ELSEVIER

Biological Psychology 47 (1998) 159–179

BIOLOGICAL
PSYCHOLOGY

Event-related potentials can reveal differences between two decision-making groups

Tim R.H. Cutmore *, Tammy D. Muckert

School of Applied Psychology, Griffith University, Brisbane 4111, Queensland, Australia

Accepted 14 August 1997

Abstract

Previous research has shown that a complex decision is dependent on an underlying utility metric that is used by decision making processes to accumulate preference for one alternative. This study postulated that a state of indecision may arise if this underlying metric is poorly organized. The underlying metric was examined with a paired comparison task while measuring event-related potentials (ERP) for subjects classified as 'career decided' and 'career undecided'. Stimuli for comparison were presented either sequentially or simultaneously. The simultaneous condition produced results consistent with the hypothesis that undecided subjects have a poorly organized value metric as revealed in both the behavioral data and the P3 component. A relationship between P3 amplitude and word distance on the underlying metric was found only for the decided group. This was interpreted in terms of the previously documented relationship between P3 and the constructs of decision confidence and task difficulty. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Decision making; ERP; P3; Career; Indecision; Equivocation

1. Introduction

Decision making takes many forms, from simple choice reaction time tasks to complex deliberative processes such as would be used by military planners, business

* Corresponding author. Tel.: +61 7 38755116; fax: +61 7 38756637

PII S0301-0511(97)00023-9

managers, physicians contemplating treatment strategies or people considering career alternatives. Although considerable literature has examined simple decision making, less is known about the cognitive factors that underlie complex decision making processes. In particular, little is known about the cognitive variables that may distinguish individuals who are very capable at deliberative decision making from those who flounder with such tasks. The purpose of this study was to investigate individual differences in decision making, particularly in relation to P3 amplitude as an indicator of the quality of an underlying metric that supports decision processes.

1.1. ERPS and individual differences

The study of individual differences in cognitive abilities has been facilitated by the use of event-related potential (ERP) measures. This is because these measures provide access to intervening processes that mediate observed behavioral differences. In addition, understanding of the cognitive correlates of an ERP (and thus the cognitive processes that the ERP reflects) is advanced by comparing groups that clearly differ in their cognitive or behavioral attributes. Some examples are briefly mentioned.

It has been observed for over four decades that attention problems occur in schizophrenic populations (McGhie and Chapman, 1961; Salisbury et al., 1994). In recent years, it has been shown that the N2 component distinguishes schizophrenic subjects from normal subjects on a difficult stimulus discrimination oddball task (Salisbury et al., 1994). A better understanding of both attention in schizophrenia and the meaning of the N2 is fostered by studies of this kind. Similarly, Klorman et al. (1990) has shown that methylphenadate, which reduces symptoms in attention-deficit disorder (ADHD) children, also results in an increased Nl amplitude during an auditory oddball task. The N1 has been argued to reflect selective attention processes (Naatanen and Picton, 1987). Finally, late ERP components such as P3 have been examined in groups at risk of alcoholism. Polich et al. (1994) conducted a meta-analysis of studies showing a reduced amplitude in this component for those subjects with a family history of alcoholism and they argue that P3 may, therefore, be a useful marker for identifying persons with a biological risk for this disease. These studies illustrate the importance of using ERPs in studies of individual differences.

Palmer et al. (1994) asserted that P3 is 'a valid index of central information processing during task-related decision making', and went on to show that P3 amplitude is inversely related to decision difficulty in same-different letter classification tasks. This finding extends previous studies postulating that P3 amplitude was inversely related to decision confidence (Hillyard et al., 1971; Ruchkin et al., 1980) and would appear to be incorporated within the construct of equivocation (Johnson, 1986, 1993). These studies typically examine the effects of stimulus and task variations on ERPs with subject variables either held constant (in within-subject manipulations) or treated as random error variation (in between-subjects manipulations). Thus, a relatively smaller P3 amplitude may be attributed to processing a stimulus that involves a difficult discrimination and subjects may report having less confidence in responding to these stimuli. However, the stimulus (or task) may also be held constant and individual differences examined. Thus, a smaller P3 observed in one group may be attributed to these subjects finding the stimulus (or task) to be more difficult to process. Although groups formed on the basis of individual differences complicates making causal inferences, identifying a theoretically meaningful difference between the groups is an important clue to why the groups differ. In particular, individual differences in an ERP may suggest differences in cognition. The logic of this approach may be stated as follows: If the groups are initially formed on the basis of a measure that bears a theoretically meaningful relationship to cognitive variables reflected by the ERP (e.g. decision difficulty), then it is reasonable to infer that the observed ERP difference is an index of the underlying construct that differentiates the groups. As in all studies of individual differences, other incidental differences between the groups that might reasonably be expected to result in the ERP difference should be taken into account in the methodology and when interpreting the meaning of the ERP difference. Factors such as age, gender, seasonal variation and time of day should be randomized or directly controlled to avoid confounding (Polich and Kok, 1995). These preliminary remarks serve to introduce two things that first must be identified before a study of individual differences in decision making can be usefully investigated using ERP measures. First, an underlying construct for differentiating the groups needs to be articulated. Second, a task that taps this construct needs to be defined.

In summary, there are important reasons for using ERP measures in the study of individual differences. The logic in comparing different groups of individuals is grounded in defining a theoretically meaningful construct for differentiating the groups and relating this to a cognitive variable that is reflected by an ERP measure. In the following sections it will be argued that two groups that differ in decision-making ability should show differences in the P3 component since this component is sensitive to task decision difficulty.

1.2. Levels of a decision-making process

Educational and counseling psychologists have long recognized a problem in some individual's ability to make a career decision (Katz, 1963; 1987; Jones, 1989; Rojewski, 1994). Katz (1963) outlined a career decision theory postulating that a complex decision such as selecting a career is motivated by an attempt to maximize the satisfaction of a set of elementary values. By values, Katz meant that some concepts such as money, security, independence, etc. are important to the individual and a career is chosen to, in some sense, fulfill them. Although Katz did not specify how these simpler elements might be combined to resolve a complex decision, there is clearly a separation of levels apparent in the theory. The career decision is a higher-level decision (that may be referred to as a macro-decision) that is based on the resolution of simpler elementary decisions (or micro-decisions) about values. If an individual is confused or uncertain regarding the ordering of career-related values, then one would anticipate that making a career decision would be problematical.

Busemeyer and colleagues have endeavored over several years to describe a model that can account for complex decision making behavior, such as career decision making and have proposed decision field theory (DFT; Busemeyer and Townsend, 1993). Unlike earlier decision theories, such as signal detection theory that address simple presence/absence decisions, DFT is a theory of human deliberative processes. One of its fundamental components is a subsystem that resolves the importance or utility of outcomes. The top portion of Fig. 1 illustrates the two levels implied by this model. As a specific example, consider an individual trying to make a career choice. There are a number of consequences in selecting a career: money, independence, helping, etc. For a particular individual, Katz's theory postulates that the relative value of these consequences are used to influence the final decision of a career choice. DFT makes this process explicit: pairs of these outcomes (e.g. money and helping) would be selected by a stochastic process and credit would accumulate from computing the utilities for each career alternative being considered. A utility is computed by summing the weights of outcomes. A weight refers to the subjective probability of a career alternative producing the outcome. For example, the career alternative 'police officer' may have a low weight assigned to the outcome money but a high weight assigned to the outcome helping. A career choice (action) is made when one of the career alternatives acquires credit beyond a threshold (a parameter in DFT). Although DFT does not address how the utilities are defined, it appears to be implicitly assumed that they can be organized on an underlying metric.

It is clear that the distribution of values (money, helping etc.) on the metric will have a direct impact on the deliberative process since the variance in the credit across career alternatives is a function of both the weights and the variance of the values. This is because the credit is a weighted composite (Mulaik, 1972, p. 71). In the extreme case, if there were no variance in the credit accumulated by the career alternatives then all alternatives would appear equally attractive and confound the deliberation process with the individual left in a state of career indecision. Although this is not likely to be the only possible cause of career indecision, it was the hypothesis investigated in this study. Fig. 1 (lower panel) presents examples of two individuals value metrics. In one case the values are well ordered and have large variance that should contribute to variance in the credit function facilitating deliberation. In the other case the values are all similar (confused) leading to less variance in the credit function and stalling the deliberation process. It was hypothesized that career decided individuals have the first metric and the career undecided have the second metric. The problem then arises as to how to assess this underlying metric.

1.3. Paired comparison tasks and P3

A subject's metric for a quantity has been assessed using comparative judgment tasks. A comparative judgment is made when two or more stimuli are evaluated and one chosen as having the highest (or lowest) rating on some dimension. For example, several digits may be presented and the task is to choose the largest (or

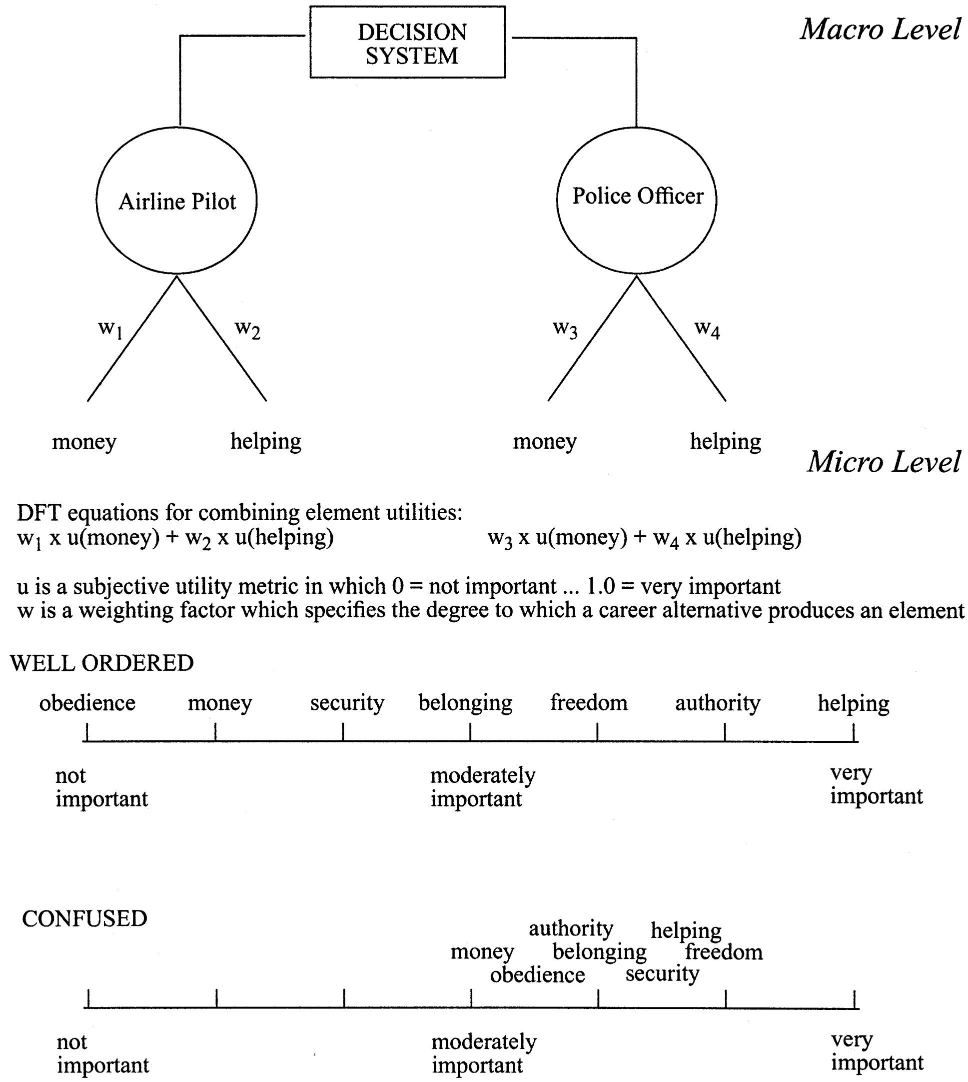

Fig. 1. Application of DFT to choosing between two career options with only two value elements considered (money, helping). At the macro level a choice is to be made; choose either the career option of airline pilot or police officer. DFT specifies equations for determining the combined utilities of each option. In the full theory, there may be several options at the macro level, several elements at the micro level and the w values may change during deliberation. The difference between the options is accumulated in the decision system until a threshold is exceeded. The theory also includes parameters which bias the decision system at the outset, accounts for primacy and recency effects of memory and approach/avoidance gradients in a continuous time model. In the lower panel two example subjective metrics are given for career values. One is well ordered and will facilitate decision making while the other is confused and may result in a state of indecision.

smallest). In its simplest form, the comparison is between two stimuli. The metric may be objective as in the case of digit comparison (Link, 1990) or subjective as in a task in which the subject must choose the more favorite of two colors. The metric for career value words is an example of a subjective metric since different people will generally have different rankings of these values. Studies of RTs in a comparative judgment task have revealed that this can be used to examine the nature of the underlying metric. Moyer and Landauer (1967, 1973) and Link (1990) have shown that the nearer the two numbers are for comparison, the longer the RT. This has been referred to as the 'symbolic distance effect'. Paivio (1975) has shown that size comparison of remembered objects produces similar results. End anchors have been shown to contribute a separate effect. If the anchor is processed first then a response may be produced without processing the other item (Potts, 1974) resulting in a fast RT.

Relating the distance effect to decision difficulty, it may be surmised that objects or numbers nearer to one another on their metric present a more difficult comparison task. Given the earlier discussion regarding task decision difficulty and P3, one would generally also expect to see a relatively smaller P3 amplitude during stimulus comparisons between items that are near on the metric. Distant stimuli could be compared with relatively high confidence, whereas, adjacent stimuli would be compared with relatively low confidence. This hypothesized relationship concerning the distance between stimuli and P3 amplitude is consistent with previous studies that postulate P3 amplitude reflects mental distance to an internal standard or 'adaptation level' (Ullsperger and Gille, 1988; Grune et al., 1993). Recently, Mecklinger et al., 1994 presented a conceptualization in terms of information extraction. Subjects were presented with spoken digits in the range of 1–5 with the task of determining whether each digit was greater or smaller than the preceding one. The P3b amplitude covaried with digit nearness to end anchors (i.e. to 1 or 5). The P3b was concluded to be an index of information extraction regarding the response since end anchors deliver full information whereas intermediate terms deliver less information; although a decision confidence interpretation is also possible for these results.

However the P3 (or P3b) amplitude is interpreted (either as an index of information extraction or as an index of decision difficulty) it is clear that the underlying metric should affect P3 amplitude during comparative judgments. Under either interpretation, if the underlying metric is confused (i.e. items are not distinctly different on the metric) P3 amplitude will generally be lower than if it is well ordered. With a confused metric an item will deliver little information regarding future responses and it will be difficult to decide whether it is higher or lower on the metric than some other item. It was not the purpose of this study to resolve this interesting theoretical issue but rather to be pragmatic in the use of the P3 as an index of the general quality of the underlying metric. Furthermore, in this initial study, the simplifying assumption was made that the career value metric is unidimensional.

In summary, comparative judgments between stimuli result in behavioral and ERP differences that reflect the standing of those stimuli on an internal metric. This metric may be objective (digit comparison) or subjective (career values). P3 amplitude appears to vary with a stimulus' 'distance' from either an internal standard or another stimulus. This effect may reflect either the difficulty of performing a comparative judgment or the information delivery of the stimulus and indeed these two interpretations may not be independent. Pragmatically, it is concluded that the P3 amplitude should be a useful means of quantifying the general nature of the underlying metric either as confused (small P3s) or well ordered (large P3s).

1.4. A prediction of P3 amplitude for two decision groups

This study compared two groups of subjects classified as able or unable to make a career decision. It was predicted that the undecided subjects would show smaller P3 amplitude than decided subjects because the value metric they use to deliberate and choose a career is confused. In particular, it was hypothesized that the symbolic distance effect should be more apparent in the decided group because this effect assumes that elements being compared are well ordered. A methodological issue that arose was how the two value elements should be presented, that is, sequentially or simultaneously. Although it was not an objective of the study to address the considerable debate over sequential (serial) and simultaneous (parallel) processing and cognition, it was of interest to determine whether the method of presentation could make a difference. It is possible, for example, that group differences may only appear under one method, or that the decision groups may have preferred processing styles. If either of these is the case then the presentation method (processing) would interact with decision group. It is not precisely known how individuals perform the internal paired comparisons during actual deliberation. This may be done by mentally comparing two (or perhaps more) values simultaneously whilst performing a macro decision. If this is the case then the simultaneous presentation method may best exemplify this process. However, including the sequential presentation method better replicates other ERP studies which have used paired comparisons, so it is important to compare these two methods. Finally, ERPs were extracted for the early components that have been associated with attentional processes in the event that group differences might be evident in the early stages of information processing. No specific hypotheses were developed for these measures and they may be considered exploratory.

2. Method

2.1. Subjects

The career decision scale was completed by 144 undergraduate students from arts and science courses. A subset of 48 (35 females) subjects with a mean age of 24 years were recruited on the basis of their scores (see details below).

2.2. Materials

The career decision scale (CDS; Osipow et al., 1976; Osipow, 1980; 1991; Vondracek, 1991) is the earliest published and most widely used scale of career indecision. It contains a 16 item indecision scale with each item requiring a rating on a four-point likert scale (1, not at all like me... 4, exactly like me). The indecision scale consists of items such as 'Several careers have equal appeal to me. I'm having a difficult time deciding among them'. It has a 2-week test-retest reliability of 0.90 and 0.82 from two separate samples (Osipow et al., 1976) and a 6-week test-retest reliability of 0.70 (Slaney et al., 1981). Percentile tables are included with the test and provide a means for selecting two extreme groups: career decided and career undecided.

The career values card sorting activity (CVCSA) developed by Knowdell (1993) is a test that yields the rank ordering of a persons career values. Forty-one cards are sorted into five categories: always valued, often valued, sometimes valued, seldom valued and never valued. Under each category the cards are placed in a rank order of importance.

Two paired comparison tasks were used and each required subjects to view two stimuli presented on the monitor and to choose the one that was higher on a specified attribute (either the larger of two digits, or the more important of two words). Digits in the range 1–5 were used as in Mecklinger et al. (1994) and 20 value words were selected from the CVCSA and from Schwartz and Bilsky (1990). list of universal values. A constraint on word selection was that there be 12 letters or fewer to reduce the risk of eye movements occurring when they were read. All digits and letters were 13 mm in height (subtending a visual angle of approx. 0.5 degrees) presented in helvetica font. All stimuli were presented on a 15 in., high resolution color monitor in white on a black background. One randomized list of 36 digit pairs (three replications of 12 pairs) was created with the constraint that the first stimulus not be an end item (1 or 5) since this eliminates the need for a comparison. Ten randomized lists of the 380 possible pairs of the 20 value words were created (both orders were used: a, b and b, a). Thus each word was used equally often and in all combinations with other words.

2.3. ERP recording and averaging

An electrode cap with tin electrode contacts in the international 10–20 positions (Jasper, 1958) was used, with a ground electrode in the mid-line halfway between Fz and Fp electrodes. Eye-related artifact (eye movement and blinks) was examined at HEOG and VEOG and forehead muscle artifact was examined at Fp1, Fp2, F7 and F8. The recording electrodes selected for analysis here were: F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. All electrodes were referenced to linked ear lobes. Prior to fitting the cap, skin contact points were cleaned with alcohol. One of two cap sizes was selected depending on head size. Each electrode site on the cap was filled with a high conductance gel. Impedance of all recording sites was brought to less than five kΩ. The cap was connected to a NeuroScan synamps system with digitization rate set at 250 Hz, band pass filtering (between 0.1 and 40 Hz) and signals amplified by a factor of 30 000. A sweep was recorded 100 ms pre-stimulus to 1000 ms post-stimulus and saved to disk. In preparation for signal averaging, sweeps containing significant (± 80 μV) artifact in one of the artifact channels between 0 and 800 ms, were discarded. This amounted to 12% of the trials, on average, being discarded. Averaged ERPs were computed for each subject on each task (digits and words) to identify components of neural activity that are synchronized to the stimulus (Coles et al., 1990).

2.4. Procedure

From the initial contact group of 144 subjects, two extreme groups were formed with 24 subjects in each. The 'undecided' group was defined by scores at the 82nd percentile or above and the 'decided' group by scores at the 15th percentile or below on the indecision scale. The undecided group was composed of seven males and 17 females and the decided group was composed of six males and 18 females. This gender distribution reflected that of the original contact sample. The groups were similar in mean age (decided, 22 years and undecided 25 years, $t(46) = 1.15$, n.s.). Subjects were tested between 09:00 and 17:00 (randomly assigned over decision groups) and all subjects were tested in the winter months of July and August thus reducing any possible influence of seasonal variation on ERPs (Polich and Kok, 1995). The 20 words used in the word comparison task (or words task) were shown to subjects in advance and they were asked to briefly define each word. This was to ensure all word meanings were understood. The subject was prepared for ERP recording and seated in a padded recliner in a shielded room. The monitor was placed 800 mm away at head height.

Subjects from each decision group were randomly assigned either the sequential or simultaneous presentation methods for both tasks and randomly assigned one of the 10 word-pair lists. For sequential presentation the first stimulus of the pair was presented for 500 ms with an interval of 500 ms to the second stimulus that remained on screen until a response was made. For the simultaneous presentation method the same stimulus pairs as in the sequential method were presented on the same horizontal axis with a space between them. These remained on screen until a response was made. For both presentation methods the interval from the response to the next stimulus was 2000 ms. Instructions to subjects stressed accuracy. This was deemed essential as there was no way to score the accuracy of the value word comparison since it depends on each subject's particular (idiosyncratic) value metric.

The 36 trials of the digits task were presented first. Subjects were instructed to compare the two stimuli and press the left button of a button pad if the first (or left, simultaneous) stimulus was larger and to press the right button if the second (right) stimulus was larger. The 380 trials of the words task was then given with similar response instructions except that the 'more important' of two personal career values was to be used to make a decision. Two rest periods were given at one and two thirds the way through this task. Twelve practice trials were given prior to each task. RTs were recorded for the words task and RTs and accuracy for the digits task. Upon completing the words task the electrode cap was removed and the subject was asked to complete the CVCSA.

2.5. Data analysis

The N1, N2, P2 and P3 components were examined separately as these reflect different cognitive processes (Coles, et al., 1990; Regan, 1989; Donchin et al., 1986; Lindholm and Koriath, 1985). ERPs were baseline adjusted on the average voltage over the 100 ms pre-stimulus interval. Peaks were identified by a computer program that scanned for the maximum voltage in predefined intervals: N1 (75–150 ms), N2 (150–260 ms), P2 (150–280 ms), and P3 (400–750 ms). These intervals were selected after manually scoring a few subjects at Fz, Cz and Pz to determine the appropriate latency ranges. Comparison of component latency and amplitude for manual and automated scoring produced nearly identical values.

Mixed analysis of variance (ANOVA, Keppel, 1991) was used with factors defined as follows: decision group (undecided, decided) and processing (sequential, simultaneous) were between groups factors; task (digits, words) was a within groups factor. Nl, N2 and P2 were examined at Cz where it is generally large and most stable. Similar to the Palmer et al. (1994) analysis of P3, Fz, Cz and Pz were analyzed to examine task and processing effects. This was followed by an examination of lateral sites at frontal, central and parietal regions to specifically test for possible interactions of group and site. Thus, in this second analysis a restricted set of results were interpreted since many effects would be redundant to the central site analysis. Following Vasey and Thayer (1987) effects involving repeated measures were corrected for violations in sphericity using the Greenhouse and Geisser (1959) procedure.

3. Results

3.1. Behavioral measures

Accuracy data for the digits task showed very few errors (less than 3%) and thus all trials were included in the averages. Table 1 shows mean frequency of choice, S.D. and range for the 20 words used in the ERP task. The maximum possible score is 38 (the number of times each item was paired with all the rest in both orders). It is clear that there was considerable variability in individual metric for this comparative judgment task. Indeed, most elements span nearly the entire range attesting to the subjective nature of this task.

3.1.1. Reaction time

The RT distributions for button presses on both the digits and words tasks were found to have moderate positive skew, therefore the natural logarithm was computed prior to analysis. The factors analyzed were group, processing and task.

Table 1
Words used in the word comparison task

Words used in the comparison task	Mean choice frequency	S.D.	Minimum	Maximum[b]
Happiness	30.52	6.20	9	38
Friendships[a]	26.63	8.15	9	37
Competence[a]	25.75	7.38	6	37
Freedom[a]	25.71	6.76	9	35
Independence[a]	25.48	5.75	11	36
Wisdom	25.17	7.86	7	37
Helping[a]	21.15	9.10	5	37
Belonging	20.98	6.96	4	32
Security[a]	20.79	9.32	5	36
Creativity[a]	20.73	7.69	4	36
Intellectual[a]	20.65	6.63	5	34
Recognition[a]	19.98	6.77	4	34
Excitement[a]	19.35	7.75	5	36
Money	15.40	10.92	0	36
Deciding[a]	14.52	6.62	1	27
Status[a]	13.52	8.42	2	35
Change[a]	11.71	7.43	1	30
Authority[a]	9.27	7.95	0	34
Location[a]	7.67	5.27	0	28
Obedience	4.94	4.22	0	23

[a] These words are from the career values card sort and were used in computing the consistency score for each subject.
[b] The maximum possible score was 38.

There was a significant main effect of task with digit comparison being faster ($F(1, 44) = 118.58$, $P < 0.001$). Table 2 shows that there is a greater difference between the two processing conditions for words than for digits, resulting in significant interaction of processing and task ($F(1, 44) = 23.88$, $P < 0.001$).

3.1.2. Consistency

Value words common to the ERP task and the CVCSA (there were 15 in common) were examined for rank ordered consistency. A rank order was derived from the ERP task by counting the number of times a word in each word pair was selected as more important. The CVCSA provides a ranking explicitly. The

Table 2
Mean (and S.E.M.) RTs (ms) to digit and word comparisons under conditions of sequential or simultaneous processing

	Digits	Words	
Sequential	809 ($\pm$38)	1092 ($\pm$70)	951 ($\pm$40)
Simultaneous	671 ($\pm$54)	1420 ($\pm$90)	1090 ($\pm$58)
	740 ($\pm$35)	1256 ($\pm$59)	

Table 3
Mean (and S.E.M.) for word value consistency scores under conditions of sequential and simultaneous processing on the ERP task

	Undecided	Decided	
Sequential	0.60 ($\pm$0.04)	0.48 ($\pm$0.08)	0.54 ($\pm$0.04)
Simultaneous	0.46 ($\pm$0.09)	0.67 ($\pm$0.05)	0.56 ($\pm$0.06)
	0.53 ($\pm$0.06)	0.57 ($\pm$0.05)	

Spearman correlation was computed for each subject and used as a 'consistency' dependent variable in a group by processing ANOVA. The pattern of means in Table 3 illustrate a significant interaction of these factors ($F(1, 44) = 4.51$, $P < 0.05$). Each processing condition was examined separately for group differences with simple main effects analysis. As predicted, in the simultaneous condition, the decided group had a significantly higher consistency score than the undecided group, ($F(1, 22) = 2.79$, $P < 0.05$, directional test).

3.2. ERP measures

The grand average wave forms for each task and decision group are shown in Fig. 2 for Cz, Pz and VEOG. Larger baseline variability for the digits task reflects the smaller number of wave forms in these averages.

3.2.1. Early components

The Nl, N2 and P2 were each analyzed at Cz and had latencies of 97, 192 and 224 ms, respectively.

3.2.1.1. Amplitude. For the N1 component, the undecided group had a larger mean amplitude (-1.96 ± 0.30 μV) than the decided group (-1.00 ± 0.27 μV), ($F(1, 44) = 6.97$, $P < 0.05$), and the sequential processing condition produced a larger mean amplitude (-1.99 ± 0.29 μV) than the simultaneous condition (-0.97 ± 0.28 μV), ($F(1, 44) = 6.17$, $P < 0.05$). For the N2 component, there were no significant amplitude effects. For the P2 component, the simultaneous condition produced a larger mean amplitude (7.72 ± 0.78 μV) than the sequential condition (4.52 ± 0.51 μV), ($F(1, 44) = 11.73$, $P < 0.001$).

3.2.1.2. Latency. For each component, there was a shorter mean latency for the digits task than the words task. The means and effects were as follows: for N1 digits 89 ± 3 ms, words 103 ± 3, ($F(1, 44) = 12.82$, $P < 0.001$),for N2 digits 185 ± 4 ms, words 199 ± 4 ms, ($F(1, 44) = 6.76$, $P < 0.05$), and for P2 digits 213 ± 6 ms, words 234 ± 5 ms ($F(1, 44) = 9.03$, $P < 0.01$).

3.2.2. P3 component global analyses

The mean latency of the P3 component was 502 ms. A four factor ANOVA was performed for amplitude and latency with the factors being decision group, processing, task and electrode site. In any case where electrode site interacted with another factor for the amplitude measure, the analysis was recomputed on vector normalized scores as recommended by McCarthy and Wood (1985), and both are reported.

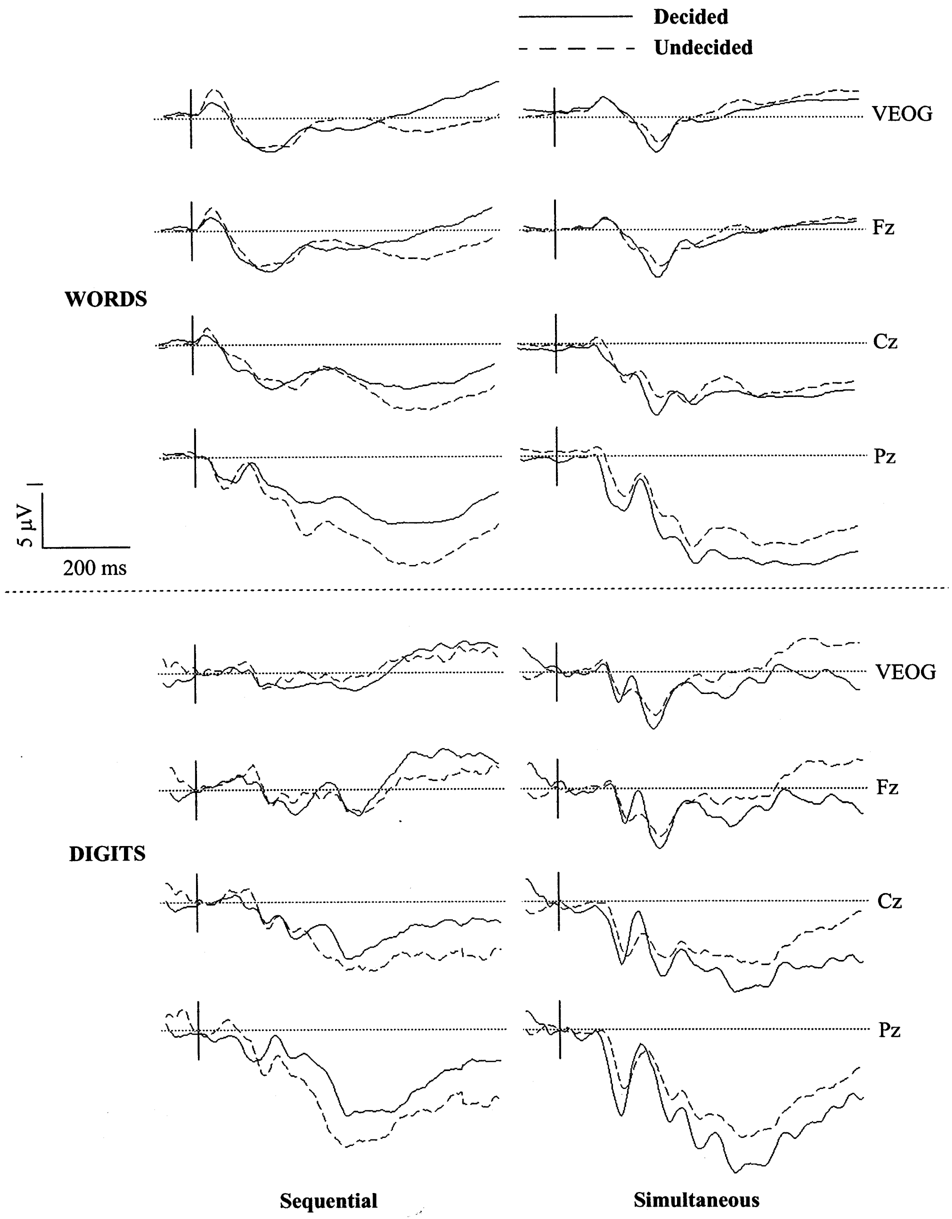

Fig. 2. Grand average wave forms for each decision group, processing condition and task at Fz, Cz, Pz and VEOG.

3.2.2.1. Amplitude. There were two main effects. The digits task produced larger mean amplitudes than the words task ($F(1, 44) = 11.40$, $P < 0.01$) and there was an effect of site ($F(1.25, 54.84) = 105.58$, $P < 0.001$) with the largest amplitudes recorded at the Pz site ($P < 0.001$). Decision group and processing interacted ($F(1, 44) = 4.66$, $P < 0.05$), see Fig. 3. Although the effect is most apparent parietally, there was no significant interaction with electrode site. Directional tests were used to assess whether the decided group had a larger mean P3 amplitude than the undecided group. This was found for the simultaneous processing condition ($P < 0.05$). Task and processing interacted ($F(1, 44) = 3.92$, $P < 0.05$). Simple effects analysis showed the digits task to have a larger amplitude than the words task only for the simultaneous processing condition ($P < 0.01$). Task interacted with site ($F(1.30, 57.28) = 3.94$, $P < 0.05$), but this failed to reach significance for the vector normalized amplitudes.

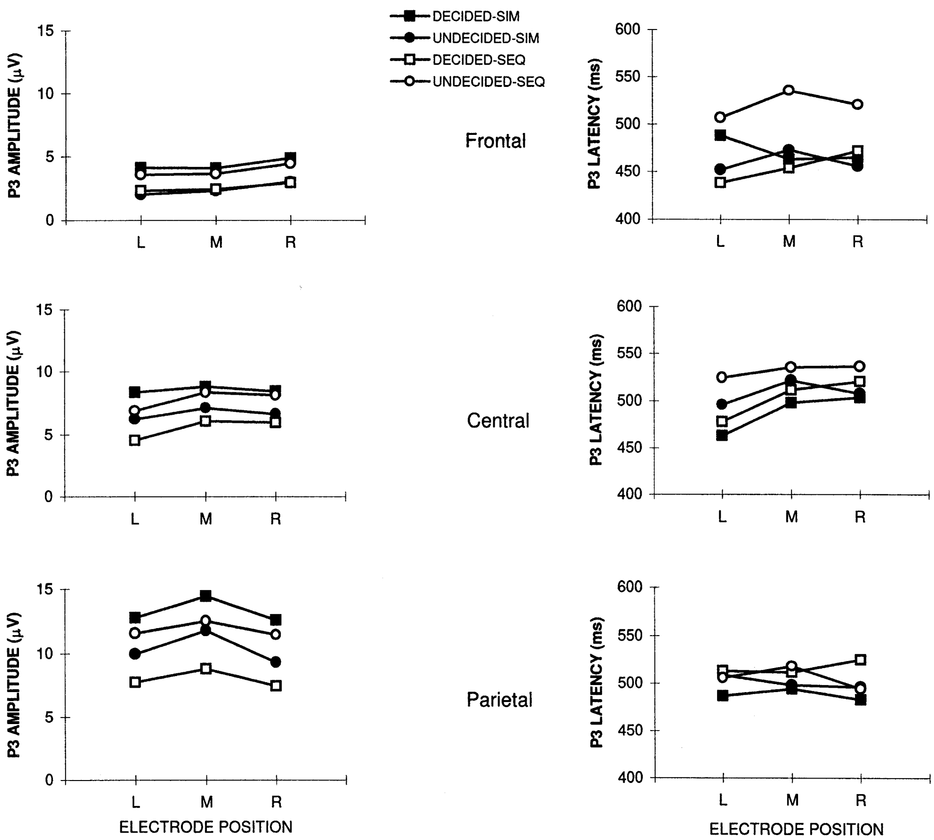

Fig. 3. P3 amplitude (left) and latency (right) for each decision group and processing condition averaged over tasks. Frontal, central and parietal regions are shown for left (L) midline (M) and right (R) electrode sites.

3.2.2.2. Latency. There were two significant main effects. Digits produced a shorter latency than words ($F(1, 44) = 43.35$, $P < 0.001$) and there was an effect of site ($F(1.95, 85.79) = 11.04$, $P < 0.001$). Decision group and site interacted ($F(1.95, 85.79) = 3.56$, $P < 0.05$) and decision group, site and processing interacted ($F(1.95, 85.79) = 3.58$, $P < 0.05$), see Fig. 3. Post hoc analysis failed to resolve the locus of these effects. However, they may be due to uncertainty in peak location at Fz where P3 amplitudes are relatively small and not well defined.

Analysis of lateral sites in frontal, central and parietal regions, testing specifically for any lateralization differences between the decision groups, did not produce any significant effects for either amplitude or latency measures.

3.2.3. P3 amplitude and word metric distance

The preceding results of the P3 component showed the largest amplitude to be at Pz. A group by processing interaction was also found with a predicted group difference found for the simultaneous processing condition. To examine more specifically the hypothesized relationship between P3 amplitude and decision confidence in using the underlying value word metric, an analysis at Pz was undertaken to compare decision groups under the simultaneous condition. This was done to determine whether the decided group (compared with the undecided group) would show a better linear relationship between the 'symbolic distance' separating word pairs and P3 amplitude. Decisions involving greater inter-word distances should incur greater confidence and thus greater P3 amplitude. The position of each word on the value metric was defined as the rank ordered preference over other words. This was computed by compiling the frequency with which a word was chosen in the word pairs. Next, the distance between each word pair was defined by the difference in ranking. For example, if one word of a pair was ranked 1 (most valued) and another word was ranked 10, then the difference between them would be 9. To acquire ERP averages the word pair distances were divided into three sets. Set 1 was composed of all word pairs that had a distance of 1–6, set 2, 7–12 and set 3, 13–19. As argued in the introduction, the decided group was anticipated to show a better linear relationship between a measure of decision confidence (P3) and inter-word distance because their underlying value metric is better ordered. The analysis was performed with and without the end anchors (words ranking either 1 or 20) and the same pattern of results were found. The interaction between decision group and word distance was significant ($F(1.79, 39.9) = 3.28$, $P < 0.05$). This was decomposed with trend analysis, as recommended by Keppel (1991), and the only significant finding was a linear trend for the decided group ($P < 0.05$) A similar analysis was conducted on RTs but no group differences were found.

4. Discussion

The purpose of this study was to determine whether ERP components, the P3 in particular, can be used to understand individual differences in ability to make complex decisions. Katz (1963) and Busemeyer and Townsend (1993) provided a theoretical framework that permitted the prediction that high-level decision making problems (being in a state of indecision) may be due to difficulties at a more elementary level at which values (utilities) are organized on a metric. Specifically, it was hypothesized that undecided subjects (for example, in trying to make a career decision) may have a more confused value metric. The utilities on this metric are used, according to Busemeyer and Townsend (1993) to drive the decision process towards a resolution. The results are first discussed with respect to the evidence that confirms this hypothesis. This is followed by a section that briefly reflects on the tasks used to examine this hypothesis and finally the implications for theories of complex decision making.

4.1. Decision group differences

For the early components (N1, N2 and P2), the only one that revealed a decision group difference was the N1 amplitude. Undecided subjects had a larger N1 amplitude than decided subjects. In a review of auditory N1, Naatanen and Picton (1987) identified selective attention as a factor that contributes to N1 amplitude. In particular, the review suggested that the N1 amplitude may be increased by 'prior preparation for a demanding task' (p 404). Harter and Aine (1984) have a shown similar effect of attention with visual stimuli. Therefore, a possible reason for undecided subjects showing a larger N1 amplitude was that the paired comparison tasks were experienced as more difficult or demanding than for decided subjects thus requiring a greater degree of attention or processing resources.

Previous studies have found two or more late positive components in choice reaction time tasks (e.g. Falkenstein et al., 1993; Mecklinger et al., 1994; Donchin et al., 1975). The paired comparison tasks in the present study yielded one clearly visible late positive component at a latency of about 500 ms. This component appears to be similar to the P3b component of Mecklinger et al. (1994) who used a similar paired comparison paradigm. It also compares favorably with the P3 component of Palmer et al. (1994) which found that P3 amplitude covaries with task difficulty and used the construct of a posteriori uncertainty ('equivocation') to account for this result. Johnson (1986) presented a triarchic model in which the P3b was identified with the P3 and argued for the influence of information transmission (also described in terms of 'equivocation', 'uncertainty' and 'stimulus discriminability', p374) on this component. Thus, the P3 component identified in this study may be similar to the P3b component in other studies.

Before examining the P3 results in detail it should be recalled from the introduction that the way in which stimuli are presented (and thus processed) in a paired comparison task may be an important factor. In particular, differences between decision groups may be sensitive to this methodological variable. This was confirmed in both the behavioral and ERP analysis with significant processing by group interactions. Group differences that are consistent with the hypothesis that undecided subjects have problems with their value metric were only revealed when values were compared in the simultaneous condition. It is unclear, at present, why the sequential condition tended to produce the opposite effect. However, Palmer et

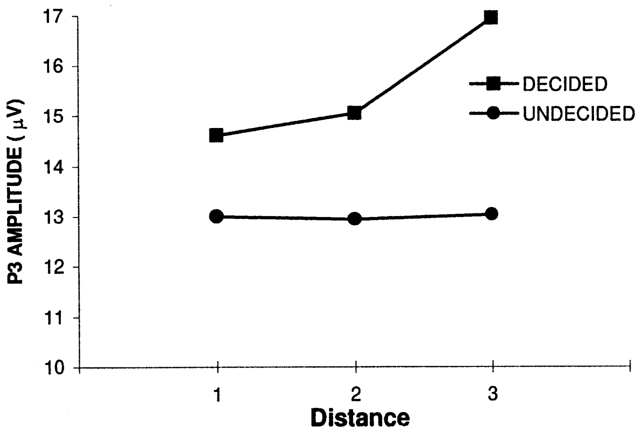

Fig. 4. Mean P3 amplitude at Pz for decided and undecided subjects under the simultaneous condition. The subjective distance between words in a pair was grouped into three categories with '1' being the shortest distance. A significant linear trend was found only for the decided subjects.

al. (1994) have noted that simultaneous presentation avoids memory requirements that are inherent to serial presentation and thus may provide a more direct assessment of decision making factors. The serial presentation method may thus mask true group differences on decision making variables such as the underlying metric or may introduce effects independent of the decision process. The discussion, therefore, focuses on the results of the subjects in the simultaneous condition. It should be noted that this was also the method of stimulus presentation used by Palmer et al. (1994).

The behavioral consistency results (the rank correlation between the CVCSA and ERP tasks) showed decided subjects to have greater consistency in their word metric (Table 3, simultaneous condition). Thus, decided subjects would appear to be utilizing some underlying metric which is more stable across different contexts. The P3 results showed a global effect with decided subjects having larger P3 amplitude than undecided subjects (Fig. 3, simultaneous condition). Having noted these effects, a more focal analysis of P3 at Pz for the simultaneous condition revealed that this component showed a linear relationship (the only statistically significant trend) with value word distance only for the decided group, as illustrated in Fig. 4. Previous studies (e.g. Polich, 1987; Palmer et al., 1994) have shown the P3 to be related to task decision difficulty with larger P3 amplitudes occurring with easier task decisions or stimulus discriminations. Earlier studies indicated that the P3 amplitude may be inversely related to the lack of confidence (Hillyard et al., 1971; Ruchkin et al., 1980) or equivocation (Johnson, 1986; Palmer et al., 1994) in making a decision. In the present study, the P3 may be interpreted in this way. With a value metric that is clearly defined (that is, items are distinctly separable and organized in a stable manner) it should be relatively easier to choose the higher item when the distance is large than when items are adjacent. The easier decision would be made with greater confidence and certainty and this would be reflected in a larger P3 amplitude. This component, which was maximal in the parietal region, showed this amplitude relationship for subjects who were hypothesized to have such a clearly defined value metric. It can be concluded, therefore, that the P3 component may be a useful marker that distinguishes these two decision groups when value words are compared simultaneously. Previous behavioral studies (e.g. Potts, 1974; Scholz and Potts, 1974) have shown that the metric presumed to underlie these decisions will contain end anchors that contribute effects that are separable from items between the anchors. The analyses performed in the present study showed the P3 relationship with word distance whether end anchors were included or not. Thus, this P3 relationship appears to function mainly of word distance and not an artifact of end anchor effects. This is, perhaps, not surprising given the large number of items (20) in the value word set. No group difference in RTs over word distance was found but this may reflect the instruction set which emphasized accuracy because of the subjective nature of the judgments that precludes detection of errors.

4.2. Task effects

The P3 group differences in the global analysis were evident across tasks (digits and words) with no interaction evident with this factor. The present study did not use a large number of digit comparisons (mainly to avoid subject fatigue) and thus a distance analysis of digits was not possible. It would be of interest to replicate these results with a digits task in which the relationship between P3 and inter-digit distance could be examined for two decision groups. Until this is undertaken it may be tentatively concluded that this subgroup of people may manifest a more general relationship between P3 amplitude and underlying metrics used to make decisions. That is, these subjects may be generally less confident in making paired comparison decisions, at least under simultaneous presentation. There appears to be no literature addressing the possibility that decision making difficulties may be general rather than specific in university populations.

The analysis of all components showed that the digits task produced a component peak sooner than the words task. This was evident as early as N1. If the information processing reflected at later components depends on that of earlier ones, this may explain the reappearance of latency differences at several ERP components as well as the RTs. This finding is not surprising given that more information is carried by the word stimuli than the digits. The difference at N1 may simply reflect that this difference is important at early perceptual stages (e.g. attentional filtering) of processing. The digits task also produced a larger P3 amplitude than the words task only for the simultaneous processing condition, reflecting the greater ease of comparing digits than value words. This result underscores the importance of the stimulus presentation method and reaffirms the possibility that under sequential presentation the P3 amplitude may reflect factors other than decision confidence.

4.3. Implications for theories of decision making

Complex decision processes such as those involved in making a career decision are common in many daily activities and thus it is important to understand the factors that moderate such decision processes. An early theory of career decision making presented by Katz (1963) suggested that underlying career values drive the process of making a career decision. Busemeyer and Townsend (1993) summarized a history of attempts to develop a more precise theoretical formulation for complex decision making (which they term 'deliberation'). One of the initial components of their theory was the postulation of utilities (and by implication some metric on which the utilities are organized). This component was maintained throughout several revisions of the decision making theory and remains an important component of the most recent version, DFT. The results of analysis of P3 provide evidence for the basis of this latent variable, the utility metric. Furthermore, Katz's postulate that career values are fundamental to the career decision making process appears substantiated. If the value metric is not clearly defined then the relationship between distance of items on the metric and amplitude of synchronous neural activity (P3) fails to hold. Since the P3 can be interpreted to reflect decision confidence then this failure of P3 activation may indicate a fundamental problem that may manifest itself in a state of indecision. In DFT this results in a delay in resolving the deliberation and in Katz's theory can result in chronic career indecision. More generally, a method has been found that may permit the study of internal metrics. A simplifying assumption of a unidimensional metric was made in the present study to make this initial study tractable. However, it is possible that the metrics underlying many types of deliberative processes are multidimensional. The distance in n-dimensional space between items may then be a better predictor of P3 amplitude.

In summary, previous studies have examined simple decision making processes using ERPs and have shown P3 to be an important indicator of the cognitive factors involved and for studying individual differences. The present study addressed the question of whether P3 might be a marker for discriminating between two decision making groups, those who can or cannot resolve a complex decision. The P3 component was found to distinguish the two groups provided a task was used that promoted simultaneous processing of stimuli. Furthermore, a linear relationship was found between P3 amplitude (interpreted as a measure of ease of decision making and thus decision confidence) and a distance measure for words being compared and this was only shown for a group identified as being decisive.

Acknowledgements

We thank Professor Mark Dadds and Ms Madeleine Brabant for comments on earlier drafts of this paper. This research was supported by an equipment grant from Griffith University Major Research Facilities Fund to T.R.H.C. A preliminary report of a portion of these results was presented to the 1994 meeting of the Australasian Psychophysiology Society.

References

Busemeyer, J.R., Townsend, J.T., 1993. Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. Psychol. Rev. 100, 432–459.

Coles, M.G.H., Gratton, G., Fabiani, M., 1990. Event related potentials. In: Cacioppo, J.T., Tassinary, L.G. (Eds.), Principles of Psychophysiology: Physical, Social and Inferential Elements. Cambridge University Press, New York, pp. 413–454.

Donchin, E., Tueting, P., Ritter, W., Kutas, M., Heffley, E., 1975. On the independence of the CNV and P300 components of the human averaged evoked potential. Electroencephalogr. Clin. Neurophysiol. 38, 449–461.

Donchin, E., Karis, D., Bashore, T.R., Coles, M.G.H., Gratton, G., 1986. Cognitive psychophysiology and human information processing. In: Coles, M.G.H., Donchin, E., Porges, S.W. (Eds.), Psychophysiology: Systems Processes and Applications. Guilford Press, New York, pp. 244–267.

Falkenstein, M., Hohsbein, J., Hoorman, J., 1993. Late visual and auditory ERP components and choice reaction time. Biol. Psychol. 35, 201–224.

Greenhouse, S.W., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrica 24, 95–112.

Grune, K., Mecklinger, A., Ullsperger, P., 1993. Mental comparison: P300 component of the ERP reflects the symbolic distance effect. Neuroreport: Int. J. Rapid Commun. Res. Neuroscience 4, 1272–1274.

Harter, M.R., Aine, C.J., 1984. Brain mechanisms of visual selective attention. In: Parasuraman, R., Davies, D.R. (Eds.), Varieties of Attention. Academic Press, London.

Hillyard, S.A., Squires, K.C., Baur, J.W., Lindsay, P.H., 1971. Evoked potential correlates of auditory signal detection. Science 172, 1357–1360.

Jasper, H.H., 1958. Report of the committee on methods of clinical examination in electroencephalography. Electroencephalogr. Clin. Neurophysiol. 10, 370–375.

Johnson, R. Jr., 1986. A triarchic model of P300 amplitude. Psychophysiology 23, 367–384.

Johnson, R. Jr., 1993. On the neural generators of the P300 component of the event-related potential. Psychophysiology 30, 90–97.

Jones, L.K., 1989. Measuring a three-dimensional construct of career indecision among college students: a revision of the vocational decision scale—the career decision profile. J. Couns. Psychol. 36, 477–486.

Katz, M., 1963. Decisions and Values: a rationale for secondary school guidance. College Entrance Examination Board, New York.

Katz, M., 1987. Theory and practice: The rationale for a career guidance workbook. The Career Dev. Q. 36, 31–44.

Keppel, G., 1991. Design and Analysis: A Researchers Handbook. (3rd Ed.) Prentice-Hall, Englewood Cliffs, NJ.

Klorman, R., Brumaghim, J., Salzman, L.F., Strauss, J., Borgstedt, A.D., McBride, M.C., Loeb, S., 1990. Effects of methylphenidate on processing negativities in patients with attention-deficit hyperactivity disorder. Psychophysiology 27, 328–337.

Knowdell, R.L., 1993. Career Values Card Sort, California: Career Research & Testing.

Lindholm, E., Koriath, J.J., 1985. Analysis of multiple event related potential components in a tone discrimination task. Int. J. Psychophysiol. 3, 121–129.

Link, S., 1990. Modeling imageless thought: the relative judgement theory of numerical comparisons. J. Math. Psychol. 34, 2–41.

McCarthy, G., Wood, C.C., 1985. Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. Electroencephalogr. Clin. Neurophysiol. 62, 203–208.

McGhie, A., Chapman, J.S., 1961. Disorders of attention and perception in early schizophrenia. Br. J. Med. Psychol. 34, 102–116.

Mecklinger, A., Ullsperger, P., Molle, M., Grune, K., 1994. Event-related potentials indicate information extraction in a comparative judgement task. Psychophysioloy 31, 23–28.

Moyer, R.S., Landauer, T.K., 1967. The time required for judgements of numerical inequality. Nature 215, 1519–1520.

Moyer, R.S., Landauer, T.K., 1973. Determinants of reaction time for digit inequality judgements. Bull. Psychon. Soc. 1, 167–168.

Mulaik, S.A., 1972. Foundations of Factor Analysis. McGraw-Hill, New York.

Naatanen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. Psychophysiology 24, 375–425.

Osipow, S.H., 1980. Career Decision Scale Manual. Odessa, FL: Psychological Assessment Resources.

Osipow, S.H., 1991. Response to Vondracek, Dorn, and Hackett. J. Couns. Dev. 70, 332–333.

Osipow, S.H., Carney, C.G., Barack, A., 1976. A scale of educational-vocational undecidedness: A typological approach. J. Vocat. Behav. 9, 233–243.

Palmer, B., Nasman, V.T., Wilson, G.F., 1994. Task decision difficulty: effects on ERPs in a same-different letter classification task. Biol. Psychol. 38, 199–214.

Paivio, A., 1975. Perceptual comparison through the mind's eye. Mem. Cognit. 3, 635–647.

Polich, J., 1987. Task difficulty, probability and interstimulus interval as determinants of P300 from auditory stimuli. Electroencephalogr. Clin. Neurophysiol. 68, 311–320.

Polich, J., Kok, A., 1995. Cognitive and biological determinants of P300: an integrative review. Biol. Psychol. 41, 103–146.

Polich, J., Pollock, V.E., Bloom, F.E., 1994. Meta-analysis of P300 amplitude from males at risk for alcoholism. Psychol. Bull. 115, 55–73.

Potts, G.P., 1974. Storing and retrieving information about ordered relationships. J. Exp. Psychol. 103, 431–439.

Regan, D., 1989. Human Brain Electrophysiology: evoked potentials and evoked magnetic fields in science and medicine. Elsevier, New York.

Rojewski, J.W., 1994. Career indecision types for rural adolescents from disadvantaged and nondisadvantaged backgrounds. J. Couns. Psychol. 41, 356–363.

Ruchkin, D.S., Sutton, S., Kietzman, M.G., Silver, K., 1980. Slow wave and P300 in signal detection. Electroencephalogr. Clin. Neurophysiol. 50, 35–47.

Salisbury, D.F., O'Donnell, B.F., McCarley, R.W., Shenton, M.E., Benavage, A., 1994. The N2 event-related potential reflects attention deficit in schizophrenia. Biol. Psychol. 39, 1–13.

Scholz, K.W., Potts, G.R., 1974. Cognitive processing of linear orderings. J. Exp. Psychol. 102, 323–326.

Schwartz, S.H., Bilsky, W., 1990. Toward a universal structure of human values. J. Pers. Soc. Psychol. 53, 550–562.

Slaney, R.B., Palko-Nonemaker, D., Alexander, R., 1981. An investigation of two measures of career indecision. J. Vocat. Behav. 18, 92–103.

Ullsperger, P., Gille, H.-G., 1988. The late positive component of the ERP and adaptation level theory. Biol. Psychol. 26, 299–306.

Vasey, M.W., Thayer, J.F., 1987. The continuing problem of false positives in repeated measures ANOVA in psychophysiology: a multivariate solution. Psychophysiology 24, 479–486.

Vondracek, F.W., 1991. Osipow on the Career Decision Scale: Some comments. J. Couns. Dev. 70, 327.

.