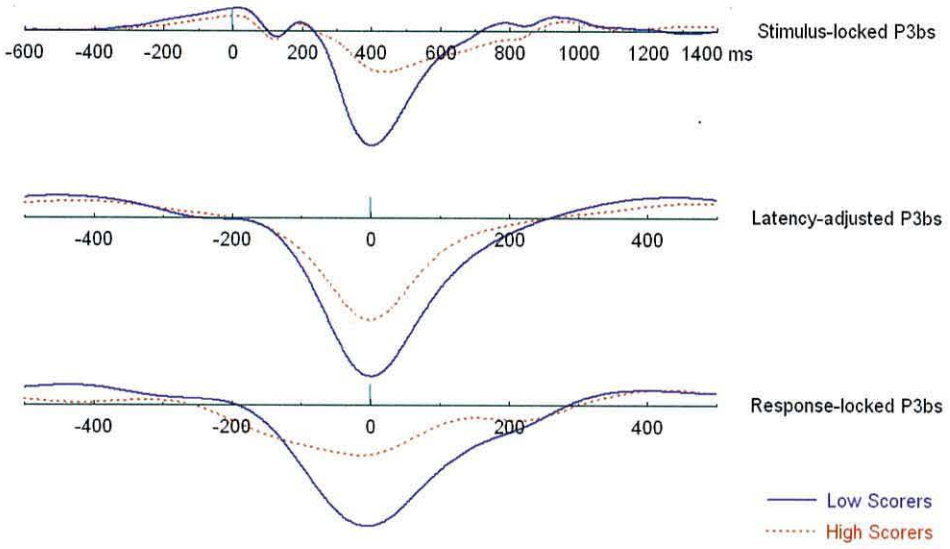

Fig. 4. Stimulus-locked, latency-adjusted, and response-locked group average waveforms for factor 1. Amplitudes represent the sum of voltage at all electrodes, weighted by the factor pattern matrix

The inferential statistics for the single-trial analysis found group effects on median latency, with shorter latencies for low scorers ($F_{(1,24)}$=4.518, p=.044, h_p^2=.158); median amplitude, with larger amplitudes for low scorers ($F_{(1,24)}$=5.082, p=.034, h_p^2=.175); standard deviation of latency, with greater latency jitter in high scorers ($F_{(1,24)}$=13.411, p=.001, h_p^2=.358); and amplitude variability, with the unexpected finding of increased variability in low scorers ($F_{(1,24)}$=4.865, p=.037, h_p^2=.169).

There was also a medium to large correlation between P3b latency and RT in low scorers (average correlation=.34) but only a weak correlation between these variables in high scorers (.16), a group difference which reached statistical significance ($F_{(1,24)}$=4.796, p=.038, h_p^2=.167). Coefficients between P3b amplitude and RT, however, were close to zero for both groups, and as such no group differences were found (F<1).[6]

Results for the response-locked data were similar to stimulus-locked results (see Table 5). Interestingly the increased standard deviation of P3b latency seen in the stimulus-locked data of the high ISV group was also found in the response-locked data ($F_{(1,24)}$=14.607, p=.001, h_p^2=.378). Correlations between response-locked P3b latency and RT were slightly higher for low scorers (average correlation=.12) than high scorers (.04), but this difference did not reach significance. P3b amplitude and RT were again weakly associated in both groups. Finally, there were no differences between groups in terms of the number of trials used (F<1), suggesting comparable signal-to-noise ratio. Descriptive statistics for stimulus-locked and response-locked single-trial data can be found in Tables 4 and 5 respectively.

Discussion

To our knowledge, the present study is the first to identify differences in ERPs between healthy participants grouped on the basis of temporally stable individual differences in reaction time variability. Significantly increased P3b latency jitter was seen in participants with high ISV

---

[6] As the range which we picked P3bs from was from 250-750ms, and high scorers would have a greater number of RTs after 750ms, this could be seen as placing an artificial limit of the strength of the correlation coefficients between RT and P3b latency in high scorers. In order to assess the impact of this, we recomputed correlations after removing trials where RTs were longer than 750ms. This only increased high scorer coefficients from .16 to .18, suggesting that the group difference is not merely an artefact of peak picking range.

compared to those with low ISV. Despite rigorous measures to control for the confounding effects of latency jitter, significant differences between high and low ISV participants in amplitude were found in the P3 time range for several latent variables within the data. Our findings extend work by Segalowitz *et al.* (1997), showing that a relationship between ISV and P3b amplitude can be found in healthy participants, and that such a relationship, while exaggerated by latency jitter distortion, is not merely an artefact of it. Reduced correlation coefficients were also found between RT and P3b latency in the high ISV group, compared to low ISV participants. No systematic relationship was seen in either group between RT and P3b amplitude however. This discussion will examine what these results suggest about the neural correlates of ISV, and evaluate the strengths and weakness of our methodology.

What can we infer about the neural correlates of ISV?

The data show reduced amplitude of P3b in high ISV participants, even when controlling for latency jitter with the single-trial analysis. One explanation for this comes from computational models of catecholaminergic modulation of signal-to-noise ratio in neural networks. Li *et al.*'s (2000) model of catecholamine function found that by reducing catecholamine activity, modelled as a gain parameter, output became more variable, possibly analogous to an increase in ISV, and within-network variability also increased. Usher *et al.* (1999) similarly found that in order to mimic empirical findings with their model of noradrenergic attention it was necessary to simulate electrotonic coupling, thereby increasing spatial coherence. These findings become important when we consider than the P3b, like all EEG measureable at the scalp, is generated by the synchronised firing of thousands of cortical macrocolumns. Simulations, such as Nunez and Srinivasan's (2006), show that a key determinant of the scalp potential amplitude is the degree of spatial aggregation; holding the overall level of activity constant and varying the degree of spatial

coherence has dramatic effects on scalp amplitude. Taken together, the putative increase in within-system variability and the dependence of ERP amplitude on spatial coherence could well explain the reduced P3b amplitude seen in the high ISV group.

Another possibility is that high scorers found the task more difficult than low scorers. There is evidence that P3 amplitude is reduced by increasing task difficulty (Comerchero & Polich, 1999), increasing working memory load (Watter, Geffen & Geffen, 2001), and by introducing concurrent tasks (Nash & Fernandez, 1996). Johnson (1986) links these effects to attention allocation and equivocation, an information theoretic concept measuring the fidelity of information transfer. While we did not directly manipulate task load or difficulty in the present study, our grouping strategy may have had this effect implicitly. It is, however, important to note that P3 reduction as a result of poor information transfer is generally found when task difficulty is increased enough to impact performance (Comerchero & Polich, 1999), whereas the present study found differences in the absence of accuracy differences.

A final explanation, similar to the first account but on a higher scale, is that the apparent fall in P3b amplitude is best viewed as a decoupling of stimulus and response-related subcomponents of the P3. It has been hypothesised that the P3b is composed of separate 'subroutines', some related to stimulus-evaluation and others related to response-selection (Falkenstein, Hohnsbein, & Hoormann, 1994). When ISV is high, these subcomponents may overlap less than usual, and due to the presumably similarly positioned and aligned dipoles that putatively underpin these subcomponents, this may result in an apparent fall in amplitude.

No less important is our finding that variability of P3b latency is increased for high scorers in both stimulus and response-locked single-trial analyses. Indeed the effect size (h_p^2=.358) was comparable to that found between the ISV scores of ADHD patients and controls by Klein *et al.* (2006). If ISV affected only stimulus evaluation, we might expect to see increased latency jitter in stimulus-locked P3bs alone. If, on the other hand, ISV affected only response

selection, we might expect to just see increased latency jitter in response-locked P3bs. Finding increased jitter in both suggests that individual differences that modulate ISV affect both stimulus-evaluation and response-selection processes. This pattern of results may suggest that ISV is determined by a fairly global property of the nervous system, such as neuromodulation, as suggested above, or myelination (Walhovd & Fjell, 2007).

Another interesting finding, potentially related to the previous one was the group difference in the strength of the correlation between P3b latency and RT. While low scorers showed a robust relationship between P3b latency and RT with medium-to-large effect sizes, suggesting functional significance of P3b latency, this relationship was far weaker in high scorers. This finding, taken together with the finding of increased latency jitter in stimulus and response-locked P3bs, is suggestive of a global increase in neural noise. An increase in noise in either stimulus evaluation or response selection alone could not have led to our latency jitter findings, while general fluctuations affecting both processes equally would be inconsistent with the reduced high scorer correlation coefficients.

Finally our data revealed an unhypothesised greater P3b amplitude variability in low scorers. This could fit with work by McIntosh *et al.* (2008), who found a negative correlation between behavioural ISV and multiscale entropy and dimensionality in EEG activity, measures which could also be viewed as reflecting increased amplitude variability. This raises interesting questions on the role of neural variability in promoting higher order stability in behaviour. That said, the lack of correlation between P3b amplitude and RT casts doubt on the functional significance of P3b amplitude in determining RT.

What were the strengths and weaknesses of the methodology?

The present study employed a PCA and filter-based approach to denoising the data for single-trial analysis. We feel that this analytic strategy distinguished between latency and amplitude effects that were confounded in the traditional average ERPs. That being said, there remain details of our approach that warrant further evaluation.

Chief among these is our decision to carry out the PCA on average ERPs, rather than on single-trial data. We justify this as a more conservative strategy than applying PCA to single-trial data, an approach which has not been verified by the literature. It is however, important to acknowledge the assumption that is implicit in basing our spatial filters on averaged waveforms, namely that the single-trial topography of the ERPs of interest is captured by the average waveform. We would argue that while potentially interesting induced activity would presumably be poorly characterised by time-domain average ERPs, the P3b is an evoked component and its topography in average ERPs should at least be a good approximation of its topography in single trials.

A second important feature of our analysis was that a single set of spatial filters was derived for all participants. A case can be made for carrying out separate PCAs for the two groups, or even for each individual participant, as these 'bespoke' spatial filters would presumably be an optimal fit for each group or participant's data. Equating the factors derived from such an approach across groups and participants would however be fraught with ambiguity and the 'one-size-fits-all' approach was chosen on the grounds of clarity, interpretability, and ease of comparison between groups.

Finally, drawing a link between ISV, its electro-cortical correlates and one of the best established cognitive correlates of ISV, psychometric intelligence (e.g., Schmiedek *et al.*, 2007) would have provided another approach to the neural basis of human intelligence. This attempt

was not undertaken, however, as our participant sample comprised of samples of highly pre-selected and rather IQ-homogeneous university students, and because the participants tested here already participated in a rather demanding series of three testing sessions.

Conclusion

Measuring the ERPs of participants who, by definition, show unusually high variability in the timing of cognitive processes is something of a technical challenge, as most ERP techniques assume strong time-locking of cognition to stimulus or response. We controlled for this by conducting a PCA-denoised single-trial analysis to obtain 'jitter-free' amplitude measurements, as well as estimates of latency jitter. High ISV was found to be associated with not only increased latency jitter, but also reduced amplitude of a factor with P3b topography. These findings are consistent with the idea of ISV as a measure of noise in neural processing. We believe this paper makes a strong case for implementing single-trial analyses, not just in studies of ISV, but also in ERP studies comparing populations where ISV has been shown to be elevated (ADHD, schizophrenia, older adults *etc.*) to controls.

Chapter Eight

On the Unity of Variability:

Structuring Individual Differences in Intra-Subject Variability of Reaction Times Across Cognitive Tasks and Sensory Modalities

Abstract

While most reaction time research uses measures of central tendency (*e.g.* means, medians), there is evidence that intra-subject variability of reaction times (ISV) has the potential to reveal additional information about neural processing. While most research has assumed that ISV is a relatively unitary construct, this assumption has not been extensively tested. This study employed a factor analytic approach to assess the validity of this assumption. Seventy-one participants carried out visual and auditory versions of simple and choice reaction time tasks. A single factor explained the majority of the variance in ISV, suggesting that ISV was largely unitary, with only minor secondary factors dividing auditory and visual tasks, and explaining additional variance in simple reaction times. The present study suggests that while the source of individual differences in ISV may interact with the brain's sensory systems, it should be investigated primarily at a supra-modal level.

Introduction

While the study of inter-individual differences has been key to empirical psychology ever since Galton (1869/1881) and Stern's (1911) seminal publications, the systematic investigation of intra-individual variability has proven less popular, despite programmatic suggestions (Fiske & Rice, 1955) and recurrent calls to focus on the subject (*e.g.*, Jensen 1992). This neglect is possibly a consequence of the long tradition in empirical psychology of focussing on the central tendencies of behavioural measures such as reaction times (RTs; Posner, 1978) which considers within-person variability in RTs as noise that is unrelated to the phenomena under investigation. Time and again, however, psychological theorists have emphasised that RTs carry meaningful information in addition to what is conveyed by measures of central tendency such as arithmetic mean or median (henceforth collectively called *AvRT*), and that intra-subject variability of reaction times (ISV), measuring inconsistency in performance across time, may reflect important cognitive and neural processes underlying RTs (*e.g.*, Jensen, 1992).

Indeed, outside the domain of traditional experimental research, ISV has recently become a key topic. Increased ISV, relative to controls, has been found in a number of psychiatric disorders, including attention deficit hyperactivity disorder (ADHD; Klein, Wendling, Huettner, Ruder, & Peper, 2006), schizophrenia (Vinogradov, Poole Willis-Shore, Ober, & Shenault, 1998) and Alzheimer's dementia (Burton, Strauss, Hultsch, Moll, & Hunter, 2006). Its importance has been demonstrated in relation to a variety of cognitive constructs such as IQ (Larson & Alderton, 1990), working memory (Schmiedek, Oberauer, Wilhelm, Süβ, & Wittmann, 2007), and executive function (Unsworth, Redick, Lakey, & Young, 2009). ISV is also a key area in gerontologic research, where it is thought to be a marker of cognitive ageing (Hultsch, MacDonald, & Dixon, 2002; MacDonald, Nyberg, Sandblom, Fischer, & Bäckman, 2008) and

even a predictor of impending death in longitudinal designs (MacDonald, Hultsch, & Dixon, 2008).

Sceptics might suggest that increased ISV in low IQ, psychiatric disorders, and cognitive ageing is merely an artefact of increased AvRT, and thus at best of secondary interest. The evidence suggests otherwise. Jensen (1992) carried out a principal components analysis (PCA) on RT data. He found that while the largest component included the variance shared by MnRT and SDRT, the second largest showed bipolar loading on these variables, suggesting that they measure correlated but independent constructs. This second component loaded more heavily on SDRT than MnRT, suggesting that SDRT was capturing additional variance. Jensen (1992) and Klein *et al.* (2006) argue that differences in ISV drive differences in AvRT. Due to RTs being capped at the fast end of the distribution by physiological constraints but largely uncapped at the slow end, an increase in variability can lead to an increase in AvRT, but not *vice versa*. Furthermore, it appears that individual differences in RTs are much larger in the tail of the RT distribution than in the faster end (Unsworth *et al.,* 2009). Measures of dispersion, such as the standard deviation, are inherently more sensitive to extreme scores than measures of central tendency and thus such measures better represent these tail RTs.

There is, however, a lack of consensus on whether ISV represents a unitary construct or not. Are individual differences in ISV caused largely by a single factor, or several separable factors? Is ISV a global property of the central nervous system, or is it local to specific networks? A number of neuroscientific models of ISV: myelination (Jensen, 1992; Russell *et al.*, 2006), Signal-to-noise ratio (Li, Lindenberger, & Frensch, 2000), and default mode interference (Sonuga-Barke & Castellanos, 2007) accounts; assume fairly unitary causes of ISV, implying that the construct has a certain general unity. In contrast, several articles from the aging and head injury literatures have argued explicitly that ISV is not unitary (Stuss, Pogue, Buckle, & Bondar, 1994; Shammi, Bosman, & Stuss, 1998; & Strauss, MacDonald, Hunter, Moll, & Hultsch, 2002).

While much of this disagreement may stem from different operational definitions of ISV – all three papers look at variability across occasion, as well as or instead of within occasion, while other areas of the ISV literature tend to focus on within occasion – it remains an open question. The issue of ISV's unity, or lack thereof, is a question well suited for psychometric approaches, but there has been a paucity of studies employing these approaches to ISV. MacDonald, Hultsch, and Dixon (2003) report using a factor-weighted composite of ISV scores for data reduction purposes, suggesting that ISV can be meaningfully reduced to a single factor, but do not provide details on the factorial structure uncovered. Klein *et al.* (2006), however, found that although one component explained the variability data across tasks in ADHD patients, two were needed in controls.

Another issue with the assumption that ISV is unitary is that previous ISV research has employed tasks using only stimuli from only one sensory modality (usually vision). When only one modality is used to deliver stimuli, it is possible that the variability of interest enters at a modality-specific stage of processing, rather than centrally. If this were the case, one could argue that ISV is local to specific subsystems, rather than a global or central property of the brain.

It is also important for the question of unity to assess whether ISV from different types of speeded task load onto the same factor, although MacDonald *et al*'s (2003) use of a data reduction approach suggests that it does. Both cognitive modelling (Ratcliff & Tuerlinckx, 2002) and event-related potential (Vogel & Luck, 2000) approaches suggest that tasks featuring a decision component involve different processes than those that do not. While these processes may be more central than the early modality-specific processing mentioned above, the lack of a common ISV factor related to both of these types of response would cast doubt on the view of ISV as a global construct.

The present study seeks to assess whether ISV can be considered a global and unitary construct. This will be achieved by employing a latent variable approach to the structure of ISV

across modalities (vision and audition) and tasks (simple and choice RT). If ISV represents a unitary construct, a single factor should explain variance not only across different tasks, but also across different sensory modalities. However, if multiple significant factors emerge, this may indicate that there are multiple sources of ISV that may be task- and modality-specific, which future theories of ISV would need to account for.

Methodology

Participants

Seventy-six students from Bangor University participated in exchange for course credit. Five participants were excluded due to performing at below chance accuracy on one or more tasks leaving 71 participants (age 21.9 ± 4.8 years, 70 right-handed, 28 male) in the final sample. Permission to conduct the study was obtained from the school ethics and research governance committee. All participants gave informed consent prior to participating and were debriefed afterwards.

Apparatus and materials

The experiment was conducted on a PC running E-Prime (Psychology Software Tools Inc., Pittsburgh, U.S.) Visual stimuli were presented on a 17" monitor and auditory stimuli were presented using closed-back headphones. Responses were made using a standard QWERTY keyboard.

Stimuli and procedure

Participants carried out visual and auditory versions of two tasks, a simple reaction time task (SRT) and choice reaction time task (CRT) for a total of four tasks (task order was counterbalanced). All tasks had 150 pseudorandomly ordered trials (the CRTs also had five practice trials). In all tasks, stimuli were presented for 500ms and the participant had until the subsequent stimulus appeared to respond. Stimulus onset asynchrony varied randomly between 1500ms and 2500ms, with a mean asynchrony of 2000ms. Participants were asked to respond to all stimuli as quickly and as accurately as possible.

In the auditory SRT (AS), participants listened to a series of 640Hz sinusoidal pure tones, and responded to each with a press of the space bar. In the visual SRT (VS) participants watched a series of white circles (visual angle approximately 3°) and responded to each with a press of the space bar.

In the auditory CRT (AC) participants listened to a series of 840Hz and 440Hz sinusoidal pure tones and responded with a different key to each (response hands were counterbalanced across participants). In the visual CRT (VC) participants watched a series of blue (L=100, A=8, B=-60) and yellow (L=100, A=8, B=60) circles (visual angle approx. 3°) and responded with a different key to each (response hands were counterbalanced across participants).

Data Analysis

An accuracy measure was derived for each participant on each task (correct trials/all trials) before trials where responses were incorrect, absent, or pre-emptive (RT<120ms) were filtered out of the data. MnRTs, and SDRTs were then computed for each participant on each task.

Repeated measures ANOVAs were run with MODALITY (auditory and visual) and TASK (simple and choice) as within subject factors and MnRT and SDRT as dependent variables.

To assess the internal reliability of the measures, separate MnRTs, and SDRTs were derived for odd and even trials, and odd-even correlation co-efficients were computed for each variable on each task. SDRTs for the four tasks were then correlated with one another and a principal axis factoring with oblimin rotation (kappa = 0) was run on the resulting correlation matrix (normal Q-Q plots of all SDRT variables suggested non-normal distributions, so principal axis factoring was used rather than maximum likelihood estimation, as not to violate the normality assumption of the latter technique). An oblique method of rotation was chosen as the factors underlying ISV could plausibly be correlated and under these circumstances an oblique rotation would give more accurate results.

The question of how many factors to extract is a controversial one. The Guttman-Kaiser criterion (Guttman, 1954; Kaiser, 1960) recommends extracting only those factors which have eigenvalues greater than one, while Jolliffe (1986) suggests 0.7 as a cut-off. Another approach is to identify the point of inflection on a scree plot and extract all preceding factors (Cattell, 1966). These approaches, however, are primarily useful for conducting psychometrically sound data reduction, and as our study was concerned with whether ISV was unitary or not, it seemed more appropriate to extract the maximum number of factors (in this case three), assess whether a single factor explained the vast majority of the variance, and identify where the remaining variance loaded.

Results

Table 1 shows medians and standard deviations for the four tasks. The repeated measures ANOVAs confirmed that participants were slower ($F(1,70) = 702.827$, $p = .000$, $h_p^2 = .909$) and

more variable ($F(1,70) = 83.444$, $p = .000$, $h_p^2 = .544$) in their responses to CRTs than they were to SRTs. They also appeared to have responded slower ($F(1,70) = 5.469$, $p = .022$, $h_p^2 = .072$) and more variably ($F(1,70) = 33.412$, $p = .000$, $h_p^2 = .323$) to the auditory tasks than the visual tasks, which may have been due to a minority of participants who responded much slower to auditory than visual tasks. There was also a significant MODALITY * TASK interaction on SDRT ($F(1,70) = 10.108$, $p = .002$, $h_p^2 = .126$) but not MnRT (Fs<1), reflecting the high SDRTs for the AC task.

	MnRT	SDRT
AS	280.02 ± 20.63	70.24 ± 35.76
VS	262.75 ± 30.11	60.77 ± 27.72
AC	420.01 ± 93.48	106.59 ± 42.98
VC	400.23 ± 47.53	80.69 ± 23.88

Table 1. Descriptive statistics for reaction time parameters.

	MnRT	SDRT
AS	.996	.567
VS	.959	.608
AC	.989	.878
VC	.960	.785

Table 2. Odd-even reliability coefficients for trimmed and untrimmed reaction time parameters.

Table 2 shows odd-even reliability coefficients for MnRT and SDRT on each of the four tasks. MnRTs were uniformly higher than SDRTs, showing reliability coefficients above 0.95 reliability on all tasks. SDRT reliability coefficients are adequate for CRTs, but mediocre for SRTs – presumably due to the large impact of a few very slow trials on the SDRT falling into either the odd or even halves and reducing the reliability. Correlation coefficients, shown in Table 3, were significant between SDRTs on all tasks.

	AS SDRT	VS SDRT	ACSDRT	VC SDRT
AS SDRT	1	.378	.514	.383
VS SDRT		1	.476	.446
ACSDRT			1	.701
VC SDRT				1

Table 3. Raw correlation matrix for SDRTs.

The principal axis factoring revealed that initial communalities were acceptable for CRTs (ACRT: 0.576, VCRT: 0.507) but were low for SRTs (ASRT: 0.287, VSRT: 0.273), probably reflecting the poor reliability of the untrimmed SDRTs. The Kaiser-Meyer-Olkin measure of 0.730 and significant Bartlett's test of sphericity suggested that the data were appropriate for factor analysis. The first factor had an eigenvalue of 2.464, and explained 61.592% of the variance. The following two factors had eigenvalues of 0.645 and 0.611 (16.127% and 15.286% of variance respectively) and the last had an eigenvalue of .280, explaining the remaining 6.994% of variance. Three factors were extracted and rotated obliquely.

Table 4 shows the post-rotation factor loadings for the three extracted factors and their intercorrelations. Loadings suggest that a single factor explains the majority of the variance of the CRTs, while a second, bipolar factor weakly separates the two modalities. A third factor seems to explain the common variance between the two SRTs, but is highly correlated with factor one.

	Factors		
	1	2	3
AS SDRT	0.081	0.259	0.479
VS SDRT	0.15	-0.078	0.655
AC SDRT	0.859	0.169	0.002
VC SDRT	0.808	-0.149	0.038
Factor 1	1	0.181	0.816
Factor 2	0.181	1	0.248
Factor 3	0.816	0.248	1

Table 4. Pattern matrix and factor correlation matrix for principal axis factoring of reaction time standard deviations.

Discussion

The present study set out to asses to what extent individual differences in ISV were stable across tasks and sensory modalities. To that end we employed a primarily correlative and factor analytic approach, employing ANOVAs to ensure that our tasks were somewhat heterogeneous. The ANOVAs supported our task selection, with significant effects of task and modality on SDRTs suggesting that our tasks were not excessively similar. Finally the correlations and factor analyses suggested that, using our selection of tasks and modalities at least, ISV is a unitary construct with 61% of variance accounted for by a single factor. This has important theoretical implications, and also suggests that data reduction may be an appropriate technique in future ISV research.

In addition to the large first factor, we found two small factors that seemed to represent modality and task-specific variance respectively. Despite our use of an oblique rotation, the modality factor appeared to be orthogonal to the two other factors; suggesting that the main source of ISV appears to be relatively modality-independent. This is consistent with a unitary and global model of ISV and suggests that the modality-specific variance has a separate source from 'factor 1' ISV.

According to our pattern matrix, the primary factor in our data loaded onto the CRTs, while factor three loaded on the SRTs. This could be interpreted as evidence for a relationship between task complexity, akin to the relationship between *g* loading and ISV found by Kranzler (1992).Unlike factor two, however, factor three was highly correlated with factor one. This, in addition to the small amount of variance it explains, makes it likely that factor three represents a case of overextraction, rather than a meaningful source of variance.

We argue that the factorial structure identified by this study suggests that the main contribution to ISV is largely independent of modality and task. ISV is thus a relatively unitary

and global construct. Furthermore, this suggests that studies aiming to measure ISV should consider employing a latent variable approach in order to tap this modality and task-independent source of variance. By doing this, the key source of ISV variance can be disentangled from variance associated with specific tasks and methods, yielding theoretically and psychometrically 'cleaner' dependent variables.

Although our results are somewhat different to two previous studies using principal components analysis on ISV, this is easily accounted for. Jensen (1990) conducted a principle components analysis in order to assess the independence of MnRT from SDRT, and the results are thus not comparable to our own. Klein *et al.* (2006) found a single major factor in ADHD patients, but two factors in controls. However, similarly to our study, the larger factor loaded on the tasks with two response alternatives while the smaller factor explained the tasks where participants either responded or inhibited their response, dependent on the stimuli. Their study only presented tasks in the visual domain and so there was no cross-modal variance to be explained. As the tasks used were also quite different, being generally more complex, to those in the present study, the samples were quite dissimilar, and the statistical approaches were different, differences between the results of these studies are easily explained.

This study was not without limitations. ISV has become an important topic in a number of clinical areas and, as our data were collected from healthy university students, it is not clear how well they will generalise to the various clinical and neuropsychological populations where ISV is thought to be relevant. Klein *et al.* (2006) raise the possibility that there may be differences in the factorial structure of ISV between certain populations, but further work is needed to assess this.

Also on the subject of generalisability, it is unclear whether ISV has the same structure across all cognitive domains. We examined ISV on a group of simple tasks, but future research could employ a similar methodology and assess whether variability in performance is unitary across more complex tasks. It would also be interesting to examine whether similar results would

be obtained using a more heterogeneous battery of tasks. A larger battery of tasks would also provide superior overdetermination for the factor analysis, which would improve the accuracy of factor extraction.

On a more technical front, factor analysis calls for a large samples, and our sample of 71 was smaller than many factor analytic designs. McCallum, Widaman, Zhang, and Hong (1999) found that with high communalities, sample sizes as low as 60 could achieve highly accurate factor recovery; it was cases of low communality where sample size and overdetermination became important. Our communalities were high for CRTs but somewhat mediocre for SRTs. However, our factors had face validity, lending further support to our approach.

To conclude, ISV appears to be fairly unitary with respect to the comparison of tasks in the visual and auditory modalities, and only minor modality-specific sources of variance that could not be explained by the first factor were found. There is some evidence that correlated but separate factors underlie SRTs and CRTs, but it is unclear how meaningful this distinction may be. Future studies could build upon this work by investigating the structure of ISV across a larger and more heterogeneous battery of tasks, tapping a variety of cognitive, neuropsychological, and psychomotor domains. This would help to assess to what extent ISV could be considered a global property of the central nervous system. Future work might also consider employing factor-analytic data reduction in order to measure the latent variable underpinning ISV, rather than using individual tasks. This would allow the investigation of ISV at a supra-modal and supra-task latent variable level, with associated psychometric advantages.

Chapter Nine

A Supra-Task Latent Variable Approach to Intra-Subject Variability and Electroencephalographic Data

Abstract

Individual differences in intra-subject variability in reaction times (ISV) have been attributed to global and unitary biological bases, such as catecholaminergic function, white matter volume, and the activity of the default mode. Such a conceptualisation of ISV lends itself to a latent variable modelling approach to measurement, but this has so far only been used with behavioural data. In the present study, participants carried out two oddball tasks, each repeated in two modalities, while having an EEG recording. Supra-task latent variables were derived for ISV, as well as several single-trial P3b parameters, and the true score correlation coefficients between these constructs were estimated. These data support previous work, showing strong connections between P3b latency and reaction times, and suggest that a latent variable approach to ISV may not just be useful with behavioural data, but also psychophysiological data.

Intra-subject variability in reaction times (ISV) has become an increasingly important topic in a number of areas of neuroscience. The trait has been identified as a possible marker for ADHD (Klein, Wendling, Huettner, Ruder, & Peper, 2006), brain trauma (Stuss, Murphy, Binns, & Alexander, 2003), and cognitive aging (Lövdén, Li, Shing, & Lindenberger, 2007).

A number of models have been posited regarding ISV's neural underpinnings, and can be broadly speaking categorised as neural noise, oscillatory, and lapse models (see Chapter 4). These models differ in what they assume ISV represents, but all of these models point towards a unitary and global source for individual differences in ISV. ISV is thus seen as a global property of the central nervous system, rather than being local to discrete networks such as sensory systems or cognitive domains.

The conception of individual differences in ISV as being caused by global and unitary properties of the central nervous system fits with our own data. Data presented in the previous chapter suggest that a single factor, derived using principle axis factoring, can account for most of the variance in ISV over two tasks each carried out in two sensory modalities. The high level of shared variance across tasks and modalities is consistent with this global and unitary model of ISV.

If one takes the view of ISV as a global trait which manifests, albeit with some task-specific variance, in any RT task, then the subject lends itself well to a latent variable approach (see Loehlin, 1987). ISV, in this conceptualisation is not measured directly but is instead a latent variable which must be estimated from observed variables. By measuring the common variance in ISV across a number of tasks, the underlying global trait can be more accurately measured than by using any single task. This approach has its roots in the 'hotchpotch principle' endorsed by Spearman (1904) for the study of psychometric intelligence. The advantage here is that it is possible to separate the common variance, thought to be due to the latent variable, from task-

specific variance including measurement error, which is not, and so ISV can be measured more accurately (Loehlin, 1987).

This approach to ISV has been employed before on behavioural data (*e.g.* Ram, Rabbitt, Stollery, & Nesselroade, 2005) but not to explore the relationship between ISV and EEG-derived variables that were described in Chapter 7 of this volume. EEG parameters, however, are equally vulnerable to sources of measurement error, indeed the sources that determine EEG can be well described as latent variables that cause the observed variables on the scalp. The present study attempts to explore the link between ISV and certain parameters of the P3b, by modelling both ISV and single-trial P3b variables as latent variables, and determining canonical correlations between them.

Methodology

Ethical approval was obtained from the School of Psychology ethics and research governance committee at Bangor University.

Participants

Participants were 59 student volunteers from Bangor University. Participants were not tested if they reported any psychiatric or neurological conditions, and their data were excluded if fewer than 20 clean trials existed on any task (N=1). Data from 58 participants (32 female, 3 left-handed, mean age = 21.98, SD = 2.84) were used in the final analysis.

All participants gave written informed consent to participate after being fully briefed on the study.

Apparatus and Materials

Stimuli were presented on a 17'' monitor (with a copper-shielded power unit), connected to a stimulation PC running E-Prime V. 1.2 (Psychology Software Tools, USA). EEG activity was recorded with a sampling rate of 1000Hz and a low pass filter of 250Hz, using 63 Ag/AgCl electrodes (Falk Minow, Munich) of the international 10-10 system (American Electroencephalographic Society, 1991) attached over both hemispheres, with Cz as the recording reference, AFz as ground, as well as two infra-orbital electrodes for a total of 65 electrodes. Impendence was reduced to < 5kΩ and electrodes were connected to two BrainAmp DC amplifiers (0.1μv resolution, Brain Products, Germany), which were connected to a recording PC running Brain Vision Recorder (Brain Products, Germany). Data were collected inside a sound-attenuated Faraday cage.

Stimuli and Procedure

Participants performed four different oddball tasks: Visual and auditory versions of a two-stimulus oddball task, and visual and auditory versions of a repeating oddball task. Stimuli for the visual tasks were circles of ~2° visual angle in light and dark shades of blue (Light blue: L=80, A=-60, B=-60; Dark blue: L=80, A=8, B=-60), while stimuli for the auditory tasks were high and low sinusoidal tones (740Hz and 440Hz respectively).

In the two-stimulus oddball tasks, participants attended to a series of stimuli and were asked to respond to one type of stimuli (*e.g.* dark blue circles) with a key press using one hand and stimuli of the other type (*e.g.* light blue circles) with a key press using the other. 80% of

stimuli, hereafter called standards, were of one type while the remaining 20% of trials, hereafter called oddballs, were of the other.

In the repeating oddball task, participants attended to a series of stimuli which on 80% of trials presented the stimulus that was not seen in the preceding trial. On the remaining 20% of trials however, the stimulus from the previous trial *was* presented. Participants were asked to respond to the two trial types, repeats and non-repeats, with a key press using a different hand.

In both the two-stimulus oddball tasks, the colours and tones used as oddballs and standards were counterbalanced across participants, and for all tasks the response hands used for the two classes of response were counterbalanced across participants. Each task consisted of 500 trials (400 standards and 100 oddballs) with a jittered stimulus-onset asynchronicity, averaging 2000ms but ranging between 1250ms and 1750ms. Participants were asked to respond as quickly as possible to all stimuli.

Data Analysis

EEG analysis was primarily carried out using Brain Vision Analyzer (Version 2.0, Brain Products, Germany), but Dien's ERP-PCA toolkit (Dien, 2009) was used for the principal components analysis. The data analysis steps described below are summarised in Table 1 below.

Data for each task were average referenced and 0.1-50.0Hz filtered with symmetrical 24dB per octave slopes. Data contaminated by direct current offset corrections made during recording were rejected, before an Infomax independent component analysis, trained on 200s of data starting from 100s into each dataset, was applied to the data. Weightings for components representing eye activity, heartbeats, muscle activity, movement artefacts, or bad channels were set to zero before back-projection (Jung *et al.*, 2000).

	Process	Details
1	Compute average ERPs	Raw data preprocessed and standard average ERPs derived in order to characterise topography of P3b.
2	Infomax-PCA	PCA with Infomax rotation applied to concatenated averages of all tasks, to identify electrode weightings for P3b factor.
3	Single-trial analysis	Single-trial P3bs identified for P3b factor, and medians and SDs of amp. & lat. calculated for each participant on each task.
4	Create supra-task latent variables.	Principal axis factoring used to estimate supra-task latent variables for median and SDs of P3b lat., P3b amp, and RT.
5	Estimate latent variable correlations	The six supra-task factors were correlated with each other.

Table 1. Summary of analysis steps.

Any data where amplitude still ranged by more than 100μv, or less than 0.5μv, within 200ms were marked as bad. The data were then filtered 0.5-4.0 Hz and data for oddball trials were segmented into stimulus-locked epochs of 600ms pre-stimulus to 1400ms post-stimulus, with the period between -600 and -400ms in each epoch used as a baseline. Epochs containing any data marked as bad, those based around an incorrect response, or those where the response was faster than 120ms or slower than 1400ms were discarded.

Conventional average ERPs were first derived by averaging stimulus-locked oddball trials (Step 1 in Table 1). The data were then prepared for single-trial analysis in the manner described in Chapter Seven briefly summarised again below.

Stimulus-locked participant averages for all four tasks were concatenated[7] before Dien's ERP-PCA toolkit (Dien, 2009) was used to run spatial PCA on the data (Step 2 in Table 1). Previous studies have shown scalp topography of the P3b to be independent of modality (*e.g.* Katayama & Polich, 1999), so we would argue the use of a single topography for these tasks is reasonable. Based on a parallel Scree test (Horn, 1965), we extracted six factors and Infomax-rotated them. Figure 1 shows factor topographies, with the depth of colour representing the strength of the weighting of each electrode given by the factor pattern matrix. Factor 1 appears to have topography most similar to a standard P3b (midline parietal) and so all further analyses are based upon this factor.

We then applied the factor pattern matrix for Factor 1 to the single-trial data. P3b peaks were picked for each trial (Step 3 in Table 1), defined as the time-point with the maximum voltage between 250 and 750ms post-stimulus. Latency and amplitude information for each trial were recorded and used to create 'jitter-free' amplitude and latency measurements (median single-trial amplitude and latency), as well as measurements of amplitude and latency variability (intra-participant standard deviation of amplitude and latency).

While the data were now in factor space, rather than electrode space, they were still on the within-task level. In order to examine the variance shared across tasks, principal axis factoring was used to produce supra-task latent variables of the RT and P3b measures (Step 4 in Table 1).

---

[7] Due to memory constraints, data were down sampled to 250Hz prior to being filtered, segmented, averaged, and exported for analysis using PCA. Due to the low-pass filter we employed, this should not lead to an appreciable loss of information. The spatial weightings derived were applied to data with the original sampling rate of 1kHz.

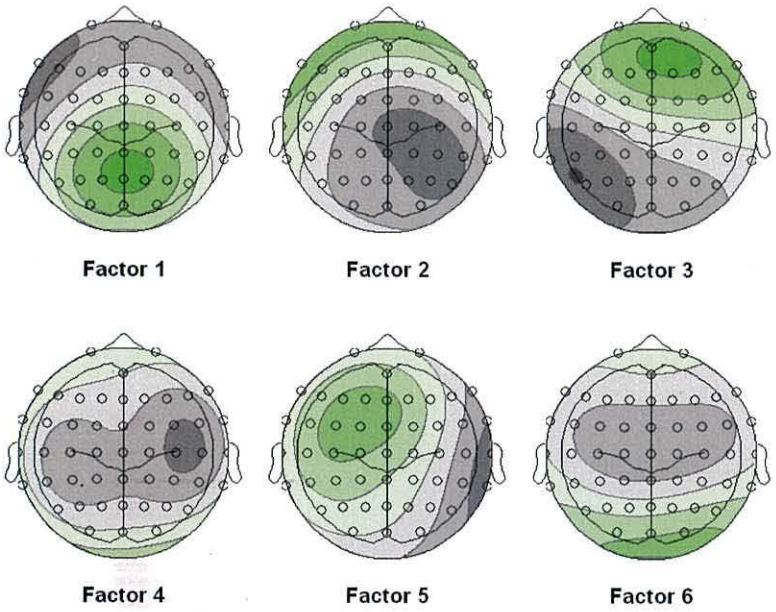

Fig. 1. Infomax-rotated topographies for PCA components (Lambda = 0.0001). Positive weightings are in green and negative weightings are in black.

Separate principal axis factorings were run on the following variables for each of the four tasks to reduce them down to a single factor: MnRT, SDRT, median P3b latency, standard deviation of P3b latency, median P3b amplitude, and standard deviation of P3b amplitude. Proportions of variance explained by each factor for each task are summarised in Table 2. Correlation coefficients were then computed between each supra-task variable (Step 5 in Table 1).

	Variance Explained				Factor Matrix for Factor 1			
	Factor 1	Factor 2	Factor 3	Factor 4	Aud. Rep.	Aud. 2S	Vis. Rep.	Vis 2S
Median Latency	61.656	16.389	13.934	8.020	0.751	0.639	0.732	0.675
SD Latency	52.643	18.721	16.633	11.983	0.535	0.555	0.800	0.542
Median Amplitude	77.137	9.067	8.286	5.510	0.865	0.868	0.806	0.797
SD Amplitude	75.583	11.382	9.565	3.470	0.732	0.997	0.793	0.766
MnRT	70.252	15.091	8.440	6.217	0.856	0.672	0.734	0.845
SDRT	61.456	19.159	13.234	6.151	0.563	0.643	0.796	0.785

Table 2. Variance explained by principal axis factorings conducted separately on ERP and RT variables, and factor matrices for first factor for each variable.

Correlation coefficients were also computed for the observed parameters on each task and these coefficients were averaged across all four tasks for comparison with the latent variable approach.

On the basis of our previous work, we hypothesised that SDRT would be positively correlated with median P3b latency, and P3b latency jitter; and negatively correlated with median P3b amplitude and P3b amplitude variability.

Results

For illustrative purposes, Figure 2 shows scalp topographies and waveforms from electrode Pz for all four tasks, while Figure 3 shows waveforms for all four tasks on Factor 1. Table 3 shows correlation coefficients between the extracted latent variables and observed variables (with the latter Z-transformed, averaged across all four tasks, and retransformed into Pearson's *r* scores). As hypothesised, significant relationships are evident between SDRT and latency jitter, and SDRT and median P3b amplitude. There are also strong relationships between the ERP parameters, most notably a strong relationship between median P3b amplitude and latency jitter, mirroring the equivalent relationship between MnRT and ISV found in RTs; and the highly significant negative relationship between median P3b amplitude and P3b amplitude variability. The observed correlations show a similar pattern to the estimated latent variable correlations, but the latter are uniformly stronger.

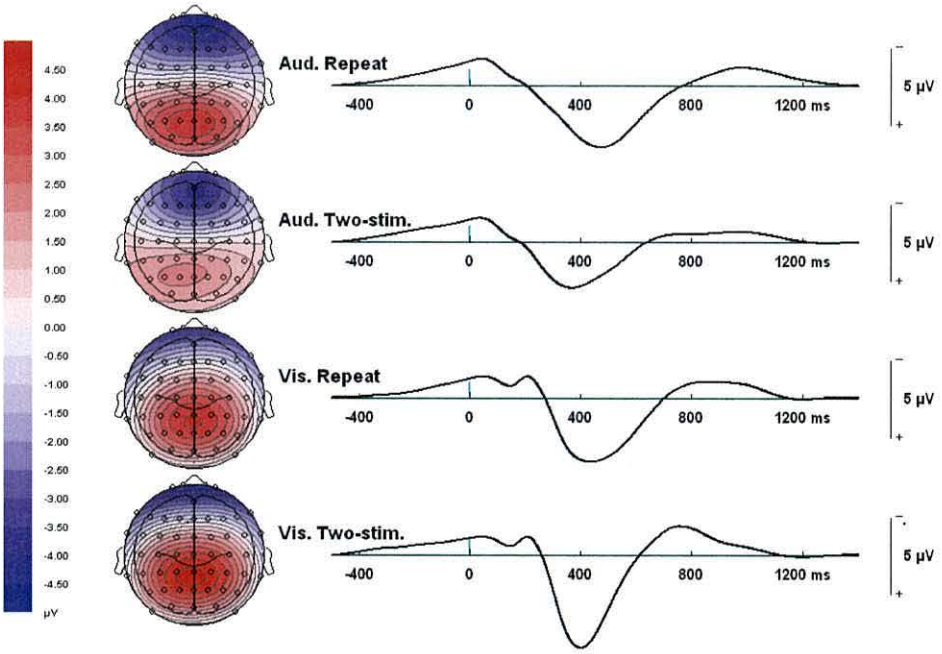

Fig. 2. Stimulus-locked average ERPs for Pz, topography maps show mean amplitude between 450 and 500ms post-stimulus (Lambda = 0.0001).

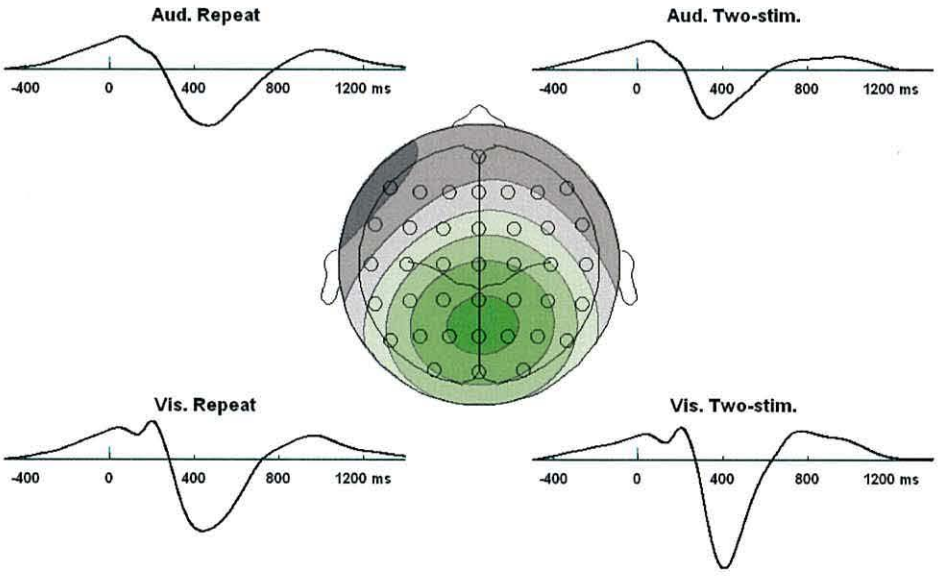

Fig. 3. Stimulus-locked average ERPs for Factor 1 of PCA. Factor 1 topography displayed in centre (Lambda = 0.0001).

		SD P3b Lat.	Mn P3b Amp.	SD P3b Amp.	MnRT Odd.	SDRT Odd.
Median P3b Latency	Latent	.621**	.292*	-.217	.727**	.629**
	Observed	.407*	.239*	-.117	.601**	.395**
SD P3b Latency	Latent		.435**	-.231	.646**	.667**
	Observed		.445**	-.115	.494**	.495**
Median P3b Amplitude	Latent			-.866**	.212	.231
	Observed			-.754**	.189	.183
SD P3b Amplitude	Latent				-.110	-.113
	Observed				-.049	-.032
MnRT Oddballs	Latent					.823**
	Observed					.695**
SDRT Oddballs	Latent					
	Observed					

Table 3. Correlation coefficients between supra-task latent variables, shown in the rows marked latent, and correlation coefficients for observed variables (Z-scaled, averaged across tasks, and retransformed into Pearson's *r* scores) shown in the rows marked observed. A single asterisk denoted significance at the .05 level, a double asterisk denoted significance at the .01 level.

Discussion

The present study applied a latent variable approach to behavioural measures of ISV and their electroencephalographic single-trial P3b correlates, in order to more accurately measure the relationship between these constructs. Our data replicate some previous findings, showing a strong positive relationship between RT and P3b latency, but fail to replicate others, finding no significant negative relationship between ISV and P3b amplitude (and indeed a trend in the opposite direction). A very strong negative relationship was also found between median P3b amplitude and P3b amplitude variability. The use of a latent variable approach was able to identify stronger inter-parameter relationships than raw intra-task correlations, suggesting that a latent variable framework may be a good way of approaching the study of ISV, both conceptually and methodologically.

The results showed a strong relationship between the latency of the P3b and RT (also see Chapter 7). This is consistent with a long tradition of work suggesting that the P3b represents a subset of the cognitive operations that underpin making a correct response (*e.g.* Kutas McCarthy, & Donchin, 1977; Gerson, Parra, & Sajda, 2005). Indeed, while this is not the primary thrust of

the present line of research, comparing participants with high and low ISV may be a useful approach for research into the functional significance of the P3b, and could perhaps help to settle the longstanding debate over its tactical and strategic roles.

While traditional models of the role of the P3b posit that it represents a strategic working memory-updating process, unrelated to response-selection, contradictory accounts have long argued for a tactical role for the P3b. Verleger (1988) argues that the P3b represents the closing of a perceptual epoch following the identification of an anticipated stimulus. Pritchard, Houlihan, and Robinson (1999) argue that while P3b onset may be a good index of the speed of stimulus processing, the P3b itself partially represents response-selection processes, which shows a temporal relationship with stimulus-evaluation due to the dependence of response-selection on stimulus-evaluation. Nieuwenhuis, Aston-Jones, and Cohen's (2005) account places the P3b somewhere between stimulus-evaluation and response-selection, arguing that the P3b represents the cortical projections of a noradrenergic decision making process. Single-trial methods represent a useful approach for disentangling these options, as they allow the relationship between RT and P3b latency to move beyond measures of central tendency to higher moments of their respective distributions, giving important evidence for the likely underlying dynamics.

The study did not replicate the inverse relationship between ISV and P3b median amplitude, described in Chapter Seven (also see Segalowitz, Dywan, and Unsal, 1997; and Di Russo and Spinelli, 2010). As the tasks employed were comparable and the analysis approach was highly similar, this suggests that the conflicting findings may be due to differences in the design of the study. Where Study Two compared groups from the opposite ends of the ISV continuum, the present study employed a correlational approach to the variables. This may imply that the relationship between ISV and P3b amplitude is non-linear, with a fall in P3b amplitude only at high levels of ISV. Conversely, the effect found in Chapter Seven was only of moderate

size, and the presence of a trend in the opposite direction in the present study may suggest that this is simply not a strong or stable effect.

Median P3b amplitude and variability in P3b amplitude show a large unhypothesised negative relationship. This is perhaps analogous to the relationship between MnRT and ISV, but with a different direction of skew. Since this relationship was unhypothesised, it will require replication before speculations can be made as to its possible significance.

One detail of our method that warrants further discussion is the decision to derive one set of PCA weightings for all four tasks, rather than a set of weightings for each task individually. Like the decision to derive one set of components for all participants, rather than a set for each participant (see Chapter Seven and Chapter Ten), this amounts to a trade off between specificity and comparability. Separate sets of weightings for each task may better characterise the precise topography of the P3bs associated with each task, but will lead to ambiguities when comparing results across tasks. This decision was, however, not only about this trade off. While the topography of a P3b is thought not to vary radically across modalities (Katayama & Polich, 1999), subtle topographic differences may occur due to the specific demands of each task. As the aim of the present study is to explore the relationships between the common latent variables that underpin different tasks, it seemed more appropriate to choose a single set of weightings that characterise the 'supra-task' P3b, and by implication its neural generators.

The study supports the use of latent variable approaches in neuroscience. The use of supra-task latent variables to measure these constructs fits well with the conceptualisation of ISV as a global property of the brain and the larger correlation coefficients found using these latent variables, as compared to the raw correlation coefficients, could be taken as possible evidence for the appropriateness of such a conceptualisation. Indeed a single-factor solution fitting the major variables is itself an argument for such an approach. Such latent variable approaches are not new in ISV research (*e.g.* Schmiedek *et al.,* 2007), but remain rare in neuroimaging and

psychophysiological studies, probably due to the relatively large sample sizes necessary to conduct these analyses. Such approaches may represent a useful alternative to the ANOVA-based approaches more common in these areas of research.

To conclude, by moving from the observed variable domain, to the supra-task, latent variable domain, the relationship between RT and ERP parameters can be more accurately evaluated. This not only represents a potentially fruitful methodological approach for the study of the neural bases of ISV, but also supports the idea of ISV as a pervasive supra-modal trait.

Chapter Ten

Overall Discussion.

The present thesis describes a series of studies into the subject of intra-subject variability of reaction times and its electroencephalographic correlates. This final chapter will first summarise the main findings of each of the four empirical chapters, before reviewing and evaluating the various methodologies employed, and discussing some of the theoretical implications and future directions of this work.

Study One

A key initial concern for any empirical endeavour is establishing the reliability of measurement techniques. This is particularly pressing for the current programme of research as ISV was long viewed as meaningless error variance, in line with information theory (Shannon & Weaver 1949) and classical test theory (Novick, 1966). Furthermore, a great many approaches exist for quantifying ISV; it is not clear how they compare in terms of their reliability, and to what extent these measures are largely redundant with regard to one another. A third issue for ISV measurement is the level of data aggregation needed to obtain reliable measurements. Quantities of RTs that are adequate to produce reliable estimates of measures of central tendency may not be adequate for measures of ISV.

Study One thus compared the test-retest and odd-even reliabilities of a selection of ISV metrics, on a battery of widely used cognitive tasks. The intercorrelations, and thus redundancy, of these metrics were also assessed on the task with the greatest number of available trials - the go-no go task. Finally a Monte Carlo simulation was run to establish the relationship between number of trials used to compute a metric and its reliability for SDRT and MRT.

The results of this first study identified good reliability for a number of measures of ISV. This level of reliability appeared, however, to be highly contingent on the number of trials used

to compute these measures, with some of the tasks with the fewest trials showing inadequate levels of reliability. The measures of ISV that did show adequate reliability appeared to be highly intercorrelated, and thus arguably redundant to one another. This was not just the case for measures that were obviously mathematically similar, such as SDRT and MAD, but also for measures that theoretically measured different aspects of the RT distribution, such as SDRT and ex-Gaussian tau. The Monte Carlo simulation verified the apparent relationship between trial number and reliability, and also showed a much slower rise in reliability for SDRT as compared to MRT. This suggests that levels of data aggregation that have proved sufficient for measures of central tendency may yield unreliable estimates of ISV, a fact underappreciated in the literature.

Study Two

The second study addressed the neural correlates of ISV. A number of studies have found reduced P3b amplitude in populations with increased ISV stemming from brain injuries (Segalowitz, Dywan, & Unsal, 1997; Di Russo & Spinelli, 2009). It was unclear, however, whether ISV and P3b amplitude would be related in non-brain injured populations, or whether increased ISV and reduced P3b amplitude were unrelated consequences of brain injury. A second open question from these studies was the extent to which the apparent reduction in ISV was in fact an artefact of latency jitter.

This study recruited participants who had shown particularly high or low ISV in the first study to carry out a working memory oddball task while having an EEG recording. Principal components analysis was employed in order to denoise the data, allowing P3b amplitude and latency to be identified in single trials.

Increased latency jitter was indeed found in the high ISV group, but the jitter-free measurements of P3b amplitude also showed a significant between groups difference, with lower

amplitude in high ISV than low ISV participants. A reduced correlation was also found between RT and P3b latency in high ISV participants, suggesting that ISV may be adding delays to both stimulus evaluation and response selection processes, a speculation also consistent with the increased latency jitter in response-locked P3bs identified in high scorers.

Study Three

Most neural models of ISV assume a unitary and global cause to individual differences in ISV, an assumption which Study Three set out to test. If ISV is relatively global, or at least central, individual differences in ISV should be relatively consistent across sensory modalities and tasks. In contrast, the presence of separable ISV factors in different modalities and tasks would suggest that ISV is relatively domain-specific and plural.

In Study Three, participants carried out a simple RT task and a choice RT task in the visual and auditory domains. Principal axis factoring was then carried out on participants' SDRTs in order to assess the cross-modal and cross-task structure of individual differences in ISV.

The results of the principal axis factoring were consistent with a single factor model of ISV, with only very weak factors representing modality and task. Such results are consistent with a unitary model of ISV, suggesting that individual differences generally enter at the supra-task, supra-modality level. The results of this study also suggest that a latent variable approach may be appropriate for the study of ISV.

Study Four

The fourth and final study of the present thesis drew on the results of the prior three studies to synthesise psychometric and EEG methods in the study of ISV. Modelling both ISV and the P3b parameters as latent variables, a relatively large sample of 58 participants were tested

on a battery of oddball RT tasks while having an EEG recording, and the single-trial approach developed for Study Two was applied to the data. Several RT and P3b parameters were computed for each task and the observed variable for each parameter on the four tasks were separately principal axis factored, giving a supra-task latent variable for each parameter. Correlation coefficients were then computed between all of these latent variables in order to estimate their true score correlation coefficients. These latent variable correlation coefficients were then compared to coefficients computed using the observed variables, in order to compare the two approaches.

Latent variable correlation coefficients found strong relationships between the parameters, replicating the results of Study Two. A strong relationship was found between SDRT and P3b latency variability, and between MnRT and median P3b latency. Stronger correlation coefficients were found between latent variables than between observed variables, demonstrating the advantage of the technique. This latent variable approach to both RT and neurophysiological data represents a promising option for future work into ISV.

Evaluation of methodologies

The present thesis employed a variety of methodological approaches over the four studies. In this section, the advantages and limitations of the approaches will be discussed, and potential refinements will be explored. The first set of methods to be discussed is the principle components-denoised single-trial analysis, employed in Studies Two and Four. The latent variable approach used in Studies Three and Four will then be evaluated.

Principal components-denoised single-trial analysis

The use of averaged ERPs can lead to misleading results when a significant level of latency jitter exists in evoked potential. This can be particularly misleading when an experiment compares participants with different levels of latency jitter in their evoked responses. As there was strong reason to expect latency jitter to covary with ISV, Studies Two and Four were analysed on a single trial basis, using an Infomax-rotated PCA pattern matrix to reduce the signal to a single channel, and to improve the signal-to-noise ratio. This approach yielded interesting results, but it is worth scrutinising some of the decisions made when designing the analysis protocol.

While the approach used employed the Infomax rotation algorithm, which has proven highly successful in EEG research, the manner of implemention was somewhat different to the way it is conventionally applied. The Infomax algorithm is essentially no different to any other PCA rotation procedure. Where it is unusual is that it is generally applied without the initial step of a PCA. The Infomax is instead usually run on *sphered* data, where higher order correlations have been removed from the data in order to speed up and stabilise convergence but the data is left at the sensor level.

When the Infomax is carried out on the sensor level, the assumption of stationarity implies that a separate ICA must be run for each participant. Any attempt to compare across participants involves identifying comparable factors[8] for each participant, either by 'eyeballing' the data or by use of some form of statistical clustering approach. The Infomax is also prone to making fine distinctions between factors which may in fact represent the same underlying construct, in contrast to more traditional PCA rotations which tend to erroneously combine

---

[8] Again the term *factor* will be used instead to refer to the statistical constructs derived by PCA, rather than the traditional term *component*. This is to avoid confusion with ERP components.

distinct sources of variance. This means that, without prior data reduction, some participants are likely to have multiple factors representing the same construct while others have one, or indeed none. This can make comparisons highly ambiguous.

Furthermore, the sheer volume of factors yielded by a separate ICA of each participant raises practical difficulties with making factor-selection decisions transparent to a reader. With more than a few participants, it is simply unfeasible to present all factors in an academic journal article. Even if this were possible, it is doubtful if such decisions would receive the necessary scrutiny.

Our approach instead applied a PCA as a preprocessing step, before applying the Infomax as a rotation procedure to data on the factor level for all participants simultaneously, an approach described by Dien, Khoe, and Mangun (2007). The substantial reduction in data accomplished by this initial PCA, as well as the application of the Infomax to factor-level data give this approach substantial advantages, and some disadvantages, over the approach described above.

Firstly moving the data from the sensor domain to the factor domain allows the Infomax rotation to be carried out on all participants at once. This has obvious advantages over having to identify comparable factors in each participant's data as one and the same set of weightings can be used for each participant. The small number of factors yielded is also compatible with the space limitations of academic journals and allows readers to easily evaluate the writer's choice of factors. Most importantly of course, the factors selected will be, by definition, comparable across participants, allowing generalities to be drawn without the risk of 'comparing apples and oranges'.

Secondly the data reduction carried out by the PCA makes oversplitting of factors, a common problem with the Infomax algorithm, much less likely. This also facilitates the application of the rotation to the whole sample at once, without an initial PCA step, it is likely that separate versions of factors would be identified for each participant, invalidating the whole premise of group ICA.

There are of course drawbacks to the PCA approach. Firstly there are advantages to creating bespoke filters for each participant – the filters may be a better fit for the peculiarities of their data, and there may be qualitative differences between scalp topographies in different groups. It is crucial to remember that the end product of the procedure is a set of spatial weightings, so if participants have different topographies there may be no single set of weightings which effectively characterises all participants' topographies. Additionally data reduction is something of a double edged sword; large highly distinct components, such as the P3b studied here, are unlikely to be lost in data reduction but the same may not be the case for more subtle components.

Finally it is important to reiterate that, although they were later applied to single-trial data, the PCA approach derived its spatial weightings from averaged ERPs. This means that EEG phenomena that were poorly time or phase-locked to stimulus-onset would be poorly characterised by this approach. Again, this is unlikely to be a problem for the P3b, but may be for other components. It may be possible to modify this approach to allow factors to be derived from single-trial data, although computer memory issues may need to be overcome here.

In short, one's choice of factor decomposition approach should depend on the details of the study in question. Running a more traditional ICA approach will yield as many factors as there are channels and thus give a wide view of the components present in the data. This makes such an approach useful for data-driven exploratory analysis. Also, running a separate ICA on each participant can be more appropriate in settings where a single bespoke set of weightings for each participant may lead to better results, certain brain-computer interface applications for example.

The PCA approach, in contrast, has advantages when there is a clear component of interest. Here it is easy to identify a single factor, representing a component such as the P3b, across a number of participants, and to relatively unambiguously compare participants on this factor. [9]

Besides factor derivation, there is also the issue of peak detection approach to consider. This is an important issue in single-trial analysis, as the poor signal-to-noise ratio of single trials leads to ambiguities with peak identification. The approach used here once the desired factor had been obtained was relatively simple – stringent filtering followed by maximum amplitude detection in a time window where the P3b was likely to manifest.

Such an approach is not the most sophisticated one available – curve fitting and signal-noise discrimination approaches exist (see Jaśkowski & Verleger, 2000) – but, given the high amplitude of the P3b and the highly stringent low-pass filter employed, such an approach appeared sufficient. Visual inspection of the identified peaks suggested that the obvious peak was picked in the vast majority of trials, and it is not clear that a more complicated approach would have performed better here. Given this, it seemed reasonable to employ an approach that most closely mimicked the peak detection approaches common in the average ERP literature. That said, in order to employ this approach on lower amplitude ERP components, which may not be as easy to isolate with low-pass filtering, it would be interesting to experiment with different peak detection approaches.

The choice of time window is also an important choice. Most of what is known about the time course of the P3b comes from the average ERP literature, but single-trial P3bs may show

---

[9] It is worth noting that nothing about employing a PCA as a preprocessing step prevents the running of separate Infomax-PCAs to create separate filters for different groups or participants. This of course does not take advantage of being able to run Infomax rotations on group level data, but there are certain instances where this may still be desirable, such as where factors are likely to be oversplit.

more latency variability than standard approaches suggest. Setting a suitable time window is a trade-off: too short and there is the risk of missing early or late P3bs, too wide and one both stretches the reader's credibility as to what can be reasonably called a P3b and risks introducing spurious peaks into the analysis. The time window (250-750ms post-stimulus) was intended as a happy medium between too long and too short, but future work may want to more closely examine the time ranges in which P3bs may occur in single-trials.

Latent variable approach to ISV

The other main methodological approach taken in this thesis was the use of a latent variable approach to measuring ISV. Latent variable methods have become a highly successful technique in modern behavioural science, allowing estimation of constructs that are difficult to measure directly, but can be measured indirectly through several observed variables (Loehlin, 1987). Such an approach fits well with a conceptualisation of ISV as a global characteristic of the nervous system. ISV is present in every task which records RTs, but each different task will contain task-specific variance and measurement error. Latent variable modelling can extract the communal variance across a number of tasks, and calculate supra-task estimates of ISV, which are less subject to measurement error.

Such an approach has great potential in the study of ISV, but it is important to address some of the caveats. Firstly, one must note that while latent variable methods may be able to identify similar patterns in ISV (*i.e.* unity, modality-specificity *etc.*), it cannot show that such ISV has the same neural underpinnings in all participants. For example, while one participant's high ISV may reflect poor axonal myelination, another participant's might reflect catecholaminergic dysregulation. If these neural bases led to similar profiles of task performance, factor analysis would group these as a single factor.

Secondly, such methods tend to require large samples. This can prove problematic if the populations of interest are rare or otherwise difficult to recruit, such as participants with a certain psychiatric or neurological disorder. This can also be difficult to orchestrate for psychophysiological or neuroimaging studies, where data can be time-consuming and costly to collect.

Thirdly, while latent variable approaches can remove measurement errors from estimates of latent variables, high levels of measurement error have a negative impact on the communality – a measure of shared variance. McCallum, Widaman, Zhang, and Hong (1999) have shown that one of the most important determining factors of the sample size needed to accurately recover latent variables is the level of communality. Thus in order to avoid needing a, possibly prohibitively, large number of participants, accurate and reliable measurements of ISV are important. This goes back to comments made in the discussion of Study One that, in light of the high intercorrelation of most metrics of ISV, reliability may be a good criterion to choose metrics by.

Theoretical implications

This thesis has a number of theoretical implications. These can be grouped into trait-theoretic and electrophysiological implications.

Trait-theoretic implications

The psychometric work in this thesis develops the concept of ISV as a trait. Study One showed that ISV shows good levels of stability of individual differences, at least in the short term. Study Three showed that in addition to temporal stability, ISV showed a good level of cross-task and cross-modal consistency. Study Four showed that this supra-task construct was closely

related to supra-task electroencephalographic parameters. While this latent variable approach is clearly a methodological decision, it is also implicitly a statement of a certain theoretical conceptualisation of ISV. Modelling ISV as a single latent variable implies that ISV is a pervasive, global, and stable property of a person, which manifests in all RT tasks, and perhaps in all situations where attention is sustained. By this view, the specific task may be viewed as being of secondary importance to the global trait. This is a bold claim, and one which will require stringent verification across a number of cognitive domains. Future work should investigate to what extent individual differences in variability parameters are consistent across disparate cognitive domains. If this context-free view of ISV appears to be vindicated, latent variable approaches will represent a powerful approach to future work on ISV.

Neural implications

The other main strand of this thesis was the study of the electroencephalograpic correlates of ISV, specifically the P3b component of the ERP. The results of Studies Two and Four suggest a strong relationship between RT and P3b latency. Study Two showed a modest correlation between single-trial RTs and P3b latencies in participants with low ISV, which was substantially attenuated in participants with high ISV. The more highly aggregated and less measurement error prone latent variables found in Study Four found strikingly high correlation coefficients between RT and P3b latency parameters.

The P3b is one of the most widely studied ERP components, with a number of models claiming to explain what it represents. Most of these models, however, are primarily cognitive, rather than neurophysiological, such as Donchin's (1981) model of the P3b representing "context updating" to unexpected events and Verleger's (1988) explanation of the P3b as the "closing" of a perceptual epoch after an expected event. Other frameworks, such as Johnson's (1986) triarchic

model, have focussed more on identifying factors that modulate P3b amplitude than what it represents *per se*. Physiological models of the P3b have largely been confined to lesion studies and attempts to localise its neural generators. A recent model by Nieuwenhuis, Aston-Jones, and Cohen (2005) has made a strong case that the P3b may represent a downstream cortical manifestation of phasic firing of noradrenergic neurons in the locus coeruleus. This hypothesis has strong support from invasive electrophysiology research in primates, which has shown that these neurons respond to stimuli in much the same way as a P3b (*e.g.* Usher, Cohen, Servan-Schreiber, Rajkowski, & Aston-Jones, 1999). Nieuwenhuis *et al.* argue that this activity represents the response of the locus coeruleus to a neural decision making process. This view has parallels with Carpenter's LATER model (1981), and suggests that future attempts to model P3b latencies in the same ways that RTs are beginning to be modelled (such as with the LATER model and the ex-Gaussian distribution) may prove illuminating.

To return to the present data, it appears that participants with high ISV show a weaker relationship between P3b latency and RT. The fact that higher latency jitter is seen relative to both stimulus onset and response, points to increased variability in both stimulus-evaluation and response-selection. This in turn points to a general increase in noise, rather than noise added to a specific neurocognitive process, an explanation which fits well with the psychometric results of this thesis.

The aetiology of increased ISV

While the work in this thesis, linking ISV to the P3b, is compatible with a noradrenergic model of ISV, it is important not to consider these results in isolation. The literature reviewed in Chapter Four details a number of possible aetiological factors, including catecholaminergic dysregulation (Stefanis *et al.*, 2005), white matter abnormalities (Walhovd & Fjell, 2007), and

structural damage to the prefrontal cortex (Stuss, Murphy, Binns, & Alexander, 2003). Without a theoretical account for why such characteristics should co-occur, it seems reasonable to posit that a number of biological factors can lead to high ISV. It may accordingly make sense to think of ISV as being an index of neurocomputational noise, rather than of a specific biological property *per se*. By this functionalist account, ISV represents noisy neural processing, which can result from deficient catecholaminergic neuromodulation, decreased myelination, other factors, or indeed combinations of factors. Thus participants with uniformly high ISV may show considerable aetiological heterogeneity.

Where would such an explanation leave ISV research? Is it not contradictory to argue that ISV is a unitary construct, before hypothesising that it represents a non-specific noise parameter in the brain? The argument made by this thesis is that there is no contradiction. ISV may represent a helpful measure of signal-to-noise ratio in neural processing, independent of its aetiology from person to person. While participants with high ISV may vary on a neural scale, on the system-wide scale, they may be largely indistinguishable. ISV, and neural noise by extension, may thus represent a functional property rather a biological property.

Alternatively, perhaps different causes do lead to subtle behavioural differences. While standard measures such as SDRT, used in the majority of ISV research, may be similarly influenced by different biological sources of noise, perhaps other metrics will find differences. Time series measures may find periodicities at different frequencies for ISV stemming from white matter abnormalities than catecholaminergic deficits. Emerging non-linear analysis methods for understanding neural data (*e.g.* Kelly, Heathcote, Heath, & Longstaff, 2001) may find differences in fractality or in the 'colour' of noise spectra of RT data. Perhaps simply the extent to which ISV is increased will show aetiological differences, with certain causal factors being associated with more substantial increases in ISV (see Geurts *et al.*, 2008). Future investigations should employ multimodal methods, in neuroimaging and psychopharmacology to

identify different aetiologies, and in quantification of ISV in order to identify behavioural differences between different aetiologies.

Conclusion

To sum up, the study of ISV seems at first glance rather esoteric, but addresses many of the pressing issues of modern neuroscience: the clash between traditional fronto-executive model of focussed attention and modern accounts of dynamic self-organisation; the computational role of neurotransmitters; and neurometric diagnosis of psychiatric disorders to mention but a few.

Perhaps as a result of the wide-ranging relevance of ISV, the literature has a tendency towards being somewhat parochial. Future investigations of ISV should attempt to explore the generalities between different manifestations of high ISV. Conversely, attempts should be made to falsify the hypothesis put forward in this thesis that ISV represents a general property, by comparing different aetiologies of ISV in a more fine-grained fashion. Such an approach could yield interesting results for computational neuroscience, psychiatric diagnostics, and the philosophy of psychology.

American Electroencephalographic Society (1991) American Electroencephalographic Society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology, 8,* 200-202.

Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994) Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *The Journal of Neuroscience, 14(7)*, 4467-4480.

Aston-Jones, G., & Cohen, J.D. (2005) Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *The Journal of Comparative Neurology, 493,* 99-110.

Aue, W.R., Arruda, J.E., Kass, S.J., & Stanny, C.J. (2009) Cyclic variations in sustained human performance. *Brain and Cognition, 71(3),* 336-344.

Baumeister, A.A., Kellas, G. (1968) Intrasubject response variability in relation to intelligence. *Journal of Abnormal Psychology, 73(5)*, 421-423.

Bell, A.J., & Sejnowski, T.J. (1995) An information maximising approach to blind separation and blind deconvolution. *Neural Computation, 7*, 1004-1034.

Bellgrove, M.A., Hester, R., & Garavan, H. (2004) The functional neuroanatomical correlates of response variability: evidence from a response inhibition task, *Neuropsychologia, 42*, 1910-1916.

Berger, H. (1929) Über das Elektroenkephalogramm des Menschen, *Archives of Psychiatry, 87,* 527–57.

Birkett, P., Sigmundsson, T., Sharma, T., Toulopoulou, T., Griffiths, T.D., Reveley, A., & Murray, R. (2007) Reaction time and sustained attention in schizophrenia and its genetic predisposition. *Schizophrenia Research, 95,* 76-85.

Brotman, M.A., Rooney, M.H., Skup, M., Pine, D.S., & Leibenluft, E. (2009) Increased intrasubject variability in response time in youths with bipolar disorder and at-risk family members. *Journal of the American Academy of Child and Adolescent Psychiatry, 48(6)*, 628-635.

Brown, W. (1910) Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3,* 296-322.

Bunce, D., Anstey, K.J., Christensen, H., Dear, K., Wen, W., & Sachdev, P. (2007) White matter hyperintensities and within-person variability in community-dwelling adults aged 60-64 years. *Neuropsychologia, 45*, 2009-2015.

Burton, C. L., Strauss, E, Hultsch, D. F., Moll, A., & Hunter, M. A. (2006) Intraindividual Variability as a Marker of Neurological Dysfunction: A Comparison of Alzheimer's Disease and Parkinson's Disease, Journal of Clinical and Experimental Neuropsychology, 28 (1), 67-83.

Buzy, W.M., Medoff, D.R., & Schweitzer, J.B. (2009) Intra-individual variability among children with ADHD on a working memory task: An ex-Gaussian approach. *Child Neuropsychology, 15(5),* 441-459.

Carpenter, R.H.S. (1981) Oculomotor Procrastination. In D.F. Fisher, R.A. Monty and J.W. Senders (Eds.), *Eye Movements: Cognition and Visual Perception*. Hillsdale: Lawrence Erlbaum, 237-246.

Carpenter, R.H.S. (1994) SPIC: a PC-based system for rapid measurement of saccadic responses. *Journal of Physiology (Proceedings)*; 480: 4P.

Castellanos, F.X., Sonuga-Barke, E.J.S., Scheres, A., Di Martino, A., Hyde, C., & Walters, J.R. (2005) Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. *Biological Psychiatry, 57,* 1416-1423.

Caton, R. (1875) The electric currents of the brain. *British Medical Journal, 2,* 1444.

Cattell, R. B. (1966) The scree test for the number of factors. *Multivariate Behavioural Research,* 1 (2), 245-276.

Chatfield, C. (1980) *The analysis of time series: An introduction.* London: Chapman and Hall.

Cleghorn, J.M., Kaplan, R.D., Szechtman, B., Szechtman, H., & Brown, G..M. (1990) Neuroleptic drug effects on cognitive function in schizophrenia. *Schizophrenia Research, 3*, 211-219.

Comerchero, M. D., & Polich, J. (1999) P3a and P3b from typical auditory and visual stimuli. *Journal of Clinical Neurophysiology, 110*, 24-30.

Cooley, J.W., Tukey, J.W. (1965) An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation, 19*, 297-301.

Cooper, R., Winter, A.L., Crow, H.J., & Walter, W.G. (1965) Comparison of subcortical, cortical, and scalp activity using chronically indwelling electrodes in man. *Electroencephalography and Clinical Neurophysiology, 38,* 191-196.

Cortina, J. M. & Nouri, H. (2000). *Effect size for ANOVA designs*. Thousand Oaks: Sage.

Coyle, T.R. (2001) IQ is related to the worst performance rule in a memory task involving children. *Intelligence, 29,* 117-129.

Coyle, T.R. (2003) A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence, 31*, 567-587.

Diascro, M. N., & Brody, N. (1993). Serial versus parallel processing in visual search tasks and IQ. *Personality and Individual Differences*, 14, 243-245.

Dien, J., Spencer, K.M., & Donchin, E. (2004) Parsing the late positive complex: Mental chronometry and the ERP components that inhabit the neighbourhood of the P300. *Psychophysiology, 41*, 665-678.

Dien, J., Khoe, W., & Mangun, G.R., (2007) Evaluation of PCA and ICA of simulated ERPs: Promax and infomax rotations. *Human Brian Mapping, 28*, 742-763.

Dien, J. (2009) The ERP PCA toolkit: An open source program for advanced statistical analysis of event-related potential data. *The Journal of Neuroscience Methods, 187*, 138-145.

Di Russo, F., & Spinelli, D. (2010) Sport is not always healthy: Executive brain dysfunction in professional boxers. *Psychophysiology, 47*, 425-434.

Donchin, E. (1981) Surprise!... Surprise? *Psychophysiology, 18,* 493-513.

Donders FC (1868): Twee werktuigen tot bepaling van den tijd, voor psychische processen benoodigd (Two instruments for determining the time required for mental processes), *Onderzoekigen gedaan in het Physiologisch Laboratorium Utrecht*, 2, 21-25.

Egner, T., & Gruzelier, J.H. (2004) EEG biofeedback of low beta band components: Frequency-specific effects on variables of attention and event-related brain potentials. *Clinical Neurophysiology,115,* 131-139.

Eichele, T., Calhoun, V.D., Moosmann, M., Specht, K., Jongsma, M.L.A., Quian Quiroga, R., Nordby, H., & Hugdahl, K. (2008) Unmixing concurrent EEG-fMRI with parallel

independent components analysis. *International Journal of Psychophysiology, 67 (3)*, 222-234.

Elul, R. (1972) The genesis of the EEG. *International Review of Neurobiology, 15,* 227-272.

Eysenck, H. J. (1982). Introduction. In Eysenck, H. J. (Ed.), *A model for intelligence.* Berlin: Springer

Falkenstein, M., Hohnsbein, J., & Hoormann, J. (1994) Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential. *Electroencephalography and Clinical Neurophysiology, 92*, 148-160.

Fiske, D.W., Rice, L. (1955) Intra-individual response variability. *Psychological Bulletin, 52(3)*, 217-250.

Flehmig, H.C., Steinborn, M., Langner, R., Scholz, A., & Westhoff, K. (2007) Assessing intraindividual variability in sustained attention: reliability, relation to speed and accuracy, and practice effects. *Psychology Science, 49 (2)*, 132-149.

Fourier, J. J-B. (1822) *Theorie analytique de la chaleur.* Essones: Firmin Didot.

Fuentes, K., Hunter, M.A., Strauss, E., & Hultsch, D.F. (2001) Intraindividual variability in cognitive performance in persons with chronic fatigue syndrome. *The Clinical Neuropsychologist, 15(2)*, 210-227.

Gaillard, A.W.K. Problems and paradigms in EEG research. *Biological Psychology, 26,* 91-109.

Galton, F. (1869/1881) *Hereditary genius: An inquiry into its laws and consequences*, New York: D. Appleton & Co.

Gerson, A., Parra, L., & Sadja, P. (2005) Cortical origins of response time variability during rapid discrimination of visual objects. *NeuroImage, 28,* 342-353.

Geurts, H.M., Grasman, R.P.P.P., Verté, S., Oosterlaan, J., Roeyers, H., van Kammen, S.M., & Sergeant, J.A. (2008) Intra-individual variability in ADHD, autistic spectrum disorders and Tourette's syndrome. *Neuropsychologia, 46,* 3030-3041.

Gilden, D.L. (1997) Fluctuations in the time required for elementary decisions. *Psychological Science, 8, (4)*, 296-301.

Gilden, D.L., & Hancock, H. (2007) Response variability in attention-deficit disorders. *Psychological Science, 18(9),* 796-802.

Guttman, L. (1954) Some necessary conditions for common-factor analysis. *Psychometrika, 19 (2)*, 149-161.

Heathcote, A., Brown, S., & Cousineau, D.. (2004) QMPE: Estimating Lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behavior Research Methods, Instruments, and Computers, 36(2),* 277-290.

Heiser, P., Frey, J., Smidt, J., Sommerlad, C., Wehmeier, P.M., Hebebrand, J., and Remschmidt, H. (2004) Objective measurement of hyperactivity, impulsivity, and inattention in children with hyperkinetic disorders before and after treatment with methylphenidate. *European Child & Adolescent Psychiatry, 13*, 100-104.

Hervey, A.S., Epstein, J.N., Curry, J.F., Tonev, S., Arnold, L.E., Conners, C.K., Hinshaw, S.P., Swanson, J.M., & Hechtman, L. (2006) Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychology, 12(2)*, 125-140.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6,* 65-70.

Horn, J.L. (1965) A rationale and test for the number of factors in factor analysis. *Psychometrica, 30 (2)*, 179-185.

Hultsch, D.F., MacDonald, S.W.S. & Dixon, R.A. (2002) Variability in reaction time performance of younger and older adults. *Journal of Gerontology: Psychological Sciences, 57B (2)*, 101-115.

Iqbal. (1930) *The reconstruction of religious thought in Islam*. Chicago: Kazi.

Jaśkowski, P., & Verleger, R. (2000) An evaluation of methods for single-trial estimation of P3 latency. *Psychophysiology, 37,* 153-162.

Jensen, A.R. (1982). Reaction time and psychometric *g*. In Eysenck, H. J. (Ed.), *A model for intelligence.* Berlin: Springer

Jensen, A.R. (1992) The importance of intraindividual variation in reaction time. *Personality and Individual Differences, 13 (8)*, 869-881.

Johnson, K.A., Kelly, S.P., Bellgrove, M.A., Barry, E., Cox, M., Gill, M., & Robertson, I.H. (2007) Response variability in Attention Deficit Hyperactivity Disorder: Evidence for neuropsychological heterogeneity. *Neuropsycholgia, 45(4)*, 630-638.

Johnson, K.A., Barry, E., Bellgrove, M.A., Cox, M., Kelly, S.P., Dáibhis, A., Daly, M., Keavey, M., Watchorn, A., Fitzgerald, M., McNicholas, F., Kirley, A., Robertson, I.H., Gill, M. (2008) Dissociation is response to methylphenidate on response variability in a group of medication naïve children with ADHD. *Neuropsychologia, 46*, 1536-1541.

Johnson, R. (1986) A triarchic model of P300 amplitude. *Psychophysiology, 23,* 367-384.

Jolliffe, I. T. (1986) *Principle Components Analysis*. New York. Springer-Verlag.

Jung, T-P., Makeig, S., Humphires, C., Lee, T-W., McKeown, M.J., Iragui, V., & Sejnowski, T.J. (2000) Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*, 163-178.

Kaiser, H. F. (1960) The application of electronic computers to factor analysis. *Educational and Psychological Measurement,* 20 (1), 141-151.

Katayama, J., & Polich, J. (1999) Auditory and visual P300 topography from a 3 stimulus paradigm. *Clinical Neurophysiology, 110,* 463-468.

Kelly, A., Heathcote, A., Heath, R., & Longstaff, M. (2001) Response-time dynamics: Evidence for linear and low-dimension nonlinear structure in human choice sequences. *The Quarterly Journal of Experimental Psychology, 54A(3)*, 805-840.

Kelly, A.M.C., Uddin, L.Q., Biswal, B.B., Castellanos, F.X., & Milham, M.P. (2008) Competition between functional brain networks mediates behavioral variability. *NeuroImage, 39,* 527-537.

Klein, C., Wendling, K., Huettner, P., Ruder, H., & Peper, M. (2006) Intra-subject variability in attention-deficit hyperactivity disorder. *Biological Psychiatry 60,* 1088-1097.

Koles, Z.J. (1998) Trends in EEG source localization. *Electroencephalography and Clinical Neurophysiology, 106,* 127-137

Kranzler, J. H. (1992) A test of Larson and Alderton's (1990) worst performance rule of reaction time variability. *Personality and Individual Differences,* 13 (3), 255-261.

Kuntsi, J., Stevenson, J., Oosterlan, J., & Sonuga-Barke, E.J.S. (2001) Test-retest reliability of a new delay aversion task and executive function measures. *British Journal of Developmental Psychology, 19*, 339-348.

Kutas, M., McCarthy, G., & Donchin, E. (1977) Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*, 792-795.

Larson, G.E. & Alderton, D.L. (1990) Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence, 14*, 309-325.

Lee, S-H., Song, D-H., Kim, B-N., Joung, Y.S., Ha, E.H., Cheon, K-A., Shin, Y-J., Yoo, H.J., Shin, D.W. (2009) Variability of response time as a predictor of methylphenidate treatment response in Korean children with attention deficit hyperactivity disorder. *Yonsei Medical Journal, 50(5),* 650-655.

Leth-Steensen, C., King-Elbaz, Z., & Douglas, V.I. (2000) mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychologica, 104,* 167-190.

Li, S-C., Lindenberger, U., & Frensch, P.A., (2000) Unifying cognitive aging: From neuromodulation to representation to cognition. *Neurocomputing, 32-33,* 879-890.

Linden, D.E.J. (2005) The P300: Where in the brain is it produced and what does it tell us? *Neuroscientist, 11,* 563-576.

Lövdén, M., Li, S-C., Shing, Y. L., & Lindenberger, U. (2007) Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. *Neuropsychologia, 45*, 2827-2838.

Loehlin, J.C. (1987) *Latent variable models: An introduction to factor, path, and structural analysis*. Hillsdale: Lawrence Erlbaum.

MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003) Performance variability is related to change in cognition: evidence from the Victoria Longitudinal Study. *Psychology and Aging*, 18 (3): 510-523.

MacDonald, S.W.S., Hultsch, D.F., & Dixon, R.A. (2008) Predicting impending death: Inconsistency in speed is a selective and early marker. *Psychology and Aging, 23 (3),* 595-607.

MacDonald, S. W. S., Nyberg, L., Sandblom, J., Fischer, H., & Bäckman, L. (2008) Increased response-time variability is associated with reduced inferior parietal activation during episodic recognition in aging. *Journal of Cognitive Neuroscience,* 20 (5): 779-786.

MacDonald, S.W.S., Cervenka, S., Farde, L., Nyberg, L., & Bäckman, L. (2009) Extrastriatal dopamine D2 receptor binding modulates intraindividual variability in episodic recognition and executive functioning. *Neuropsychologia, 47*, 2299-2304.

Makeig, S., Bell, A.J., Jung, T-P., Sejnowski, T.J. (1996) Independent components analysis of electroencephalographic data. In Touretzky, D.S., Mozer, M.C., Hasselmo, M.E. (Eds.) *Advances in neural information processing systems 8*. Cambridge, MA: MIT Press.

McCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999) Sample size in factor analysis. *Psychological Methods*, 4 (1), 84-99.

McIntosh, A.R., Kovacevic, N., Itier, R.J. (2008) Increased brain signal variability accompanies lower behavioural variability in development. *PLoS Computational Biology, 4(7)*: e1000106

Nash, A. J., & Fernandez, M. (1996) P300 and allocation of attention in dual-tasks. *International Journal of Psychophysiology, 23,* 171-180.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J.D. (2005) Decision making, the P3, and the locus-coeruleus norepinephrine system. *Psychological Bulletin, 131,* 510-532.

Novick, M.R., (1966) The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3,* 1-18.

Nunez, P. L., & Srinivasan, R. (1981/2006) *Electrical fields of the brain 2nd ed.* Oxford: Oxford University Press.

Nyquist, H. (1928) Certain topics in telegraph transmission theory. *Transactions of the AEEE, 47*, 617-644.

Parra, L.C., Spence, C.D., Gerson, A.D., & Sajda, P. (2005) Recipes for the linear analysis of EEG. *NeuroImage, 28(2),* 326-341.

Pham, D.T., Möcks, J., Köhler, W., & Gasser, T. (1987) Variable latencies of noisy signals: Estimation and testing in brain potential data. *Biometrika, 74,* 525-533.

Plosney, R. (1963) Reciprocity applied to volume conductors and the EEG. *IEEE Transactions on Biomedical Engineering, 10,* 9-12.

Posner, M., I. (1978) *Chronometric Explorations of Mind*, Hillsdale, N.J., Laurence-Erlbaum Associates.

Pouget, P., Wattiez, N., Rivaud-Péchoux, S., & Gaymard, B. (2009) A fragile balance : Perturbation of GABA mediated circuit in prefrontal cortex generates high intraindividual performance variability. *PLoS One, 4 (4)*, e5208.

Pritchard, W.S., Houlihan, M.E., & Robinson, J.H. (1999) P300 and response selection: A new look using independent-components analysis. *Brain Topography, 12 (1)*, 31-37.

Rabbitt, P., Osman, P., Moore, B., & Stollery, B. (2001) There are stable individual differences in performance variability, both from moment to moment and from day to day. *The Quarterly Journal of Experimental Psychology, 54A (4)*, 981-1003.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., & Shulman, G.L. (2001) A default mode of brain function, *Proceedings of the National Academy of Sciences of the United States of America, 98(2),* 676-682.

Ram, N., Rabbitt, P., Stollery, B., & Nesselroade, J.R. (2005) Cognitive performance inconsistency: Intraindividual change and variability. *Psychology and Aging, 20(4)*, 623-633.

Ratcliff, R., & Tuerlinckx, F. (2002) Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin Bulletin and Review,* 9 (3): 438-481.

Rentrop, M., Rodewald, K., Roth, A., Simon, J., Walther, S., Fiedler., Weisbrod, M., & Kaiser, S. (2010) Intra-individual variability in high-functioning patients with schizophrenia. *Psychiatry Research, 178*, 270-32.

Russell, V. A., Oades, R. D., Tannock, R., Killeen, P. R., Auerbach, J. G., Johansen, E. B., & Sagvolden, T. (2006) Response variability in attention-deficit/hyperactivity disorder: A neuronal and glial energetics hypothesis. *Behavioral and Brain Functions*, 2 (30).

Salthouse, T.A. (1998) Relation of successive percentiles of reaction time distributions to cognitive variables and adult age. *Intelligence, 26(2)*, 153-166.

Särelä, J., & Vigário, R. (2000) Overlearning in marginal distribution based ICA: Analysis and solutions. *Journal of Machine Learning Research, 1,* 1-48.

Schaul, N. (1998) The fundamental neural mechanisms of electroencephalography, *Electroencephalography and Clinical Neurophysiology, 106,* 101-107.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süβ, H-M., & Wittmann, W.W. (2007) Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136 (3)*, 414-429.

Schwartz, F., Carr, A.C., Munich, R.L., Glauber, S., Lesser, B., & Murray, J. (1989) Reaction time impairment in schizophrenia and affective illness: The role of attention. *Biological Psychiatry, 25(5)*, 540-548.

Segalowitz, S.J., Dywan, J., & Unsal, A. (1997) Attentional factors in response time variability after traumatic brain injury: an ERP study. *Journal of the International Neuropsychological Society, 3,* 95-107.

Shakow, D. (1977) *Schizophrenia: Selected papers*. New York: International Universities Press.

Shammi, P., Bosman, E., & Stuss, D. T. (1998) Aging and variability in performance. *Aging, Neuropsychology, and Cognition,* 5 (1), 1-13.

Shannon, C.E. (1949) Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers, 37(1)*, 10-21.

Shannon, C.E., & Weaver, W. (1949) *The mathematical theory of information.* Urbana: University of Illinois Press.

Singer, W. (1993) Synchronization of cortical activity and its putative role in information processing and learning. *Annual Reviews in Physiology, 55,* 349-374.

Shipley, B.A., Der, G., Taylot, M.D., & Deary, I.J. (2006) Cognition and all-cause mortality across the entire adult age range: Health and lifestyle survey. *Psychosomatic Medicine, 68,* 17-24.

Simmonds, D.J., Fotedar, S.G., Suskauer, S.J., Pekar, J.J., Denckla, M.B., & Mostofsky, S.H. (2007) Functional brain correlates of response time variability in children. *Neuropsychologia, 45,* 2147-2157.

Smulders, F.T.Y., Kenemans, J.L., & Kok, A. (1994) A comparison of different methods for estimating single-trial P300 latencies. *Electroencephalography and Clinical Neurophysiology, 92,* 107-114.

Sonuga-Barke, E. J. S., & Castellanos, F. X. (2007) Spontaneous attentional fluctuations in impaired states and pathological conditions: A neurobiological hypothesis. *Neuroscience and Biobehavioral Reviews*, 31, 977-986.

Spearman, C. (1910) Correlation calculated from faulty data. *British Journal of Psychology, 3,* 171-195.

Spencer, K.M. (2005) Averaging, detection, and classification of single-trial ERPs. in Handy, T.C. (Ed.) *Event-related Potentials: A methods handbook.* Cambridge: MIT Press.

Spencer, S.V., Hawk, L.W., Richards, J.B., Shiels, K., Pelham, W. E., & Waxmonsky, J. G. (2009) Stimulant treatment reduces lapses in attention among children with ADHD: The effects of methylphenidate on intra-individual response time distributions. *Journal of Abnormal Child Psychology, 37,* 805-816.

Stefanis, N.C., van Os, J., Avramopoulos, D., Smyrnis, N., Evdokimidis, I., & Stefanis, C.N. (2005) Effect of COMT Val^{158}Met polymorphism on the continuous performance test, identical pairs version: Tuning rather than improving performance. *American Journal of Psychiatry, 162*, 1752-1754.

Stern, W. (1911). *Die differentielle Psychologie in ihren methodischen Grundlagen*. Leipzig: Barth (Reprint 1994, Bern: Huber).

Strauss, E., MacDonald, S. W. S., Hunter, M., Moll, A., & Hultsch. (2002) Intraindividual variability in cogntive performance in three groups of older adults: Cross-domain links to physical status and self-perceived affect and beliefs. *Journal of the International Neuropsychology Society*, 8, 893-906.

Stuss, D.T., Stethem, L.L., Hugenholtz, H., Picton, T., Pivik, J., & Richard, M.T. (1989) Reaction time after head injury: Fatigue, divided and focused attention, and consistency of performance. *Journal of Neurology, Neurosurgery, and Psychiatry, 52,* 742-748.

Stuss, D.T., Pogue, J., Buckle, L., & Bondar, J. (1994) Characterization of stability of performance in patients with traumatic brain injury: Variability and consistency on reaction time tests, *Neuropsychology*, 8 (3), 316-324.

Stuss, D.T., Murphy, K.J., Binns, M.A., & Alexander, M.P. (2003) Staying on the job: the frontal lobes control individual performance variability. *Brain, 126*, 2363-2380.

Surwillo, W.W. (1975) Reaction-time variability, periodicities in reaction-time distributions, and the EEG gating-signal hypothesis. *Biological Psychology, 3,* 247-261.

Teicher, M.H., Lowen, S.B., Polcari, A., Foley, M., McGreenery, C.E. (2004) Novel strategy for the analysis of CPT data provides new insight into the effects of methylphenidate on attentional states in children with ADHD. *Journal of Child and Adolescent Psychopharmacology, 14(2)*: 219-232.

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D., L. (2010) Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence,* 38, 111-122.

Usher, M., Cohen, J.D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999) The role of locus coeruleus in regulation of cognitive performance. *Science, 283,* 549-554.

Van Ordern, G.C., Holden, J.G., & Turvey, M.T. (2003) Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132(3),* 331-350.

Verleger, R. (1988) Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences, 11,* 343-356.

Verleger, R., Jaskowski, P., & Wascher, E. (2005) Evidence for an integrative role of P3b in linking action to perception. *Journal of Psychophysiology, 19,*165-181.

Vinogradov, S., Poole, J. H., Willis-Shore, J., Ober, B. A., & Shenault, G. K. (1998) Slower and more variable reaction times in schizophrenia: what do they signify? *Schizophrenia Research 32*, 183-190.

Vogel, E. K., & Luck, S., J. (2000) The visual N1 component as an index of a discrimination process. *Psychophysiology*, 37, 190-203.

Wagenmakers, E-J., Brown, S. (2007) On the linear relationship between the mean and standard deviation of a reaction time distribution. *Psychological Review, 114(3)*, 830-841.

Walhovd, K. B., & Fjell, A. M. (2007) White matter volume predicts reaction time instability. *Neuropsychologia, 45,* 2277-2284.

Walter, W.G., Cooper, R., Aldridge, V.J., McCallum, W.C., & Winter, A.L. (1964) Contingent negative variation: An electric sign of sensori-motor association and expectancy in the human brain. *Nature, 203,* 380-384.

Watter, S., Geffen, G. M., & Geffen, L. B. (2001) The n-back as a dual task: P300 morphology under divided attention. *Psychophysiology, 38,* 998-1003.

Weissman, D.H., Roberts, K.C., Visscher, K.M., & Woldorff, M.G. (2006) The neural bases of momentary lapses in attention. *Nature Neuroscience, 9(7),* 971-978.

Williams, B.R., Hultsch, D.F., Strauss, E.H., Hunter, M.A., & Tannock, R. (2005) Inconsistency in reaction time across the life span. *Neuropsychology, 19(1),* 88-96.

Williams, B.R., Strauss, E.H., Hultsch, D.F., Hunter, M.A., & Tannock, R. (2007) Reaction time performance in adolescents with attention-deficit/hyperactivity disorder: Evidence of inconsistency in the fast and slow portions of the RT distribution. *Journal of Clinical and Experimental Neuropsychology, 29(3),* 277-289.

Winterer, G., Musso, F., Vucurevic, G., Stoeter, P., Konrad, A., Seker, B., Gallinat, J., Dahmen, N., & Weinberger, D.R. (2006) COMT genotype predicts BOLD signal and noise characteristics in prefrontal circuits. *NeuroImage, 32,* 1722-1732.

Woody, C.D. (1967) Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical and Biological Engineering, 5,* 539-553.