INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Explorations in Memory Psychophysics Using Expanded Judgement

by

Peter H. Aubin

A thesis submitted to
The Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Department of Psychology

Carleton University

Ottawa, Ontario

September, 1998

Copyright

1998, Peter H. Aubin

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-37041-0

The undersigned recommend to the
Faculty of Graduate Studies and Research
Acceptance of the thesis:

"Explorations in Memory Psychophysics Using Expanded Judgement"

submitted by

Peter Aubin

in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Department of Psychology

Chair

Thesis Supervisor

External Examiner

Carleton University
September 28, 1998

Abstract

Memory psychophysics was explored using an expanded judgement version of the method of constant stimuli. In Experiments 1 and 2, observers compared a standard line length (perceptual or remembered) with the average of a population of lines by controlling the number of randomly selected observations. The variability of comparison distributions was manipulated to examine the view that memory judgements are slower and less accurate because of increased representational variability. The expanded-judgement paradigm replicated some of the properties of perceptual and remembered visual extents characterised by the method of constant stimuli. Serial-position effects emerged in memory for discriminability, number of observations, sampling and post-sampling latencies. Decisional processing occurred primarily after sampling as shown by the lack of distance effects during sampling, but by the presence of clear distance effects for post-sampling latencies, number of observations and accuracy.

A dissociation in the trade-off between sampling speed and amount of sampled information suggested that differences between memory and perception may not be solely due to representational variability.

Using an experimenter-controlled task, Experiment 3 disconfirmed diffusion theory and, contrary to interference theory, showed that intervening stimuli that are highly similar to the study stimulus can refresh rather than disrupt its representation.

Findings are related to effects typically found in traditional psychophysical tasks and predicted by evidence accrual theories.

Acknowledgements

I would like to extend my heartfelt appreciation to the many people whose support and guidance have made this dissertation possible.

Thanks to all the members of my examination committee: Bill Petrusic, Tony Marley, Ann Laubstein, Bill Jones and Lew Stelmach, for their helpful insights and challenges.

A special thanks to my advisor, Bill Petrusic, for communicating his enthusiasm for psychophysics, love of cooking, endless patience, unfailing guidance and warm friendship. My achievements would not have been possible without his support. More generally, thank you for “all in the fullness of time”. To Elaine Petrusic for her support, patience and her warm reassuring smile, many thanks.

My gratitude to Bill Jones and Kim Matheson for their support, guidance, insight and the opportunities to further my professional career.

Thank you to all the members of Bill’s lab, for their support, encouragement and insight through the years. Much was shared and much was learned.

My deepest appreciation and love to my wife, Donna Mailloux, for the unfailing presence, patience, understanding and support that gave me strength, and for her confidence and commitment to a vision of a future built-up from the present. Finally, but very importantly, a very special and warm thank you (and hugs) to my two daughters, Mayusha and Megan, for their warmth, love and insight, but most of all for giving me the opportunity and awesome privilege to be a dad to two such wonderful persons.

My thanks to the Natural Sciences and Engineering Research Council of Canada without whose funding this research would not have taken place.

Table of Contents

	Page
List of tables	8
List of figures.....	10
Overview.....	14
Introduction.....	17
Expanded Judgement Paradigms.....	19
Experimenter-controlled.....	19
Subject-controlled.....	21
Replications and Extensions.....	22
Models of the Comparison Process.....	26
Single-observation (Strength) Models.....	26
Multiple-observation Models.....	27
Fixed-sample models.....	27
Variable-sample Stochastic Counter Models.....	29
Variable-sample Continuous Accumulators.....	33
Memory Psychophysics.....	35
Properties of Symbolically Represented Simple Visual Stimuli.....	35
Paired-associate Training.....	35
The Underlying Nature of Memory Representations.....	36
Classic Method of Constant Stimuli with Remembered Standards.....	39
Chronometric Properties of Remembered Stimuli.....	47
Similarity Scaling in Memory Psychophysics.....	55

	Page
Experiment 1.....	57
Introductory Comments.....	57
Thurstonian-based Expanded Judgement Model 1.....	59
Memory and the Effect of Variability on n	61
Expected Value as the Basis for Sampling.....	62
Thurstonian-based Expanded Judgement Model 2 for Perception.....	70
Explicit Sampling Based on Evidence Accrual.....	74
Expanded Judgement Using Ordinal Semantic Codes.....	77
Method.....	82
Results.....	96
Discussion.....	113
Experiment 2.....	116
Method.....	116
Results.....	116
Discussion.....	141
Experiment 3.....	146
Introduction.....	146
Diffusion Theory Versus Interference Theory.....	146
Diffusion-based Expanded Judgement Model.....	148
The Effect of Diffusion and of Comparison Variability on Discriminability...	150
Variance estimate for the memory condition.....	154
Variance estimate for the perceptual condition.....	156

	Page
Diffusion estimate using pair comparisons.....	159
Method.....	160
Results.....	168
Discussion.....	181
References.....	187
Endnotes.....	204
Appendix.....	214

Table	Description	Page
VI	Diffusion Parameter Estimates for Each Subject	169
VII	According to the Decay Model of Forgetting: the Amount of Stored Stimulus Information (α), the Decay Rate (ϕ) and the Corresponding Diffusion Rate $\left(1/\phi\right)$ for Each Subject	173

List of Figures

Figure	Description	Page
I	Effect of sample size (n) on discriminability (D) for perceptual and for memory standards under low and high comparison variances for small and large standard-comparison differences as predicted by the Thurstonian-based expanded-judgement model 1.	63
II	Effect of sample size on expected value for perceptual and for memory standards under low and high comparison variances for small and large standard-comparison differences as predicted by the Thurstonian-based expanded-judgement model 1 incorporating expected value.	67
III	Effect of sample size (n) on expected value for perceptual and for memory standards under low and high comparison variances for small and large standard-comparison differences as predicted by the Thurstonian-based expanded-judgement model 1 incorporating expected value.	68

Figure	Description	Page
IV	Effect of sample size (n) on discriminability separately for perceptual standards and for memory standards under low and high comparison variances for small and large standard-comparison differences as predicted by the Thurstonian-based expanded-judgement model 2.	71
V	Effect of sample size (n) on expected value for perceptual and for memory standards under low and high comparison variances for small and large standard-comparison differences as predicted by the Thurstonian-based expanded-judgement model 2 incorporating expected value.	72
VI	Discriminability of standards (Weber fractions) for each of the four standard type by comparison variability combinations.	101
VII	Standardised differences between PSEs and physical standards for each subject under perceptual and remembered judgements.	104
VIII	Distance effects for accuracy, number of observations and the last sampling time across comparison-standard differences represented by effect coding.	111
IX	Distance effects for post-summative latencies for each of the four standard type by comparison variability combinations.	112

Figure	Description	Page
X	Discriminability of perceptual and remembered standards under low comparison variability and under high comparison variability.	119
XI	Standardised differences between PSEs and physical standards for each subject under perceptual and remembered judgements.	120
XII	Serial position effects for the first sampling time, the last sampling time, mean sampling times and post summative latencies.	134
XIII	Distance effects on accuracy for memory and perception and for low and high comparison variability.	136
XIV	Distance effects on the number of observations for memory and perception.	139
XV	Distance effects on overall sampling times and post-summative latencies.	140
XVI	Effect of sample size (n) on total variances of the differences and on separate variances for perception and for different levels of diffusion under low and high comparison variances.	155
XVII	Effect of sample size (n) on discriminability for perception and for different levels of diffusion under low and high comparison variances.	157

Figure	Description	Page
XVIII	The effect of ISI on accuracy and response times for each of the three observers.	170
XIX	Variance of the study stimulus as ISI increases for observers 1 and 2.	171
XX	Observed and predicted variances of the study stimulus for subject 1.	176
XXI	Observed and predicted variances of the study stimulus for subject 2.	177
XXII	The effect of sample size on accuracy and post-summative latencies for perceptual under low comparison variability.	179
XXIII	The effect of sample size on accuracy and post-summative latencies for perceptual under high comparison variability.	180

Overview

The purpose of this dissertation was to explore the decisional processes involved in the binary comparisons of remembered stimuli, and to characterise the properties of their underlying representations using an expanded judgement paradigm. A traditional psychophysical binary judgement requires a comparison between two stimuli. An expanded binary judgement requires a comparison between a stimulus and the average of a sequence of stimuli. It was hoped that the expanded judgement paradigm would shed light on the nature of the underlying decisional processes as explained by evidence accrual theories.

Evidence accrual models describe the comparison of two line lengths as a process involving the accumulation of implicit pieces of information each of which is evaluated and classified with respect to the two possible response alternatives. The accumulation of evidence stops and an overt response is rendered when a criterion amount of evidence is attained. The number of accruals (decision time) increases with the variability of the evidence and with task difficulty. Judgements with remembered stimuli should require more accruals than judgements with perceptual standards because the representation of the former is thought to be more variable than that of the latter. The number of accruals will also increase with demands for accuracy and decrease with demands for speed.

Despite the success of evidence accrual models, the implied nature of these accruals presents a challenge. The literature review of this dissertation begins by examining the 'expanded judgement' paradigm here used as an *analogy* for the implicit evidence accrual process. In the subject-controlled case, subjects control the number and presentation rate of observations. The number of explicit observations selected is taken to reflect the

underlying nature of the decisional process and is shown to be sensitive to difficulty and variability manipulations. As a result, the manifest character of observations in expanded judgement may provide insight into how the decisional system makes use of implicit accruals. In the experimenter-controlled case, the experimenter controls the number and presentation rate of the observations. In this case, decisional latencies that follow the observation sequence are thought to reflect the end-product of an implicit accrual process that is sensitive to the properties of the explicitly presented observations.

The section on expanded judgement is followed by a review of the 'Models of the Comparison Process'. This historical account describes the evolution of theories, from single-sample to evidence accrual models that have attempted to explain the wide range of response time and response probability patterns that underlie comparative decisions.

The application of expanded-judgement to memory psychophysics requires a review of the 'Properties of Symbolically Represented Simple Visual Stimuli'. This section describes the key patterns of response time and response accuracy that characterise perception and memory. The viability of the expanded judgement paradigm was then examined for convergence with these findings. After describing paired-associate training used for creating symbolically represented physical stimuli (used in Experiments 1 and 2), two rival positions on the form of the underlying representation of remembered stimuli, the analogue view and semantic coding theory, are then briefly described. The sub-sections on the 'Classic Method of Constant Stimuli with Remembered Standards' and on the 'Chronometric Properties of Remembered Stimuli' provide a characterisation of the main patterns of response time and response accuracy that differentiate between perception and memory. Support for the analogue view provided by traditional binary

comparison paradigms is strongly supplemented by research in similarity judgements which is then briefly described. Although binary judgements provided evidence that memory properties are more precisely represented than on an ordinal scale, similarity judgements permit the explicit axiomatic demonstration of an interval level of measurement. An analogue representation is also assumed to underlie expanded judgements with remembered stimuli.

The expanded judgement paradigm is therefore developed against a backdrop of evidence accrual models that have provided a fruitful account of psychophysical comparisons. *Experiment 1* explored the basic properties of memory representations using a subject-controlled paradigm and attempted to predict optimal sample size on the basis of two Thurstonian-based expanded judgement models. Although basic psychophysical effects were replicated, some anomalous findings also emerged. *Experiment 2* was exactly the same as Experiment 1 except that real monetary costs were associated with correct and incorrect responses to ensure that observers were taking the task seriously. *Experiment 3* was designed to further characterise how decisional processes are affected by variability, a key concept thought to differentiate between perception and memory. An expanded judgement version of diffusion theory was used to predict optimal decisional sample size (decision time) under conditions where a study stimulus is presented for a specified period of time (memory condition), and then followed by an experimenter-controlled sequence of comparison observations. Such a paradigm permitted the investigation of the effects of forgetting (diffusion) on the judgmental processing.

Introduction

The first century of psychophysics assumed that comparative judgements were based on a single pair of observations drawn from a distribution of subjective differences (Cattell, 1893; Fechner, 1860; Tanner & Swets, 1954; Thurstone, 1927). Although theories based on this predominant view characterised many of the properties of comparative judgements, the complexity of the response time and response probability relationship suggested the existence of an implicit sampling process (for a review see Vickers, 1980). Thomson (1920) provided the first conceptualisation of psychophysical comparisons based on the accumulation of implicit evidence. His article in *Psychological Review* depicted an urn whose proportion of white balls (observations favouring one response) and of black balls (observations favouring the other response) were linearly related to stimulus information. In order to make a judgement, the subject draws a sample of implicit observations. The proportion of observations favouring one stimulus determines the response. Thomson (1920) stated:

On this point of view, the standard, the variable stimulus, and the physiological make-up of a subject decide the proportion of black balls in the urn, but the decision as to what proportion in the sample is to be called heavier, what undecided and what lighter depends upon a conscious act of the subject, and can be varied, if he be so disposed, at his whim; and will vary with his mood at the moment.

Thus, Thomson (1920) promoted the view that stimulus information, sensitivity and bias determine the number of observations required for the overt decision. In the final paragraph of his article, Thomson states that 'the difference between this point of view and the older one may prove to be academic only, for the disentangling of the factors

which are here distinguished may in practice be impossible." Little did Thomson know that his unassuming creative moment, outlined quite simply in a few pages, would one day form the basis of a radically new way of conceptualising decision-making.

Expanded Judgement Paradigms

The properties of comparative judgements have been most successfully explained by variable-sample models that describe decision time as comprised of a constant amount of encoding and motor time and a variable amount of decision time (accruals) (Audley & Pike, 1965; Laming, 1968; Link & Heath, 1975; Petrusic, 1992; Pike, 1968, 1971; Smith & Vickers, 1988; Stone, 1960; Vickers, 1970, 1979). Despite differing in their stopping-rules, these models posit the existence of information entities that are essentially implicit and therefore, not amenable to direct manipulation. More recently, the expanded judgement paradigm was developed as an analogy to the implicit accrual process (Irwin & Smith, 1956; Irwin, Smith & Mayfield, 1956; Vickers, Burt, Smith & Brown, 1985; Vickers, Smith, Burt & Brown, 1985). In this paradigm, the subject is required to make decisions using sequences of explicitly presented physical stimuli. The implicit sampling process can therefore be examined explicitly by manipulating the amount, variability and discriminability of the distributions of physical information a subject receives. This dissertation proposes to examine the viability of the expanded judgement paradigm as a means of characterising the decisional processes involved in comparative judgements of visual extent as remembered and as perceived.

Experimenter-controlled

The expanded judgement (E-J) paradigm was originally designed by Irwin, Smith, & Mayfield (1956) in an attempt to characterise the role of confidence in decision-making. Cartwright & Festinger's (1943) opponent process theory suggested that, in the case of a two-alternative forced choice (2AFC) task, discrimination is determined by

competition between two response tendencies whose difference is normally distributed over time and determined by the discriminability of the stimuli. These response forces are compared to a criterion difference or a restraint force. The decisional system takes successive samples of response and restraint forces until the difference between momentary response forces exceeds the momentary restraint force which itself varies randomly in a Gaussian manner. This model has also been labelled as the 'optional stopping' model (Swets & Green 1961; Vickers, 1980).

Cartwright & Festinger's (1943) model suggests that, given a fixed criterion location, increased stimulus variability should result in a larger proportion of stimulus values that exceed the criterion. The increased frequency of large stimulus values should therefore result in faster response times and greater confidence compared to stimulus distributions with smaller variances because with high variability, the criterion should be easier to exceed. In contrast, Irwin et al. (1956) predicted that if comparative judgements are akin to statistical decision-making, then confidence should decrease as stimulus variability increases. By analogy to the effect of sample size on confidence intervals, the accumulation of information should improve the quality of the decision and increase confidence levels.

Irwin et al. (1956) presented the S with a series of cards from a deck in an experimenter-controlled (ECON) task. Each card had either a positive or a negative number. The S's task was to decide whether the mean of the entire deck was greater than or less than 0. The subject provided a confidence rating after each trial using a 100-point scale. The experimenter controlled the number of cards presented (duration of the decision process), their variability (response force) and the mean of the deck

(discriminability). Experiment 1 completely disconfirmed Cartwright and Festinger (1943) and provided clear support for Irwin et al. 's (1956) view that confidence increases as a function of both the number of cards presented (decision time) and the ease of the decision. Confidence decreased as variance increased. Error rates were inversely related to confidence, the number of cards presented, the ease of discrimination, but directly to variability.

These results were replicated in Experiment 2 which made use of a 2AFC task. In this experiment, the S had to decide which of two decks of numbers had the higher mean. Each time a pair of cards was presented, the S made a judgement about which pack had the greater mean and also provided a confidence rating. The overall findings for the two experiments were similar.

Subject-controlled

Irwin & Smith (1956) provided an alternative E-J paradigm that leaves the subject free to inspect as many cards as desired. This design parallels the traditional subject-controlled (SCON) psychophysical paradigm which permits the S to inspect stimuli as long as desired before making a response. In the E-J paradigm the number of cards examined is assumed to parallel decision time. In comparison to the ECON paradigm, the SCON judgements results in different predictions.

Both Cartwright & Festinger (1943) and Irwin & Smith (1956) would predict that decision time should decrease as the difference between the mean of the deck and 0 increases. Opponent process suggested that the difference in response tendency should have a greater chance of exceeding the criterion sooner when discriminability is

increased. Irwin et al. (1956) predicted that easier decisions require less information and therefore less time. “Optional-stopping” (Cartwright & Festinger, 1943) predicted that decision time should vary inversely with the variability of the deck because increased variability in response tendencies should result in increased chances that the response criterion would be exceeded sooner. By analogy to statistical decision-making, Irwin et al. (1956) predicted that decision time would increase with increased variability because of the greater overlap between distributions.

Results clearly argued for Irwin et al.’s (1956) statistical sampling process. The number of cards required to make a decision was directly related to the magnitude of the difference between the mean of the pack and zero. More cards were required as discriminability decreased and as the variability of the numbers increased contrary to Cartwright and Festinger’s (1943) prediction. In addition, highly significant individual differences emerged in the number of cards requested ranging from 7.6 to 107.7. Unfortunately, confidence ratings were not required in this experiment thus precluding a test of Irwin et al.’s (1956) view that large individual differences in the amount of confidence probably underlie individual differences in the amount of information required for a decision.

Replications and Extensions

Expanded judgement (E-J) exemplified by the subject-controlled (SCON) and the experimenter-controlled (ECON) paradigms were more fully developed in subsequent work by Vickers and associates (Vickers, Burt, et al., 1985; Vickers, Smith, et al., 1985). The ‘implicit’ sampling process is made explicit by presenting the observer with a

sequence of sample observations, x_i , from a well-defined probability distribution, in this case Gaussian. The sampling process can therefore be manipulated by changing the parameters of the distribution, including discriminability (μ) and variability (σ), the number of observations (n_i), the duration of each observation and their rate of presentation. In the case of the SCON paradigm the number of observations and their rate of presentation are random variables. In the case of the ECON paradigm, all the above factors are under experimental control except for the time between the last observation and the primary response, also referred to as post-summative latencies.

Vickers, Burt et al. (1985) and Vickers, Smith et al. (1985) exerted more precise control over information than did the original work of Irwin et al. (1956) and Irwin & Smith (1956) by using computer-generated visual stimuli. On each trial, a series of white horizontal lines were successively presented on either side of a vertical reference line at a rate of 3 per second. The vertical line was used as a reference line representing the value '0' on an imaginary scale of standardised scores. The inner end of each stimulus line touched the vertical reference line and extended outward. The O was required to decide whether the true average of the distribution from which the sequence was randomly drawn fell to the right or the left of the vertical line. The four possible distributions ($\mu_1 = \pm 2, \mu_2 \pm 4$) had a common standard deviation of $\sigma = 6$ permitting two levels of difficulty represented by the coefficient of determination (μ/σ) whose absolute value was either .33 or .66). Subjects also indicated how confident they were in their response by selecting a number between 1 and 5. The ECON and SCON paradigms were employed.

In the ECON paradigm, an observation sequence consisted in 6, 9, 12, 15 or 18 lines. Subjects were instructed to respond as soon as they could after the series was terminated. In the SCON paradigm, subjects were instructed to respond as soon as possible after initiating the sampling process. The main findings were as follows. In both the ECON and SCON paradigms, overall accuracy was greater in the high discriminability condition. In the ECON paradigm, the percentage of correct responses increased with the number of observations. This effect was much more pronounced for the difficult condition than in the easy condition. The effect of sample size was reversed in the SCON paradigm. The percentage of correct responses decreased as sample size increased, an effect which was also exacerbated by difficulty. An analysis of latencies between the presentation of the last line and the subjects response (post-summative latencies) in the ECON paradigm indicated faster responding when discriminability was high. An inverse relationship between response time and the number of observations was found. In both the ECON and SCON paradigm, RTs for incorrect responses were longer than RTs for correct responses especially when discriminability was high. The discrepancies between RTs for correct and incorrect responses was greater for the slower subjects.

In general, Vicker's (1970, 1979) continuous accumulator model and Petrusic's (1992) discrete accumulator, slow-fast guessing theory, both provide a good account of these findings. We begin with a description of theories that can be applied to comparative judgements for simple perceptual and remembered stimuli. These models can be divided into two broad classes. Strength Theory emphasises the underlying distributional properties of internal representations and the locations of response criteria. Stochastic Counter Models make minimal assumptions concerning the underlying representations

but rather, relate variability in decision time to the amount of covert predecisional evidence accumulated. Models are considered with respect to their ability to account for patterns of response time and response probability (accuracy).

Models of the Comparison Process

Single-observation (Strength) Models

In order to account for imperfect judgement accuracy, strength theories assume that judgements are based on sensory representations built-up from distributions of subjective values (Guilford, 1954). Each distribution is characterised by an expectation and a variance. Although variability results from several sources (Barlow, 1956; Cattell, 1893; Jastrow, 1888; Welford, 1965) most agree that its shape can be reasonably described as Gaussian, even if for no more than mathematical convenience (Macmillan & Creelman, 1991; Vickers, 1980). Following the phi-gamma hypothesis (Urban, 1910), theories such as the law of comparative judgement (Thurstone, 1927) advanced the view that a comparative judgement is based on the difference between two observations drawn from a normal distribution of sensory differences. Consider the method of constant stimuli where V is the variable stimulus and S is the standard. A positive $V-S$ difference results in the response, 'Variable is greater than the Standard'. A negative $V-S$ difference results in the response, 'Variable is less than the Standard'. The distribution of discriminial dispersions is thus partitioned into negative and positive differences using a criterion whose value, $V - S = 0$, represents the complete lack of discrimination. According to this view, the expectations of the discriminial dispersions, $\bar{V} - \bar{S}$ are linearly related to the difference between the objective stimuli, $v-s$. Choice inconsistency and response errors occur as a result of variability in the internal stimulus representations such that the sign of the sensory difference sampled on a given trial may not correspond to the sign of the difference between the objective stimuli. In the long run however, $\bar{V} - \bar{S}$ will reflect $v-s$.

Signal detection theory (Tanner and Swets, 1954) propelled psychophysics into a new era by developing a theoretical basis for the distinction between sensitivity to stimulus information and the propensity to select a certain response (Murray, 1993). The superimposition of a decisional space over separate distributions for each stimulus situation permitted the use of hit and false alarm rates to determine independent indices for sensitivity and bias. Bias can be independently manipulated by changing base rates and/or payoffs. Sensitivity can be independently manipulated by changing the stimulus properties (Macmillan & Creelman, 1991).

Models that purport to explain judgmental behaviour must explain the behaviour of decision time and decision accuracy and, more specifically how they relate to each other (Audley & Pike, 1965). Despite SDT's success, its close ties to an underlying single-stage stimulus process prevented it from accounting for changes in accuracy that result from manipulations of the relative instructional demands for accuracy and speed (Festinger, 1943a, 1943b; Garrett, 1922; Vickers, 1980). Trade-offs between accuracy and speed are inadmissible by theories that base accuracy uniquely on discriminial stimulus differences (Fechner, 1960; Thurstone, 1927), and on a constant sampling time (Tanner & Swets, 1954). Even additional assumptions incorporating response time functions (Pike, 1973) left strength theories lacking as general models of the comparison process.¹

Multiple-observation Models

Fixed-sample models. These models are based on a suggestion by Thomson (1920) that the sensory representation of a stimulus is built up from an accumulation of

successive implicit observations within a trial, rather than from trial to trial (Fechner, 1860; Thurstone, 1927). Each observation takes a constant amount of sampling time.

This idea was applied to the classical model of signal detection primarily as an attempt to account for the effects of speed-accuracy trade-offs on accuracy and response time (Crossman, 1955; Swets, Shipley, McKey, & Green, 1959). For example, Crossman (1955) suggested that decisions are based on the average of multiple observations. Since observations are sampled at a constant rate, accuracy and RTs will increase with the number of observations taken. Speed-accuracy trade-offs are explained by the subject's ability to select a higher preset number of observations when accuracy is emphasised and a lower preset number when speed is emphasised, while maintaining a reasonable accuracy level. Error rates follow as a natural consequence of the effect of sample size on the probability of being correct.

This class of models can explain the inverse relationship between response time and accuracy when difficulty is manipulated between blocks. The number of observations on each trial is determined at the outset of a block and is maintained for the entire block in order to minimise error rates. A relatively small number of observations, and therefore less time, is required for easy discriminations. The number of observations is increased for a block of harder discriminations to maintain a low error rate. Longer response times for more difficulty judgements is therefore explained by an increase in the number of observations. These models cannot explain variations in difficulty within blocks where the difficulty of a stimulus pair on any given trial varies unpredictably (Henmon, 1906; Johnson, 1939; Kellogg, 1931; Festinger, 1943a, 1943b; Petrusic, 1989).

An alternative to these fixed-sample models are the ‘optional’ stopping models such as Cartwright & Festinger’s (1943) model described above. The observer adopts two criteria with a zone of no-decision between the two. A response is made when a response force exceeds either criterion. Response forces falling within the two criteria lead to a continuation of the sampling process. The probability of the response force exceeding either criteria is determined by the expectation of the difference between the two sensory distributions. The more similar the stimuli, the smaller the expected difference and therefore the less likely that a momentary difference will exceed either criterion. More samples and therefore increased response times will be evident on more difficult trials. Increased response times which result from an emphasis on accuracy can be explained by criteria that are more difficult to exceed. In the case of instructional emphasis on speed, the criteria are closer together. The effects of bias can be accounted for by the asymmetry of the criteria. Optional-stopping models predict that response times for correct and incorrect responses will be equal (Broadbent, 1971, p.291). Research has shown however, that response times for errors are longer than response times for correct responses when accuracy is emphasised. Emphasis on speed attenuates this difference and under extreme speed results in a reversal (Kellogg, 1931; Pike, 1971; Petrusic, 1992; Swensson, 1972).

The complex interaction between, on the one hand, correct and incorrect response times, and on the other, instructional stress for speed or accuracy resulted in the development of the more flexible stochastic counter models.

Variable-sample Stochastic Counter Models. Stochastic counter models make few assumptions regarding the form of the subjective stimulus representation (Audley & Pike, 1965; Pike, 1968). According to this view, the decisional system is characterised by the

accumulation of information that is classified as evidence for each of the possible alternatives, and by a criterion that serves as a stopping-rule for terminating the accrual process and for generating an overt response (Petrusic & Jamieson, 1978; Vickers, 1988). Each observation is assumed to take an equal amount of sampling time such that decision time is determined by the number of accruals in addition to a constant amount of motor and output time. In a 2AFC task, accrual models assume predecisional events concerning stimuli a and b. The covert response $a|b$ occurs if the evidence on a particular accrual favours the overt response $a|b$ i.e. a preferred over b. The covert response $b|a$ occurs if the evidence on a particular accrual favours the over response $b|a$. Slow-fast guessing theory (Petrusic, 1992), as discussed below, is unique in that it admits a third accumulator for non-diagnostic evidence. Accruals are accumulated until a criterion is reached. Stochastic counter models differ in their stopping-rule and consequently in how the accruals lead to an overt decision.

1. The ***Runs Model*** (Audley, 1960; Audley & Pike, 1965; Estes, 1960) has two subsystems that accumulate evidence until a criterion number, k , of uninterrupted accruals for one of the alternatives is attained. This model involves partial memory in that the occurrence of the implicit response $a|b$ erases all $b|a$ responses accrued in the other accumulator. Implicit responses $a|b$ continue to accrue until either the criterion, k_A , is reached at which point the overt response $a|b$ is made, or an implicit observation of the form $b|a$ is accrued. If the implicit response $b|a$ occurs, then all accruals of the form $a|b$ are erased. Implicit responses $b|a$ would then continue to accrue until either the criterion, k_B , is reached at which point the overt response $b|a$ is made, or an implicit observation of the form $a|b$ is accrued. When criteria are equal and low, error times are longer than

correct times for the same stimuli. When criteria are equal and high then error times and correct times are equal. These relations are opposite to those observed in empirical data. Error times are faster than correct times under speed where criteria are equal and low. Conversely, error times are longer than correct when criteria are equal and high as is the case under an accuracy set (Petrusic, 1992; Petrusic & Jamieson, 1978. See Vickers, 1980 for a review). More importantly, the runs model predicts a nonmonotone latency-probability function (LPF), a pattern observed only when decisions are rendered under speed (Petrusic, 1992; Vickers, 1980).

2. The *Difference Model* is similar to the runs model in that the two accumulators are not independent. In this case, however, both accumulators accrue implicit responses in tandem. A decision is rendered when a criterion difference between the two accumulator totals is attained. As Petrusic and Jamieson (1978) note, the overt response aPb is made whenever the number of implicit apb responses exceeds the number of bpa responses by criterion amount α . The overt response bPa is made whenever the number of implicit bpa responses exceeds the number of apb responses by criterion amount β . The difference model can be thought of as a normative model of choice behaviour because it results in the steepest increase in performance accuracy for a given number of accruals. In other words, this model predicts that less information (time) will be needed to reach a given level of performance accuracy than that predicted by any other stochastic counter model (Petrusic & Jamieson, 1978). Like the runs model, the difference model predicts a nonmonotone LPF but equal RTs for correct and incorrect responses.

3. The *Recruitment Model* (LaBerge, 1962) also has two accumulators, one for each overt response, RA and RB. Each sub-system independently accumulates implicit

observations either of the form aPb as evidence in favour of the overt response aPb , or of the form bPa in favour of the overt response bPa . The decisional process can be seen as a horse race between accumulators in that the accumulator that first exceeds its criterion generates an overt response. This model predicts a strictly monotone decreasing LP but cannot admit changes in the form of the LPF with practice (Petrusic & Jamieson, 1978). As a result, RTs for incorrect responses will always be longer than for correct responses. Note that both strength theory and the recruitment model predict the same form of LPF.

4. The Slow-fast Guessing Evidence Accrual Model. The slow-fast guessing evidence accrual model of binary comparative judgements (see Petrusic, 1992 for a full mathematical development) was originally developed by Petrusic and Jamieson (1978) in the context of preferential choice and later successfully applied to perceptual discriminations (Petrusic & Baranski, 1989). Petrusic and associates modified the recruitment model (LaBerge, 1962) by permitting an implicit response for indecision. If evidence on an accrual favours neither a nor b then it is said to favour indifference, aIb . The model has three criteria, α_1 leading to the state aPb , α_2 leading to the state bPa , and α_3 leading to the state aIb . When the overt indifference response is not permissible as is the case in a 2AFC task, the indifference state, aIb , leads to guessing aPb with probability g , and to guessing bPa with probability $1-g$. Guessing out of the state of indifference allows the accumulator model to account for the different shapes of the LPFs. When α_3 exceeds both α_1 and α_2 , then guessing is not likely and the model predicts a monotone decreasing LPF paralleling those found in experiments where accuracy is emphasised. As the guessing parameter, α_3 , becomes more conservative and eventually falls below α_1 and α_2 , the LPF flattens out and eventually forms an inverted U. Petrusic

& Jamieson (1978) examined how response time- response frequency relations change with practice by using latency-frequency curves (LFC). They found that the shape of the LFCs remained monotone across replications when choices were important. Presumably in the case of personally important choices, α_3 is liberal reflecting subjects reluctance to guess. Conversely, in the case of unimportant choices, LFCs became nonmonotone over replications suggesting an increased propensity to guess. LPFs exhibit the same patterns in the context of perceptual discriminations. Petrusic (1992) showed that slow-fast guessing theory can successfully predict that under an accuracy set, RTs for errors are slower than RTs for correct responses. An increasing emphasis on speed, results in a decrease in the discrepancy between response times for correct and incorrect judgements. With a sufficient emphasis on speed, RTs for errors times become faster than RTs for correct decisions.

Variable-sample Continuous Accumulators.

1. The Continuous Accumulator (Vickers, 1970, 1979) is a generalisation of stochastic counter models in that accruals carry continuous evidence. Similar to Laberge's (1962) recruitment model, the continuous accumulator has two accrual criteria, α and β , and two accrual totals, T_A and T_B , associated with each overt response aPb and bPa , respectively. Both models accrue pieces of evidence, but the accumulator also registers the magnitude of the stimulus information associated with each observation. The accumulator can assume normal underlying distributions although this is not a necessary assumption (Vickers & Smith, 1985). Magnitudes are obtained by computing the difference between the stimulus value, Z_n and a sensory referent, c i.e. $Z_n - c$. The sign of the difference determines the momentary pre-criterion state of the decisional system.

When $Z_n > c$, the momentary covert response is apb and the deviation is added to T_A .

When $Z_n < c$, the covert response is bpa and the deviation is added to T_B . Totals T_A and T_B behave as two independent random walks (Vickers & Smith, 1985). The response aPb is made if T_A reaches or exceeds the response criterion α . The response bPa is made if T_B reaches or exceeds the response criterion β . The accumulator successfully predicts shifts in the form of the LPF from monotone decreasing to nonmonotone as the response criteria decrease.

2. The General Random Walk Process (Link, 1975; Link & Heath, 1975) for a 2AFC task involves the accumulation of signed differences over multiple implicit observations. On a given trial, the decisional system samples pairs of implicit observations over n steps, computes the difference between their momentary values and adds these signed differences to a single total, $T(n)$. For a given pair of observations where $s(a)$ and $s(b)$ are the momentary subjective values for physical stimuli a and b ; if $s(a) - s(b) > 0$, then the increment will be positive corresponding to the response aRb i.e. a is greater than b . If $s(a) - s(b) < 0$, then the increment will be negative corresponding to the response bRa i.e. b is greater than a . These momentary differences are actually compared to a referent or indifference point which can be assumed to equal zero (Vickers & Smith, 1985). Corresponding to each overt response is a signed criterion difference such that aRb will occur if $T(n)$, the signed total after n steps, exceeds criterion K_A . The response bRa will occur if $T(n)$ exceeds $-K_B$. Link's model can only admit changes in the LPF from monotone to nonmonotone if the underlying probability distribution becomes asymmetric (Petrusic & Jamieson, 1978).

Memory Psychophysics

Properties of Symbolically Represented Simple Visual Stimuli

Paired-associate Training. Shepard and Chipman (1968) used similarity scaling of US states immediately available on a map (perception) and of states available only in memory to show that size relations in perception and in memory were closely related to each other and to the physical relations among the geographic shapes. In so doing, Shepard and Chipman (1968) promoted a second-order isomorphism emphasising the importance of the relation between, on the one hand, relations among physical objects and, on the other, relations among their corresponding internal representations. This principle stands in contrast to that of first-order isomorphism that suggests a direct relation between a objective and subjective stimuli. Such a pioneering view foresaw a richer characterisation of the properties of perception and memory that would ensue through the use of the classical psychophysical measures of response probability and response time.

In memory psychophysics, the definition of a stimulus in memory is well constrained to ensure its clear operationalisation. Paired-associate training is used to associated a different label with each stimulus (Hilgard, 1951). Labels are typically words, in most cases known as trigrams or nonsense syllables (CVCs: consonant-vowel-consonants) that have no meaning outside of the experiment. CVCs derive their meaning from the stimuli to which they are associated. Tables are available for selecting CVCs that have normed association levels for determining ease of learning (Hilgard, 1951). In some cases letters (Bjorkman et al., 1960) and names have been used (Petrusic, Baranski & Aubin, 1995). As an example, suppose an experiment with five stimulus squares of

different sizes. Each square has a different label associated with it e.g. BIX, TEG, ZOC, LEJ and BOC. The subject might be presented with one of the squares and all of the five CVC's. The subject is required to select the label that corresponds to the presented stimulus. Feedback on correctness of response is provided after each response. A learning criterion defined as a number of successively correct responses is typically required to ensure some operational definition of learning. Alternatively, a set number of presentations for each stimulus could be given without an actual learning criterion. Many memory experiments require an overlearning phase where, after reaching criterion, the subject is given an additional number of learning trials. Paired-associate training has been used in a wide variety of psychophysical paradigms including 'direct' scaling methods (Stevens, 1957), 'indirect' scaling methods (Baranski & Petrusic, 1992) and similarity judgements (Petrusic & Aubin, 1996).

The Underlying Nature of Memory Representations

Semantic Coding Theory. The semantic coding theory (SCT), developed by Banks (1977) and Banks and Flora (1977) in the context of psychophysical comparisons, suggests that although remembered stimuli are initially represented on an analogue scale, ordinal semantic codes form the basis of the comparative judgements. SCT has attempted to use ordinal coding to account for empirical judgement times previously interpreted as evidence of an underlying analogue representation. SCT suggests that the distance effect, the range effect, the semantic congruity effect and the serial-position effect are either due to predecisional or post-decisional processing, or are artefacts of the experimental paradigm.

The first phase of the semantic coding model involves the ‘perception’ and translation of the underlying analogue perceptual information, referred to as the ‘data base’, into ordinal codes. As an example, consider a binary comparison task involving four stimulus squares each associated with a different CVC during paired-associate training. The smallest square might be coded as S+, where 'S' stands for small. The next larger square might be coded as S (small but not as small as S+). The third as L (large). And finally, the largest would be coded as L+ (larger than L). These codes would then be used during the comparison phase. When squares are relatively different in size they are more likely to be encoded into different categories and so decision time is fast because of the lack of ambiguity. As squares become more similar in size, their corresponding semantic codes are more likely to be in the same categories thus precluding discriminability during the ‘discrimination of stimulus codes’ phase. As a result, new semantic codes would have to be generated by the first stage using additional perceptual information drawn from the data base. For example, if two stimuli are encoded as S, then the system attempts to differentiate between the stimuli by retrieving additional information from the data base. One of the stimuli might be recoded as S+. Once the ordinal codes are successfully differentiated, then discrimination automatically follows. Overall response times will therefore be affected by the amount of differentiation necessary. The more similar two stimuli are on the underlying continuum the more differentiation will be needed and the longer the RTs will be, as shown in the distance effect.

Mixed trials consist of comparisons between a perceptual stimulus and a remembered stimulus represented by a CVC. According to SCT, these trials will take

longer than pure memory trials where two CVCs are presented, because the perceptual stimulus must be recoded into an ordinal-based symbolic format that is compatible with the memory stimulus' semantic code. Compared to pure memory trials, judgments on mixed trials will take longer by a constant amount of time due to the translation of the perceptual stimulus into a semantic code (Banks, 1977; Banks and Flora, 1977).

Although designed to deal primarily with processing time, SCT can be extended to apply to response accuracy as well. Baranski and Petrusic (1992) suggested that SCT could, with additional assumptions, generate psychometric functions that are sensitive to stimulus differences. According to this view, the accuracy of code generation is probabilistic such that as stimuli become more similar, the generation of erroneous ordinal codes becomes more probable. A crucial limitation of SCT is that it was not designed to account for the properties of perceptual judgements.

Analogue and Imagery Models. Analogue models suggest that the memory representation of a physical stimulus preserves the metric properties of the perceptual referent and should, as a result, be processed in a similar manner by the decisional system. Remembered stimuli are therefore compared on the basis of analogue information. Evidence from models that emphasise the imagery debate (Farah, 1985; Kosslyn, 1980; Paivio, 1971) and from models that emphasise the existence of an underlying continuum (Baranski & Petrusic, 1992; Jamieson & Petrusic, 1975c; Marks, 1972; Petrusic & Aubin, 1996; Petrusic et al., 1995; Petrusic, Baranski & Aubin, 1995) converge to show that the memory representation is characterised by properties that are perceptual-like. This section focuses on the empirical literature that deals with binary comparisons (Moyer & Bayer, 1976; Petrusic et al., 1995; Petrusic, Baranski & Aubin,

1995), the method of constant stimuli (Baranski & Petrusic, 1992; Harrison & Petrusic, 1994; Petrusic, Baranski & Harrison, 1993), and similarity judgements (Petrusic & Aubin, 1996; Petrusic, 1996). The latter examines the analogue nature of the representation through an axiomatic approach and also examines the behaviour of response time and response probability under conditions that make ordinal coding highly unlikely. According to the analogue view, the distance effect, range effect, semantic congruity effect and serial-position effect occur because of the underlying metric properties of the stimuli.

Classic Method of Constant Stimuli with Remembered Standards

Baranski & Petrusic (1992) were the first to investigate the psychophysical properties of memory representations for visual extent by extending the classical method of constant stimuli (Hegelmaier, 1852) to symbolically represented confusable stimuli. The same three standard line lengths were used in perception and in memory. Each standard was centred in non-overlapping sets of five variable stimuli. Each standard was successively compared to each comparison line within its own set. On half the trials, subjects were required to decide which of two successively presented lines was 'LONGER', and on the other half, which of the two lines was 'SHORTER'. In the perceptual condition, lines were physically presented on the computer screen. In the memory condition, the standard was symbolically represented by a CVC but the comparison line was perceptually available. The memory condition had been preceded by paired-associate training. These authors demonstrated four main properties of memory representations. First, memory representations are more variable than their perceptual

counterparts as shown by elevated Weber fractions in the memory condition. Second, symbolically represented standards, and not perceptual standards, exhibit a serial-position effect characterised by inflated Weber fractions for the middle standards relative to the outer standards. Third, perception but not memory exhibits time-order errors (TOE's). Finally, although Weber's law was violated both in perception and in memory, the underlying basis of this violation differed between conditions.

Psychometric functions are typically interpreted as evidence of an analogue-based remembered representation. However, they can be generated according to SCT if one adds the assumption that correct encoding is probabilistic and depends on the similarity of stimuli on the underlying stimulus dimension (Baranski & Petrusic, 1992). Subjects could arbitrarily pick one of the variable stimuli as the standard and then discriminate each variable stimulus within a set relative to that standard. Psychometric functions with non-zero slopes would result from the probabilistic nature of cue generation such that the probability of correctly encoding a variable stimulus relative to the standard would increase as the difference between the stimuli increases.

Overall Discriminability. Weber fractions, the classic measure of discriminative sensitivity, provide an index of the minimal proportion difference between a comparison stimulus and a standard that is required for discrimination to occur. Smaller Weber fractions reflect a more precise and therefore more discriminable internal representation. Since SCT was not designed to account for perceptual judgements, comparisons between memory and perception can only be made on the assumption that perception will make use of ordinal semantic coding perhaps when stimuli are particularly difficult to discriminate. It is unclear how SCT can account for differences in Weber fractions

between perception and memory both in terms of the serial-position effect and differential overall discriminability.

The Serial-position Effect. The impaired discriminability of symbolically represented stimuli suggests that memory representations lose some of the richness of their perceptual referents. This disparity between perception and memory is complicated by the presence of the serial-position effect in memory. Baranski & Petrusic (1992) showed that Weber fractions for remembered standards exhibit a serial-position effect in the shape of an inverted-U indicating that middle standards suffer from reduced discriminability relative to the end standards. In perception, Weber fractions decreased in magnitude as the magnitude of the standard increased. The term "serial-position effect" (SPE) has been used in several contexts most notably in serial list-learning (Johnson, 1991), but also in psychophysical judgements of positional order (Holyoak & Patterson, 1981) and stimulus identification (for a review see Vickers, 1979).

One possible analogue-based explanation is the positional discriminability model of Holyoak & Patterson (1981) suggesting that each subjective magnitude is represented by a distribution of analogue positional codes with finite mean and variance. The SPE results from the greater overlap and increased variability among distributions in the middle of the range compared to items towards the end points. Johnson (1991) successfully modelled data suggesting that discriminability is determined by a distinctiveness measure (Murdock, 1960). According to this view, distinctiveness is proportional to the distance between a stimulus and the geometric mean or adaptation level (Helson, 1947) of all stimuli in the memory set, and to the strength of association between items in the set. As a result, the properties of serial-list-learning tasks will also characterise Weber fractions.

These factors include the number of standards to be learned, their spacing on the underlying continuum, and the amount of learning provided.

The SPE (Baranski & Petrusic, 1992) is problematic for SCT as it requires additional assumptions. According to this view, Weber fractions can be obtained if one assumes that correct ordinal coding is probabilistic. A type of distance effect would require increased differentiation, and therefore increased probabilities of making errors, as stimulus similarity increases. Since the proportion of accurately encoded variable stimuli depends on the difference between the standard and the comparison stimuli, SCT predicts, in the context of the method of constant stimuli, that Weber fractions should increase as the standard increases. Standard comparison differences increase as the standard increases to maintain a constant ratio. The SPE predicted by SCT behaves like the SPE obtained for perception rather than for memory. Baranski and Petrusic (1992) motivated two converging lines of research that showed that the shape of the serial-position effect can be altered and even eliminated by varying contextual and acquisitional parameters.

Set Size and Range Effects. Overall discriminability and the shape of the SPE depend both on the number of stimuli involved in the learning process (set size) and on the spacing between these stimuli on the underlying continuum (range). Petrusic, Baranski and Harrison (Experiment 1, 1993) found that discriminative sensitivity for remembered standards was greatly reduced in set size 6 compared to set size 3. Harrison and Petrusic (1994) found that Weber fractions were larger in set size 3 compared to size 2. In sharp contrast, perceptual Weber fractions were equally good regardless of set size. The source of the set size effect is more clearly understood when viewed in terms of the serial position effect (SPE). Petrusic, Baranski and Harrison (1993) found that the SPE

was greatly enhanced when six standards were employed relative to three, suggesting that when the number of items increases discriminability suffers and, consequently, learning suffers. This non-independence of remembered stimuli from other stimuli in the set was extended to include the effect of the spacing of stimuli on the underlying continuum in Experiment 2 of Petrusic, Baranski and Harrison (1993). Overall Weber fractions and the quadratic shape of Weber fractions across standards were greatly inflated in the narrow range compared to the wide range condition. This result suggests that the closer stimuli are to each other on the underlying continuum, the less discriminable they are, and, as a consequence, the more difficult they are to learn.

The Dependence of Discriminability on Learning. The non-independence of the discriminability of remembered standards from the number of standards, and the spacing between them can, however, be modified by learning. Following Baranski and Petrusic's (1992) methodology, Kennedy (1990) used the method of constant stimuli with confusable stimuli. Kennedy (1990) showed that previous exposure to three symbolically represented line standards greatly improved their discriminability relative to other standards when they were embedded among three new standards in a subsequent phase. Within the larger set totalling 6 standards, the standard that previously served in ordinal position 1 became standard 2, the previously positioned standard 2 became standard 4, and the previously positioned standard 3 became standard 5. The six-standard embedded condition was then compared to a non-embedded 6-standard condition where subjects were exposed to all six standards for the first time. Weber fractions for the embedded target standards were clearly diminished relative to the three corresponding standards in the non-embedded condition. The discriminability function across standards for the

embedded condition exhibited a saw-tooth shape where the dips in Weber fractions corresponded to the increased discriminability of the embedded target standards. In contrast, the non-embedded condition displayed the classic quadratic shaped serial-position effect.

Kennedy's (1990) findings stimulated a more precise examination of the effect of learning. Petrusic, Baranski & Harrison (1993) manipulated acquisition by giving additional learning to one group of subjects after they had satisfied the base learning criterion during paired-associate training. The comparison group was given no additional learning. Subjects were assigned to either a wide or narrow stimulus range. The massive effect of stimulus range on the serial-position effect was evident only when no additional learning was provided. Additional learning severely attenuated both the range and serial position effect with the result that memory behaved much more like perception. Overall, however, Weber fractions were still in the 10% range which is more than twice as large as those obtained for perception. Symbolically represented stimuli still do not retain all the richness of perceptual phenomena.

The differential effects of learning on the magnitude of Weber fractions are clearly damaging to SCT since learning affects the precision of the representation and not its ordinal position. Most importantly, in the case of the method of constant stimuli, the process of differentiation cannot be invoked since no comparisons are made between standards.

Time-order Errors. Baranski and Petrusic (1992) further distinguished memory from perception by showing the complete absence of time-order errors (TOEs) with symbolically represented stimuli. The most common form, the negative TOE, occurs if

accuracy in the short-long order of temporal presentation is greater than accuracy in the long-short order for the same pair of stimuli. A positive TOE results from increased accuracy in the long-short order relative to accuracy in the short-long order for the same pair of stimuli. The first stimulus is held in memory while the second is examined thus giving rise to the commonly used label of ‘perceptual-memory’ comparisons (Kinchla & Smyzer, 1967; Laming & Scheiwiller, 1985)

The TOEs, first described by Fechner (1860/1966), have been reported in a wide range of perceptual judgements involving intensive or extensive stimuli, especially duration discrimination (Allan & Kristofferson, 1974; Hellström, 1977; Jamieson & Petrusic, 1975a, 1975c; Petrusic, 1984. Also see Hellström, 1985 for an extensive review). Several views have suggested that the TOE results from constant error (Fechner, (1860/1966), fading trace (Kohler, 1923), or bias (Luce, 1959; Luce & Galanter, 1963; Wickelgren, 1968). However, differential-sensation weighting (Hellström, 1977), the predominant theory, interprets TOEs as a perceptual-memory phenomena (Hellström, 1977, 1985; Jamieson & Petrusic, 1975a, 1975b; Petrusic, 1984) susceptible to contrasting or assimilating effects induced by variations in stimulus magnitude and ISI.

The absence of TOEs with symbolically represented stimuli (Baranski & Petrusic, 1992; Petrusic, Baranski & Harrison, 1993) suggests that stimuli whose properties are stored in long-term memory and represented by a name (CVC) are different from perceptual-memory stimuli where the first stimulus is committed to short-term memory. Decision time provides a basis for the absence of TOEs. Baranski & Petrusic (1992), and later Harrison and Petrusic (1994), showed that RTs are uniformly faster when a perceptual stimulus is followed by a CVC. Presenting a perceptual stimulus allows the

decisional system to activate a parallel comparison process between that perceptual stimulus and each of the standards in memory. The subsequent presentation of the CVC merely allows the subject to make the appropriate comparative judgement. Harrison and Petrusic (1994) is of particular importance since they used the same set of comparison stimuli for all standards. The use of non-overlapping comparison sets by Baranski and Petrusic (1992) could have promoted the use of strategies thus precluding the successive comparison of stimuli necessary for the occurrence of TOEs (Hellstrom, 1977). Harrison and Petrusic (1994) showed that for equal comparison-standard (CVC) stimuli, the observed proportion of times that the variable stimulus was judged longer than the standard (CVC) closely paralleled the actual proportion of variable stimuli that exceeded the specified standard.

Weber's law states that the discriminability of stimuli located on the same physical continuum is described by a constant ratio of the just-noticeable difference to the standard (see Krueger, 1989 for a critical review). As such, absolute differences between stimuli should have no effect on discriminability. Baranski and Petrusic (1992) held the standard-to-variable ratios constant between standards but permitted the standard-variable differences to increase as the magnitude of the standard increased. In violation of Weber's law, Weber fractions for perceptual standards decreased as standards increased. In the memory condition, Weber's law was violated with respect to the outer standards. Weber fractions were larger for the largest standard compared to the smallest standard. Inner standards were affected by the SPE. Baranski & Petrusic (1992) showed, however, that violations of Weber's law for perception are mediated by the existence of TOEs that decrease as the absolute standard-comparison differences increase, but only for stimuli

presented in the long-short order. No violations of Weber's law occurred in the short-long order because these were relatively easy (Jamieson & Petrusic, 1975b). In contrast, the psychometric functions for remembered standards were characterised by identical slope patterns regardless of presentation order. In memory, Weber's law is violated to the same degree regardless of presentation order.

Summary. Research using the method of constant stimuli suggests that symbolically represented stimuli are characterised by information that is more than ordinal in precision. The SPE in the form of an inverted-U is not admitted by SCT. Symbolically represented stimuli are also affected by the context within which they are learned as determined by the number of other items present (set size) and the spacing between them (range). These contextual effects are lessened by additional learning that may act by sharpening the associative links made between labels and their perceptual referents (Johnson, 1991). Although perceptual standards are not affected by these global factors, they are influenced by local factors as shown by the presence of TOEs. The absence of TOEs in judgements of remembered stimuli results from the non-perceptual and non-successive nature of the comparison process whereby remembered representations are activated in parallel. In this case, order effects reveal themselves as strategies and biases in response time and response probability.

Chronometric Properties of Remembered Stimuli

The Distance Effect. The time to decide which of two perceptual stimuli has more or less of some property is inversely related to differences in the relevant underlying property of these stimuli.² This distance effect has been reliably demonstrated in a wide

variety of stimulus dimensions (Cattell, 1902; Henmon, 1906; Johnson, 1939; Kellogg, 1931; Munsterger, 1894; Welford, 1960). The 'inner psychophysics' of binary comparisons with remembered stimuli (Moyer, 1973) was first investigated using controlled learning conditions by Moyer and Bayer (1976). After paired-associate training, CVCs were then used for binary comparisons of circles, thus defining the memory condition. The perceptual condition required the binary comparisons of perceptually available circles. The time to judge the relative size of circles in memory was inversely related to the differences in the sizes of their perceptual referents. This 'symbolic' distance effect (SDE), which was greatly enhanced compared to the perceptual distance effect, bolstered Moyer's (1973) argument that symbolically represented physical stimuli are characterised by properties analogous to those of perceptual stimuli.

Semantic coding theory (SCT) provided a convincing argument that the symbolic distance effect could be explained by the use of differentiation to generate discriminable codes (Banks, 1977). RTs for relatively different squares will be fast because they are more likely to be coded into different ordinal categories. As squares become more similar in size, semantic codes are more likely to fall in the same categories thus precluding discriminability. As a result, the stimuli are recoded into finer ordinal categories on the basis of information retrieved from the data base. Semantic coding theory therefore predicts the SDE because of additional processing time required to re-categorise ordinal codes into finer categories.

In addition to simple comparisons between two perceptual stimuli and between two symbolically represented stimuli, Banks (1977) required mixed comparisons between one perceptual stimulus and one symbolically represented stimulus thus permitting a direct

test of SCT. According to this view, mixed trials will take longer than pure memory trials because the perceptual stimulus must be recoded into a symbolic format that is compatible with the memory stimulus' semantic code. Judgements involving pure memory trials are performed faster because both stimuli are already in a compatible format. This prediction was supported by Banks (1977) and more fully articulated in Banks and Flora (1977). As became clear in the sequel (Petrusic, Baranski & Aubin, 1995), this particular pattern of RTs is only observed under very particular range conditions.

The Range Effect. In a number of experiments with a variety of stimuli, Petrusic and Baranski (1994) were unable to replicate the Banks' (1977) results. Almost uniformly, pure perceptual comparisons were fastest, followed by mixed comparisons and then by pure memory comparisons. These findings are consistent with the analogue view that perceptual information must be retrieved from memory for comparison with the perceptual stimulus. The pure memory condition will take longest because perceptual information must be retrieved from two representations (Farah, 1985; Paivio, 1971; and Kosslyn, 1980 for a similar but image-based view).

Petrusic, Baranski and Aubin (1995) demonstrated that longer RTs for the mixed comparisons could only occur if the stimulus range is reduced enough to preclude perfect learning of the CVCs. With the imperfect learning of symbols, the analogue-like properties of the remembered stimuli are very difficult to retrieve for comparison with the perceptual stimulus. As a result, RTs for mixed comparisons take longer than pure memory comparisons where judgements are based on the ordering of learned symbols. Of great importance was the behaviour of accuracy for each condition. Since stimulus pairs

were highly confusable, pure perceptual comparisons were imperfect and in fact less accurate than pure memory comparisons. This suggests that comparisons of remembered stimuli are based on ordinal codes only when task difficulty prevents the retrieval of perceptual information. In most cases, the decisional system recovers analogue information from the memory representation to make the comparison with the perceptually available stimulus.

The importance of the work by Petrusic, Baranski and Aubin (1995) stems from the fact that the response time reversal occurred as a result of a range manipulation. Moyer and Bayer (1976) are noted primarily for their demonstration that the underlying properties of memory stimuli exist at a level of measurement that exceeds that of an ordinal scale. Moyer and Bayer (1976) presented a group of subjects with a narrow range of circles which were relatively similar in size, and a wide range of circles which were relatively different. If comparative judgements of symbolically represented stimuli make use of ordinal information only, as suggested by a strict semantic coding view, the decisional system will be insensitive to differences between stimuli. As a result, relatively narrow and relatively wide range conditions should have no effect on decision time. Moyer and Bayer (1976) found that judgements in the narrow range took longer than judgements in the wide range, both in perception (50ms) and in memory (125 ms). In fact, although providing evidence of the ordering of intervals on a supra-ordinal scale and not necessarily an interval scale representation, the 'range' effect was crucial in showing that the semantic coding theory was indeed too crudely articulated to permit such a finding.

Studies that failed to replicate the 'range' effect were poorly designed. Banks, Mermelstein and Yu (1982) failed to uncover a range effect for perception because discriminations were too easy. Henderson and Well (1985) did not even examine the occurrence of a range effect in perception. More recently, the range effect was replicated by Petrusic, Baranski and Aubin (1995). These authors showed that the 'range' effect is typically quite small (50 ms) and can be difficult to detect especially in a between-subjects design where large individual differences can lead to a significant reduction in power.

The Semantic Congruity Effect. The existence of the semantic congruity effect in memory (Jamieson & Petrusic, 1975c) has also been taken as evidence that the properties of the memory representation are very finely graded and reflect the interval properties of their perceptual referents (Petrusic & Baranski, 1989). Extending work begun by Shipley, Norris and Roberts (1945) in the context of colour preference, Audley and Wallis (1964) showed that the semantic requirements of a comparative judgement has an important impact on decision times. Audley & Wallis (1964) found that the time to select the 'BRIGHTER' of two relatively bright stimuli is faster than the time taken to select the 'DARKER' of the same pair. Conversely, the time taken to select the 'BRIGHTER' of two relatively dark stimuli is slower than choosing the 'DARKER' of the same pair. This interaction between the direction of the comparative and the location of the stimuli on the underlying physical continuum was originally called the crossover effect, and later, the semantic congruity effect (SCE). The SCE has been replicated in a wide variety of continua including, for example, pitch discrimination (Wallis and Audley, 1964), loudness (Banks & Root, 1979), colour (Holyoak & Patterson, 1981), line lengths,

heaviness, and unfilled distances (Petrusic & Baranski, 1989). The shape of the SCE is graded such that, when the stimulus pairs are equated for difficulty, the magnitude of the differences in RTs increases as stimuli become more extreme on the continuum. In addition, the location of the crossover point is contextually dependent on the location of the bisector of the presented stimulus range (Banks, 1977; Cech & Schoben, 1985; Petrusic & Baranski, 1989).

The SCE has been shown to be influenced by a number of factors. The magnifying effect of judgement difficulty on congruity (Petrusic, 1992; Jamieson & Petrusic, 1975; Petrusic & Baranski, 1989) has been replicated in memory judgements suggesting a common underlying representation and common decisional mechanisms between memory and perception (Petrusic et al., 1995; Petrusic, Baranski and Aubin, 1995). Instruction-based emphasis on accuracy amplifies SCE while emphasis on speed attenuates it (Petrusic, 1992; Petrusic & Baranski, 1989).

Several studies documented the existence of the SCE with remembered stimuli but failed to replicate statistically significant SCEs with perceptual judgements (Banks, Mermelstein and Yu, 1982; Duncan & McFarland, 1980; Marschark & Paivio, 1981). In these studies, the exclusive occurrence of SCEs in memory was interpreted as supporting evidence for the semantic coding theory (Banks, 1977; Banks & Flora, 1977). SCT predicts a particular form of the SCE (Banks, 1977; Banks & Flora, 1977) for which no evidence exists (Wallis and Audley, 1964). According to this view, if two relatively small stimuli are coded as, say, S+ and S and must be judged on the basis of the instruction 'SMALLER', response times are faster than if the instruction is 'LARGER' because in the latter case, the incompatibility of the instruction code and the stimulus codes requires

translation. Since differential response times occur as a result of additional constant translation time that occurs post-decisionally, semantic coding theory predicts that the difference in response times between instructions will be constant along the continuum with an obvious reversal in the centre of the continuum. The SCE is therefore only susceptible to manipulations that affect code translation, not stimulus properties.

Comparison difficulty is a crucial factor in the magnitude of the SCE and may explain the presence of a small but non-significant SCE for perception in figure 2 of Banks et al. (1982). Petrusic & Jamieson (1979) showed that reliable SCEs do occur in perception when comparisons are difficult. The SCEs does occur with errorless (supraliminal) perceptual judgements (Petrusic, 1992; Petrusic & Jamieson, 1979) but is of such small magnitude (20ms) that the effect is not statistically significant unless accompanied by large sample sizes. In a series of experiments, Petrusic et al. (1995) and Banks et al. (1982) showed that what is a minuscule 20 ms effect with errorless perceptual comparisons is magnified several times when the same stimuli are symbolically represented.

The demonstration of SCEs in perception does not eliminate semantic coding theory as a valid theory of the comparison process. Rather, its downfall comes from its inability to account for several characteristics of the SCE. Semantic coding theory cannot account for changes in the magnitude of the SCE as a function of variations in the relative demands for speed and accuracy which are regulated by decisional processes rather than post-decisional code translation (Petrusic, 1992). Secondly, since according to semantic coding theory the locus of the SCE is located post-decisionally during code translation, difficulty-based changes in the magnitude of the SCE could only occur if the difficulty

manipulation affects code translation (Banks, Clark & Lucy, 1975). Semantic coding theory therefore predicts an attenuated SCE as the distance between stimuli increases since stimuli are then more likely to be coded into separate categories and thus less likely to require code translation. For example, choosing the larger of two stimuli coded S and L would require no code translation and thus not exhibit SCE, since the stimulus coded as L will be selected automatically. On the other hand, stimuli located at the extreme of a continuum such as S and S+ would have to be translated to match the codes for the instruction, 'LARGER' (Banks, 1977). SCT predicts comparable SCE magnitudes for stimulus pairs coded at the same end of the continuum but that vary in difficulty (Petrusic, 1992; Petrusic & Baranski, 1989a) because all will require the same amount of code translation. Finally, the magnitude of the SCE for errors/corrections interacts with judgement difficulty and with relative demands for speed and accuracy. Petrusic (1992) clearly demonstrated that the SCE is larger for errors than for correct responses when accuracy is emphasised. However, instructional demands for speed that result in equivalent response times between correct and incorrect responses also result in equivalent SCE magnitudes between response times for correct and incorrect responses. According to SCT, there should be no SCE for errors because these can occur either due to improper code generation which is predecisional, or due to code translation failure, in which case there can be no SCE.

Any comprehensive model of the comparison process should be able to model both patterns of response time and response accuracy. This goal has yet to be achieved since the conditions under which response accuracy will exhibit the SCE are not well understood (Petrusic, 1992; Petrusic & Baranski, 1989). One possibility is that the SCE

results in part from a stroop-like effect. Banks and Root (1979) developed a post-decisional model whereby responding is delayed in the incongruent case because the stimuli are processed more quickly than the instructions. As a result, the decisional system is biased towards expecting the inappropriate form of the instruction. The congruent case leads to facilitation. However, the stroop-like effect would have to occur in the decisional locus in order to account for the properties of the SCE.

Similarity Scaling in Memory Psychophysics

The range effect demonstrated with binary comparisons suggests that the properties of remembered physical magnitudes can be characterised by an ordering of intervals (Moyer & Bayer, 1976; Banks et al., 1982; Henderson & Well, 1985; Petrusic, Baranski & Aubin, 1995). Despite not being as precise as an interval scale (Krantz, Luce, Suppes & Tversky, 1971), a supra-ordinal scale permits the recovery of a representation that is unique up to a linear transformation by applying fundamental measurement to stochastic choice data in the context of similarity-dissimilarity judgements (see Footnote 3). If the memory representation of perceptual information is precise only up to an ordinal level of measurement, then similarity judgements between remembered pairs of pairs of stimuli will not be possible and the axioms for an algebraic-difference structure will not be met (see Footnote 4 for an alternative approach). Satisfaction of the testable axioms would therefore undermine current semantic coding theories (Banks, 1977; Cech & Shoben, 1985) and provide a basis for the development of more complex theories that predict judgement behaviour based on interval level information (Baranski & Petrusic, 1992).

Petrusic (1996) applied fundamental measurement to similarity-dissimilarity judgements and found that the testable axioms (see Footnote 5) were satisfied, thereby demonstrating that both perceptual and remembered representations allow the homogeneous concatenation of intervals along a one-dimensional continuum. The tenability of the axioms significantly weakens strict propositional accounts of memory (Banks, 1977; Banks et al., 1977) and completes the landmark work begun by Moyer & Bayer (1976). The properties of a memory representation are analogous to their original perceptual counterparts in that they preserve interval scale properties.

Petrusic et al. (1995) also showed that perceptual and remembered similarity-dissimilarity judgements exhibit processing behaviour akin to that found in the context of binary comparisons. They found converging support for a positive-difference structure first demonstrated on the basis of response probability. A distance effect for perception and for memory was evident in the decrease in RTs as the difference of differences of ALSCAL-based scale values increased. Semantic congruity effects (SCE) were clearly evident both in perception and in memory (see Footnote 6). Mimicking binary comparisons, the SCE was more pronounced in memory than in perception. Finally, faster RTs for disjoint pairs compared to RTs for the corresponding overlapping pairs suggested that decisional processes in similarity-dissimilarity judgements are sensitive to factors other than actual differences of differences, perhaps even involving ratios of differences (see Footnote 7). Taken together, the distance effect, the SCE and the overlapping-disjoint effect for RTs further extend the analogue-based parallel between perception and memory to a type of judgement where the decisional processes are too complex to be accounted for by ordinal codes.

Experiment 1

Subject-controlled Sampling of Explicit Observations in Memory and Perception

Introductory Comments

As mentioned in the introduction, variable-sample accrual theories provide the most successful models of comparative judgements; however, the underlying accrual process is essentially implicit and not amenable to direct manipulation (For a review, see Vickers, 1980). The expanded-judgement (E-J) paradigm can be thought of as an analogy to variable-sample models (Vickers, Burt et al., 1985). The number of explicit observations required to make an expanded judgement should reflect the implicit accrual process and would thus vary on the basis of stimulus information (perception vs memory, discriminability) and the semantic requirements of the judgements (semantic congruity effect). As such, judgements involving remembered standards should require more observations than those involving perceptual standards since memory representations are thought to be more variable than the perceptual representations of the same physical stimuli (Baranski & Petrusic, 1992). Difficulty as defined by the difference between stimuli should also affect the comparison process by increasing the number of observations when the comparison is harder (Petrusic, 1992; Petrusic & Jamieson, 1979). More observations should be associated with incorrect responses compared to correct responses, consistent with predictions based on latency-probability functions (Petrusic, 1992). The semantic congruity effect (SCE) should reveal itself as an increase in the number of observations for the instruction 'LONGER' when line pairs are relatively short, and for the instruction 'SHORTER' when the stimuli are relatively long.

Decreased accuracy and slower decision times for judgements with remembered stimuli compared to perceptual judgements have been used as evidence of a noisier memory representation (Baranski & Petrusic, 1992). According to this view, the larger Weber fractions and longer response times in memory compared to perception, and the serial position effect for Weber fractions and response time in memory are due to greater variance in the underlying memory representation (Baranski & Petrusic, 1992; Holyoak and Patterson, 1981; Johnson, 1991; Petrusic, Baranski & Harrison, 1993). The E-J paradigm applied to the traditional method of constant stimuli provides the opportunity to explicitly examine the effect of distance, variability and semantic congruity on a process that is assumed to mimic the implicit evidence accrual process through variations in sample size and sampling time.

In the classic method of constant stimuli, the subject is presented with, say, two horizontal line lengths consisting of a standard and a comparison stimulus (Fechner, 1860; Petrusic & Baranski, 1989). A complete stimulus set typically consists of several standards. In one version, each standard has a unique set of comparison stimuli that do not overlap between sets (Baranski & Petrusic, 1992). In an E-J paradigm, the comparison stimulus is not explicitly presented as a stationary line; rather, its value is the mean of a normal distribution of line lengths with some specified σ . On any given trial, the O is presented with a standard and a sample of successively presented line lengths randomly selected from a comparison distribution. The instructions require that the O make a comparative judgement between the standard, s_k and the true mean of the comparative distribution, v_{jk} , as estimated on the basis of the sample mean.

Thurstonian-based Expanded Judgement Model 1

A Thurstonian representation is assumed to describe the sensory space. When perceiving a standard, s_k , the O randomly selects an observation, s'_{ik} , from a normal distribution of subjective values corresponding to the k th standard, with $E(s'_{ik}) = \mu_k$ and variance, $\sigma_{s'_k}^2$. Although each comparison stimulus on a given trial can be seen as giving rise to separate normal distributions of sensory values, the simplifying assumption is made that in the long-run, the average of these separate distributions yields an overall parent population distribution with $N(\mu_{jk}, \sigma_{jk})$. This assumption seems reasonable, since all observations on a given trial are randomly sampled from the same parent population. In this case, $\bar{y}_{ijk}$ is the momentary subjective mean of the i th sample drawn from the j th comparison distribution specific to the k th standard. In the long-run,

$E(\bar{y}_{ijk}) = \mu_{jk}$ with the variance of the estimate of the mean $\sigma_{\bar{y}_{jk}}^2 = \frac{\sigma_{jk}^2}{n}$, where σ_{jk}^2 is the variance of the distribution of subjective line observations and n is the number of observations requested by the O. The variances of the two distributions, $\sigma_{s'_k}^2$ and σ_{jk}^2 , are assumed to be independent.

According to this Thurstonian conceptualisation, the task requires that the O make a comparative judgement between the subjective value of the standard, μ_k , and the true mean of the subjective comparison distribution, μ_{jk} . According to this view, the O is assumed to use the mean of the sample of observations, $\bar{y}_{ijk}$, as the basis for estimating μ_{jk} . Since $\bar{y}_{ijk}$ varies from trial to trial, discriminability must take into account the difference, $\mu_k - \mu_{jk} = d$, the variability associated with the standard, σ_k^2 , and the

variability associated with the comparison distribution of estimates, $\sigma_{y_{jk}}^2$. Incorporating these concepts provides a Thurstonian-based discriminability index,

$$D_n = \frac{0 - d}{\sqrt{\sigma_k^2 + \frac{\sigma_{jk}^2}{n}}} \quad (1),$$

which is the standardised distance between the unbiased Thurstonian criterion, 0, and the expectation of the distribution of differences, d . The value of the criterion represents the boundary between two types of responses. The area to the right of the criterion represents the probability that the comparison mean is correctly judged longer than the standard. The area to the left represents the probability that the comparison mean is correctly judged shorter than the standard. The unbiased case is used for simplicity. When the expectation of the difference distribution is equal to the criterion, then $D_n=0$ indicating that in the long-run the comparison will be judged longer than the standard 50% of the time. As the criterion moves left, the comparison is judged longer than the comparison more than 50% of the time reflecting an increasing in discriminability. Increased discriminability can be due to an increase in d , or a decrease in the variability of the differences. Since $\frac{\sigma_{jk}^2}{n}$ can be reduced by increasing n , then the overall variability of the distribution of differences will decrease resulting in increased discriminability. Equation (1) therefore predicts that for a given d , D_n will increase as n increases. More specifically, more observations should be requested judgments with remembered stimuli relative to judgements with perceptual stimuli. More observations should be required when the variability of the comparison distribution is increased. Figure 1 provides a visual appreciation of the effects of sample size on discriminability for two levels of

comparison difficulty, for two levels of variability in the comparison distribution, and for perception and memory. A comparison of hard comparisons (filled circles) to easy comparisons (unfilled circles) in each of the panels shows that comparison difficulty greatly reduces the slope of the discriminability function. As a result, the decisional system should require more observations to improve discriminability.

Memory and the Effect of Variability on n

Assuming that the variability of the internal representation of comparison observations is directly related to the variability of the physical observations, then increasing σ_{jk}^2 should result in increased sampling in order to attain the same level of discriminability achieved when σ_{jk}^2 is smaller (Petrusic, 1992). This prediction is clearly indicated by the discriminability functions in Figure 1 where increasing the variability of the comparison distribution (panels B and D relative to panels A and C) results in slower improvements in discriminability as sample size increases. In a similar manner, when the standard is available only by way of a symbol learned through paired-associate training, then σ_k^2 should be larger than when the standard is perceptually available (Baranski & Petrusic, 1992). As a result of the increased variability of the memory-based differences, a larger n should result from an attempt to improve discriminability (See panels A and B versus C and D of Figure 2).

In this experiment, perception vs memory will be factorially varied with two levels of comparison variability (low and high). This 2 x 2 between-subjects factorial design will permit the simultaneous comparison of the effect of memory and comparison variability on the requested n and on sampling times. Consistent with the effects on

Weber fractions and response times reported in studies using the classic method of constant stimuli, the current SCON E-J paradigm should find a larger number of observations and longer sampling times in memory compared to perception, and in the high variability perceptual group compared to the low variability perceptual group. Increasing the variability of the comparison distribution should exaggerate the serial-position effect (SPE) for memory standards reported for discriminability and response time (Harrison & Petrusic, 1994). The SPE should occur for Weber fractions, sample size and sampling time.

Expected Value as the Basis for Sampling

The model suggested in Equation (1) is not complete however, since although it suggests that the decisional system should continue sampling to increase discriminability, it lacks a clear basis for a stopping-rule. Equation (1) suggests that the O might never stop sampling since an infinite n results in an infinitely large D_{eff} . The evidence accrual process postulated by stochastic counter models must obviously assume that covert accruals carry some form of meaningful information, otherwise the accruals would be irrelevant to the decisional process. The information content of each accrual determines response time and response probability (Petrusic, 1992). Petrusic and Jamieson (1978) showed that the a priori importance of judgements has a strong impact on the information content of each accrual and consequently on response-time/response frequency relations (LFC). The LFC simply shows that when accuracy is stressed, response times decrease as a response becomes more probable. Over repeated judgements that lack personal importance, the LFC becomes non-monotone indicating an increased tendency to guess.

PERCEPTION

MEMORY

Figure 1. Effect of sample size (n) on discriminability (D) for perceptual ($\sigma_k^2 = 9$, top panels) and for memory standards ($\sigma_k^2 = 36$, bottom panels) under low ($\sigma_{jk}^2 = 36$, left panels) and high ($\sigma_{jk}^2 = 324$, right panels) comparison variances for small ($d = 3$, filled circles and squares) and large ($d = 6$, unfilled circles and squares) standard-comparison differences as predicted by the Thurstonian-based expanded-judgement model 1.

The lack of discriminatory information on each accrual leads to a greater proportion of trials on which the decisional system guesses from a covert decisional state of indifference. When judgements are personally important however, such as in the case of gambles, the LFC remains monotone suggesting that accruals carry more information.

In evidence accrual models, e.g. slow-fast guessing theory (Petrusic, 1992), the number of accruals is determined on the basis of the difficulty of the judgements and on the basis of instructions that emphasise accuracy or speed. In this SCONE-J paradigm, one could reasonably hypothesise that the number of overt observations should continue to increase until improvement in discriminability from one sample to the next becomes meaningless. In the case of an E-J paradigm, 'meaningfulness' must be defined by artificially attaching value to each observation. In such a paradigm, the decisional system would be required to maximise accuracy to receive some valued outcome, while incorporating a workable understanding of the paradoxical effects of sample size. Increasing n improves the chances of being correct, but, since the acquisition of information incurs a cost, a larger n also decreases the overall value of the decision (Manz, 1970).

In a task of such complexity, overt instructions may not be sufficient to preclude the development of idiosyncratic strategies or, simply put, of boredom with a task that may be too difficult to understand. Edwards (1961) shows how payoffs act not only by providing general motivation because of the prospect of monetary gains, but more importantly, by inducing implicit instructions for assessing the relative desirability of changes in accuracy as a function of sample size. Expected value is a useful means of

ascribing relative weights to correct and incorrect responses because input values can be objectively controlled.

In the traditional conceptualization, expected value suggests that Os linearly combine monetary outcomes with their respective objective probabilities in order to determine the intrinsic value of a decision (Edwards, 1954b; Edwards & Tversky, 1964; Tanner & Swets, 1954). The general model is

$$EV = Wp(c) + Lp(\bar{c}) \quad (2)$$

where EV is the expected value associated with a particular judgement situation, W is the amount of money won, $p(c)$ is the probability of the event, correct, L is the amount of money lost, and $p(\bar{c})$ is the probability of the event, incorrect i.e. $1 - p(c)$.

Equation (2) can be modified in two important ways to incorporate the E-J nature of this experiment. First, $p(c)$, is defined as the inverse function of the Thurstonian measure of discriminability, such that $p(c_n) = \Phi(D_n)$ where Φ is an inverse transformation that provides the area to the right of D_n according to Tail's normal deviate. As a result, a separate EV is obtained for each n due to the associated changes in discriminability. Equation (2) therefore becomes

$$EV_n = Wp(c_n) + Lp(\bar{c}_n) \quad (3)$$

which implicitly incorporates the E-J discriminability model of equation (1). Finally, in order for EV_n to reflect the costs associated with sampling a specified n , equation (3) becomes,

$$EV_n = Wp(c_n) + Lp(\bar{c}_n) + n \times l \quad (4)$$

where n is the number of observations and l is a negative value determining the cost of each observation. This model therefore predicts that sampling will continue until EV_n levels off or until the improvement in EV_n with increasing n becomes counter-productive.

Expected value (EV) functions determined on the basis of this Thurstonian E-J model are arbitrary in that their general form is in part determined by the experimenter's choice of monetary values (Tanner & Swets, 1954). As Edwards (1961) points out, however, no general rule exists. In the current proposed experiment, values associated with the various outcomes are chosen such that the EV functions will be sensitive to the effects of comparison difficulty and variability on sample size. Figure 3 provides expected value functions for perception and memory, under low and high comparison variance for easy (large diff) and hard (small diff) comparisons. These functions incorporate the discriminability functions illustrated in figure 1. In the scenario presented in figure 2, a correct primary response earns \$0.50 less the cost incurred for the number of requested observations. An incorrect response incurs a loss of \$0.50 in addition to the cost incurred for the number of requested observations. For example, an correct response with $n=10$ would pay $\$0.50 - \$0.10 = \$0.40$. An error would cost $\$0.50 + \$0.10 = \$0.60$.

The arrows in figure 1 indicate optimal sample sizes for each EV function. As is clearly evident, predictions concerning optimal sample size are in some cases opposite those predicted by evidence accrual models. First, increasing comparison difficulty by decreasing the absolute difference between the standard and the mean of the comparison distribution leads to smaller sample sizes in three of the four graphs. The source of this effect can be better understood by examining figure 1. It is clear from the perceptual

PERCEPTION

MEMORY

Figure 2. Effect of sample size (n) on expected value for perceptual ($\sigma_k^2 = 9$, top panels) and for memory standards ($\sigma_k^2 = 36$, bottom panels) under low ($\sigma_{jk}^2 = 36$, left panels) and high ($\sigma_{jk}^2 = 324$, right panels) comparison variances for small ($d = 3$, filled circles and squares) and large ($d = 6$, unfilled circles and squares) standard-comparison differences as predicted by the Thurstonian-based expanded-judgement model 1 incorporating expected value (correct=\$0.50, error=-\$0.50, cost per line=-\$0.01).

condition with low comparison variance (panel A) that decreasing the difference between the standard and the mean of the comparison distribution mitigates the positive effect of sample size on discriminability.

Note in figure 2 that the magnitude of the improvement in EV from one sample size to the next is much smaller in the more difficult condition (filled circles) compared to the easier condition (open circles). As a result, maximal EV is reached sooner in the more difficult condition. The second counterintuitive prediction is that, relative to perception (panels A and B of figure 2), memory (panels C and D) reduces the optimal sample size. This occurs because increasing the variance of the standard reduces the beneficial impact of sample size since the latter directly affects the variance of the estimates of the comparison mean but not the variance of the standard. Consistent with evidence accrual models, more observations will be taken when the variability of the comparison distribution is increased (panels A and C versus panels B and D in figure 2) because changes in discriminability become more linear.

EV is important in maximising the chances that Os are carefully considering each observation. In a pilot study (Aubin, 1994), it was found that Os requested an unusual large number of observations at a rate that in many cases precluded perceptual processing (e.g. 100-200ms). Assigning positive values to correct responses and negative values to incorrect responses helps ensure that each primary comparative judgement is taken seriously. However, such a strategy by itself would not preclude the use of unreasonably large or small sample sizes that may not reflect the underlying discrimination process. Attaching a cost to each observation emphasises that each observation must be used judiciously. Pilot work using imaginary payoffs supported this view (Aubin, 1994).

In order to encourage Os to sample observations, a relatively high reward for each correct response is given and a low cost is charged per observation. The magnitude of the loss associated with an incorrect response is also very important in determining the magnitude of the change in EV as sample size increases (Edwards, 1961). Figure 3 illustrates a condition whereby a base value of \$0.50 is paid for a correct response and a much smaller amount is charged for an incorrect response, \$0.10.

A comparison of figures 2 and 3 shows that charging a small amount for an error (figure 3) results in a type of ceiling effect so that all EVs are relatively high and less discriminable across sample sizes. Charging an amount equal to the amount won for a correct response (figure 2) results in maximum sensitivity of EV to sample sizes and will be the strategy adopted in these experiments.

Thurstonian-based Expanded Judgement Model 2 for Perception

An alternative view suggests that on each overt sampling, the O takes an observation from the distribution of subjective values for the perceptual standard as well as an observation from the comparison distribution. This model can only apply to perception since the precision of the memory representation is determined by the associative learning process that takes place during paired-associate training not during the comparative judgement phase. It is difficult to see how taking samples from the subjective distribution for the remembered standard would in any way improve the precision with which the expectation is estimated.

Incorporating the effect of sample size on both the estimate of the comparison mean and the estimate of the standard, equation (1) becomes

PERCEPTION

MEMORY

Figure 5. Effect of sample size (n) on expected value for perceptual ($\sigma_k^2 = 9$, top panels) and for memory standards ($\sigma_k^2 = 36$, bottom panels) under low ($\sigma_{jk}^2 = 36$, left panels) and high ($\sigma_{jk}^2 = 324$, right panels) comparison variances for small ($d = 3$, filled circles and squares) and large ($d = 6$, unfilled circles and squares) standard-comparison differences as predicted by the Thurstonian-based expanded-judgement model 2 incorporating expected value (correct=\$0.50, error=-\$0.50, cost per line=\$-0.01).

$$D_n = \frac{0 - d}{\sqrt{\frac{\sigma_k^2}{n} + \frac{\sigma_{jk}^2}{n}}} \quad (5).$$

Figure 4 provides a graphic representation of this model's predictions of changes in discriminability as sample size increases.

Model 2 results in less compression of the discriminability function for perception. Incorporating equation (5) with the expected value model of equation (4) provides the following predictions for sample size. The two bottom panels in Figure 5 are the predictions from Thurstonian models 1 for the memory condition and have already been discussed. Figure 5 clearly shows that increasing the variance of the comparison distribution leads to increased sampling (see panel A versus panel B, and to a lesser extent panel C versus panel D). A difficulty effect is predicted in perception with low comparison variability (panel A) but none in memory (panel C). A reversed distance effect is predicted when comparison variability is high. Easier comparisons (unfilled squares) require more observations than harder comparisons (filled squares).

Figure 4 accounts for these effects. When variability of the comparison distribution is increased, the discriminability functions exhibit compression (panel B relative to panel A). However, easier comparisons (unfilled circles in panel A and unfilled squares in panel B) exhibit massive compression relative to the harder comparisons (filled circles in panel A and filled squares in panel B) that are limited by floor effects. Figure 5 shows that this causes relatively greater shifts in the optimal sample size for easier comparisons (unfilled circles in panel A relative to unfilled squares in panel B) compared to harder comparisons (filled circles in panel A relative to filled squares in panel B). In fact, the optimal sample size for harder comparisons decreases with increased comparison

variability because the cost of information in conjunction with minimal advantage of increased sample leads to a much smaller EV.

The clear dependence of the difficulty effect on the magnitude of comparison variability is problematic because psychophysical data clearly demonstrate the robustness of the distance effect. All models of comparative judgement predict that samples sizes (decision latency) will monotonically increase as the distance between stimuli decreases (Vickers, 1980) and that the effect should be magnified when stimulus variability is increased (Moyer & Bayer 1976).

Models 1 and 2 appear to predict counterintuitive effects because in these models, observations contribute only indirectly to discriminability by altering the variability of the underlying distribution of estimates. For a given variance, a small d leads to relatively smaller sample size-dependent improvements in discriminability and EV than that predicted by a larger d . However, both models are interesting because they suggest that sampling should stop sooner in situations where discriminability-related increases in EV are swamped by the cost of information acquisition.

Explicit Sampling Based on Evidence Accrual

The E-J models 1 and 2 may in fact be far too complex for what the decisional system can handle. The decisional system may be too burdened with the calculation of a mean to keep track of the variability of two distributions and of the effect of sample size on each. Stochastic counter models (Petrusic, 1992; Vickers, 1970) assume rather that Os perform a discrimination for each pair of observations and classify the result as evidence for one of the alternatives. The simplest example of such a model, Pike's discrete

accumulator (1968), hypothesises separate accumulators for each possible response. Each accumulator has a separate criterion amount of implicit evidence (r_A and r_B) that it requires for a response to be emitted (A and B). The first accumulator to exceed the criterion determines the response. The probability of evidence favouring the response A (R_A) is p . The probability of response A is given by

$$P_A = \sum_{j=1}^{r_B-1} (r_A + j - 1)! p^{r_A} q^j / j! (r_A - 1)! \quad \text{and the mean number of accruals}$$

conditional on A is given by

$$E(N | R_A) = \frac{1}{P(R_A)} \sum_{j=1}^{r_B-1} (r_A + j - 1)! p^{r_A} q^j / (j - 1)! (r_A - 1)!$$

Petrusic et al. (1995) provide an example of this model's ability to predict more accruals for memory than for perceptual judgements and a distance effect that is amplified in memory compared to perception.

The overt sampling process in this E-J paradigm could be viewed in similar terms. When an explicit comparison line is presented the O compares its subjective value to an observation drawn from a distribution of subjective values corresponding to the standard. When the comparison mean is close to the standard, most of the observations will be close to the standard leading to greater variability in the classification of accruals. As a result, more observations will be sampled before one of the accumulators reaches criterion. The role of instructions is actually quite complex but is here simplified for clarity (Petrusic, 1992). Emphasis on accuracy leads to higher criteria as does increased difficulty. Suppose for example that the criterion number of accruals is 5 for each accumulator; if the standard and the comparison mean are very different, then most of the

comparison observations will be very different from the standard observations and classification of the observation pairs will quickly favour one accumulator. In the case of relatively easy discriminations, it might take 6 observations before one of the accumulators reaches criterion. However, if the standard and the comparison mean are very similar, then comparison observations will be more similar to the standard observations and classification of the observation pairs will switch more often between accumulators. In the case of more difficult discriminations, 9 observations might be required for one of the accumulators to reach criterion. Basically then, more difficult decisions take more time because of the necessity for more observations.

In an E-J paradigm, difficulty varies as a function of the absolute differences between the stimuli to-be-judged, and also on the basis of the variability of the underlying sensory distributions. This general evidence accrual model of E-J predicts that both sample size and sampling times should increase as the stimulus difference decreases and as the variability of the underlying sensory distribution increases. As became evident for Thurstonian-based models 1 and 2 however, incorporating EV may reveal certain predictions that are unique to an E-J paradigm. Comparisons between an E-J paradigm and an implicit evidence accrual model are complicated by the fact that the means by which the decisional system attaches importance to implicit observations is unknown. Clearly, however, the role of monetary value in stochastic counter models is highly relevant. Petrusic and Jamieson (1978) showed that personally important decisions involving gambles ensure that each accrual continues to carry useful information over replications. In contrast, the informational content of accruals becomes increasingly

impoverished over time when judgements deal with less important choices, e.g. the aesthetics of geometrical shapes or opinions concerning college fraternities.

Expanded Judgement Using Ordinal Semantic Codes

According to semantic coding theory (SCT), ordinal semantic codes are the only stimulus information used in the actual comparative decision, although the underlying continuum from which the codes are generated is analogue in nature (Banks, 1977; Banks & Flora, 1977). As a result, when one of the stimuli is symbolically represented but the other is perceptually available, the perceptual stimulus must be ordinally encoded. Different predictions can be made with respect to the effect of d , comparison variability, serial-positions of the standards and semantic requirements, depending on how the comparison observations are used in the encoding process. In the first conceptualisation, comparison observations are averaged and then ordinally encoded for comparison with the standard's code. According to the second view, each observation is encoded separately and compared separately with the standard's code in an evidence accrual-like fashion (Petrusic, 1992). We now expand on each of these views.

1. Behavioural consequences of encoding the mean of the observations. The perceptual system calculates the mean of the perceptual observations, and then ordinally encodes the mean for comparison with the remembered standard. The process would start by sampling one observation and then encoding it for comparison with the memory standard. If the ordinal codes differ then discrimination has successfully occurred, sampling terminates, and a response can be emitted. If the ordinal codes are the same, the system would sample another observation, average it with the perceptual information pre-

associated with the first observation, and encode the average for comparison with the standard. This E-J differentiation process would continue until the standard and the comparison mean are encoded into different ordinal categories. The distance effect would reveal itself as a decrease in n as the ordinal difference between the standard and the true comparison increases, since for easier comparisons the mandatory time-consuming differentiation process would be engaged less often.

Since SCT precludes the use of analogue information in the comparison process, the distance effect on n should be the same regardless of the variance of the comparison distribution. SCE for n should not occur since, according to SCT, SCE occurs post-decisionally i.e. after the sampling process is over. Once discrimination has occurred, sampling stops and the decisional system attempts to match the instruction and the stimulus codes before outputting a response. As a result, n for both congruent and incongruent trials cannot be affected by instruction-stimulus.

Sampling time essentially takes place pre-decisionally during the encoding phase (Banks, 1977; Banks & Flora, 1977) and as such should be constant regardless of sample size, distance or variability. Sampling times reflect a constant amount of time needed to encode the stimuli and decide whether the codes are discriminable. Congruity will however be evident during the post-decisional phase between the last observation and the primary response also known as post-summative latencies. Since post-decisional, the magnitude of the SCE will not be influenced by stimulus characteristics such as distance and variability of the comparison distribution.

In summary, sample size will be affected by d , but not by σ_{jk}^2 or by semantic congruity since sample size is determined decisionally. Sampling times will be unaffected

by d , σ_{jk}^2 or SCE since they deal with predecisional encoding. Finally, post-summative latencies occur post-decisionally and will therefore reflect SCE but not d or σ_{jk}^2 .

2. Behavioural consequences of encoding each observation. In the second scenario, each explicit observation would be compared to the standard in an evidence accrual-like fashion after ordinal encoding. Each standard-observation accrual leads to a preliminary decision that is classified as evidence for one of the accumulators, i.e. comparison longer than the standard, or comparison shorter than the standard. Each accumulator has a response criterion. Observations are sampled until one of the criteria is exceeded at which point an overt response is made.

According to this accrual version of SCT, sampling times will exhibit a distance effect because they should reflect increasing use of differentiation as d decreases. Increasing the variability of the comparison distribution on the underlying analogue continuum should decrease sampling times since comparison observations are more likely to differ from the standard.

Sample size will exhibit a distance effect since, in an evidence accrual model, response criteria are sensitive to the difficulty of the discrimination here reflected in the probabilistic nature of proper code generation. When stimuli are more similar, categorisation of stimulus pairs will vary more between accumulators due to increased faulty encoding. Increasing the variability of the comparison distribution should also increase sample size since stimuli will be more likely to be erroneously categorised as d decreases. The distance effect should be magnified with increasing comparison variability.

Semantic congruity effects should be reflected in sampling times since each standard-comparison judgement also requires a match between instruction and stimulus codes. No congruity effects will be evident in sample size however, since the effect of incongruity on code translation has already been absorbed within each sampling time.

3. Accuracy. Serial-position effects (SPE) for response times and accuracy are confounded with the distance effect in most of the psychophysical studies that have examined the viability of semantic coding theory (Banks, 1977; Banks, Mermelstein & Yu, 1982; Petrusic et al., 1995). Baranski and Petrusic (1992) showed however that the serial-position effect obtained with the method of constant stimuli results from impaired discriminability for middle items relative to end-point items. The method of constant stimuli provides a stronger statement concerning the nature of the SPE since it precludes explicit comparisons between standards.

Baranski and Petrusic (1992) provide a more general version of SCT which permits the generation of psychometric functions with non-zero slope. According this view, the O only needs general hypotheses about the location of the true comparison means and of the standard on the underlying analogue continuum, and then ordinally codes each comparison stimulus relative to its standard. This categorisation process is probabilistic depending on the ease with which the comparison stimuli can be ordered. Comparison lines that differ more from the standard will lead to a greater probability of correct coding. Variations in accuracy follow in that comparison stimuli closer to the standard will be more likely to be misclassified. In situations where several standards are used and the ratios of comparison stimuli to standards are held constant between sets, but absolute differences increase as the standard increases, then SCT predicts that Weber fractions

will decrease as the standard increases. Baranski and Petrusic (1992) found a clear inverted-U serial-position effect in memory contradicting SCT theories prediction.

4. Problems with SCT. According to SCT, variations in sample size and sampling time occur as a result of a process of differentiation which recodes the stimuli until they are classified into different ordinal categories. Discrimination will of course occur as soon as stimuli are encoded into different categories. Despite the fact that the amount of differentiation is related to discriminability, the absence of analogue information during the actual discrimination process makes SCT very impractical. Banks (1977) is very specific about stating that the precision of the analogue information available in the underlying data base is very limited. The crude nature of the underlying continuum makes it unlikely that SCT can properly account for interval level variations in stimuli, since an interval scale requires the homogeneous concatenation of intervals along the continuum (Krantz, Luce, Suppes & Tversky, 1971). Differentiation is therefore likely to fail in an E-J paradigm where comparison observations vary along a finely graded continuum.

More importantly, research has shown that the various properties of the response time-response probability relations can only be accounted for by models that posit a decisional locus. As such, SCT is inadequate because its decisional locus uses ordinal codes that are insensitive to interval level information demonstrated to exist in perceptual and memory representations (Petrusic & Baranski, 1992; Petrusic et al., 1995; Petrusic, Baranski, Kennedy, & Aubin, 1996). Finally, since semantic coding theory (Banks, 1977) was designed to explain judgements involving remembered stimuli and not perceptual stimuli, it is inadequate as a general theory of comparative judgement.

Method

Subjects. 32 Carleton University undergraduate students were recruited for one two-hour session in fulfilment of course credit requirements. Os were randomly assigned to each of the four cells of a 2x2 factorial design. The four cells consisted of two types of standards counterbalanced with two levels of comparison variability. All Os had normal or corrected to normal vision, as verified by an eye chart examination, and were naive with respect to the nature and purpose of the experiment.

Apparatus. The stimuli were presented on a Zenith ZCM-1492 high-resolution video monitor (640 x 480) incorporating Zenith's Flat Technology CRT. A 33 mhz 486-dx IBM-PC controlled event sequencing, randomisation, and the recording of responses and response times (RT). All responses were made on a response panel using the index finger of the preferred hand. A small desk lamp, positioned behind the video monitor, provided sufficient light for responding without interfering with the visual display.

Stimuli and Design. Stimuli were horizontal line lengths measured in pixels. Each pixel measured .3828 mm in width and .3854 in height. Four horizontal line lengths measuring 60, 120, 180 and 240 pixels (22.97mm, 45.94mm, 68.90mm, and 91.87 mm, respectively) served as four standards (S_k). Instead of actual comparison lines, as would be the case in the traditional method of constant stimuli (Baranski & Petrusic, 1992), five comparison distributions of line lengths were defined such that μ_{jk} for each of the five distributions differed from its respective standard by -10%, -5%, 0%, 5% and 10%. The comparison distribution was Gaussian with $N(\mu_{jk}, \sigma_k)$, where μ_{jk} is the j th comparison mean

associated with the k th standard, and σ_k is a constant for all μ_{jk} 's for a given standard.

Once s_k , μ_{jk} and σ_k were determined, each of the I successive observations requested by the O were randomly selected from the comparison distribution such that

$x_i = z_i\sigma + (s + ks)$. In the latter equation, x_i is the i th observation requested by the O, z_i is a value randomly selected from a standardised Gaussian distribution with $N(0,1)$, $\sigma = \sigma_k$, s is a particular standard s_k , k is the percent difference between the standard and the comparison mean and equals either -10%, -5%, 0%, 5% or 10%, and $(s + ks)$ is the mean of the comparison distribution or μ_{jk} .

Comparison μ_{jk} 's measured 54, 57, 60, 63 and 66 pixels (20.67mm, 21.82mm, 22.97mm, 24.12mm and 25.26mm, respectively) for standard 1; 108, 114, 120, 126 and 132 pixels (41.34mm, 43.64mm, 45.95mm, 48.23mm, and 50.53mm, respectively) for standard 2; 162, 171, 180, 189 and 198 pixels (62.01mm, 65.46mm, 68.90mm, 72.35mm, and 75.79mm, respectively) for standard 3; and, 216, 228, 240, 252 and 264 pixels (82.68mm, 87.28mm, 91.87mm, 96.47mm, and 101.06mm, respectively) for standard 4.

σ_k were generated such that $\frac{s_k}{\sigma_k} = c$. Since the $s_k - \mu_{jk}$ percentage differences for the 5 J comparisons within each of the four sets is the same for all standards

i.e. $\frac{s_k - \mu_{jk}}{s_k}$ for $j=1$ to 5, and, since $\frac{s_k}{\sigma_k} = c$ between comparisons sets, then the a

priori difficulty of the comparisons defined by $z = \frac{|s_k - \mu_{jk}|}{\sigma_k}$ is held constant between

standards both in terms of ratios and differences. As a result, $\sigma_k = \{6, 12, 18, 24\}$ for standards 1 to 4 respectively.

A central motivation for this experiment was to examine the impact of σ_k on the number of observations requested and the rate at which they are sampled. Variability in the Gaussian line distributions was manipulated between two groups of Os according to the expression $m\sigma_k$, where $m=1, 3$. When $m=1$, then $\sigma_k = \{6, 12, 18, 24\}$ as listed above and the resulting Gaussian z-scores for the five comparison-standard pairs within each standard set are -1, -0.5, 0, +0.5, and +1.0. When $m=3$, then $\sigma_k = \{18, 36, 54, 72\}$ for standards 1 to 4 respectively and the resulting Gaussian z-scores for the five comparison-standard pairs within each standard are -0.33, -0.17, 0.0, +0.17, and +0.33. This experimentally induced increase in stimulus variability was intended to lead to changes in accuracy and decision times that parallel those differences found between perceptual and symbolically remembered stimuli if poorer performance with remembered stimuli is due to a noisier internal representation.

Comparison lines were successively presented in the same vertical location just below the standard. The centre of each comparison line was randomly staggered with respect to the centre of the standard to avoid endpoint strategies.

Half of the Os within each of the two variability manipulations were presented with perceptual standards. The other half of the Os (memory condition) were presented with the same standards as in the perceptual condition but only by way of consonant-vowel-consonant triads (CVCs). Four different groups of Os were formed on the basis of this 2 x 2 factorial design: Perception - 1 σ , Perception - 3 σ , Memory - 1 σ , and Memory - 3 σ .

Payoffs. Given the exploratory nature of this investigation, money did not actually exchange hands between the experimenter and the O. Research has shown that imaginary payoffs can have the same effect as actual monetary gains and losses in providing the O

with implicit instructions (Edwards, 1954a). Specifically with respect to the original expanded judgement involving the mean of packs of cards, Irwin and Smith (1957) found that imaginary payoffs induced the expected changes in sample size as a function of difficulty and variability manipulations. Here imaginary payoffs were used as follows. A correct comparison was worth a basic value of \$0.50. An incorrect comparison cost a basic value of \$0.50. Each line requested cost one penny. A correct judgement was actually worth \$0.50 less a penny for each observation requested. An incorrect response resulted in a loss of \$0.50 in addition to a penny paid out for each observation requested. For example, the value gained with a correct response with a sample size of 10 lines was $\$0.50 - \$0.10 = \$0.40$. The value lost with an incorrect response with a sample size of 10 lines was $\$0.50 + \$0.10 = \$0.60$.

Procedure. The O was welcomed into the lab by the experimenter who then asked the O to fill in the informed consent form. Afterwards, the experimenter administered a visual acuity test using a conventional eye chart with both of the O's eyes open since the comparative judgement task was performed with both eyes. This procedure was used as a general check to ensure that the O's eyesight was in fact normal or corrected-to-normal. Os were asked to wear glasses or contacts if they are prescribed. In the experimental room, the O was seated approximately two feet from the video monitor. The following instructions were given:

Instructions for the perceptual condition. "In a traditional experiment, one of two words ('SHORTER' or 'LONGER') appears as an instruction at the top of the computer screen followed a moment later by two horizontal lines positioned one above the other [write example on blackboard]. In that kind of task you would be asked to select the

longer or the shorter of the two lines. This experiment is slightly different. In this case, the line at the bottom is not actually presented. Instead it serves as the true average of a very large population of lines that differ in length. Your task is to compare the stationary line with the true average of the population of lines and decide which is longer or which is shorter depending on the instruction.

Not all of the population lines will be presented. Rather a smaller number of lines, called a sample, will be randomly selected from the entire population and presented one after the other in the same spot. Using the average of the sample as your best guess as to what the population average is, you will then decide which is longer or shorter, the population average or the stationary line. The population lines will be presented one after another in the same spot just below the stationary line. The number of lines selected from the population is under your control. Select as many lines as you need to make the best possible decision. It is important that you remember that the more lines you select, the better your estimate of the population mean and the greater your chances of being correct [draw example illustrating the effect of sample size on the sampling distribution of the mean].

Accuracy is very important in this experiment because I'd like you to pretend that you can win or lose money on each trial. Specifically, a correct comparison between the stationary line and the population mean wins you \$0.50 and an incorrect comparison costs you \$0.50. Since each line you select provides information about the population average and so improves your chances of being correct, each line will cost one penny. So although increasing the number of lines improves your chances of winning 50 cents, the cost associated with each line you select decreases the amount of money you can win. A

correct judgement earns you \$0.50 less a penny for each observation requested. An incorrect response costs you \$0.50 in addition to a penny for each observation requested. For example, say you sample 10 lines and your decision is correct then you would make $\$0.50 - \$0.10 = \$0.40$. With 10 lines, an error would cost you $\$0.50 + \$0.10 = \$0.60$.

A large number of lines increases your chances of being correct and of gaining money, but you pay for the information. A small number of lines keeps your information costs down, but increases the chances that you will be incorrect and lose 50 cents in addition to information costs. As a result, it is important that you keep this trade-off in mind throughout the experiment. I will ask you to avoid the use of strategies such as always using the same number of lines for the entire experiment. Rather, determine the number of lines you need to make the best decision separately on each trial. It is important to note that the length of the stationary line and the length of the population average will vary from trial to trial. Sometimes they will be more similar than other times.

The message you see on the screen, 'press any key to continue' will recur several times during the course of the session in order to give you breaks. The message can be cleared to start the experiment by pressing any letter or number on the keyboard using your non-preferred hand. Otherwise, all responses are made using these the buttons on this response panel. I will ask you to use the index finger of your preferred for indicating all your responses.

Each trial begins by pressing this middle key with your index finger [the middle button sits in a three-button configuration arranged in the shape of an inverted triangle]. Once you have done so, keep your index finger over the middle button until you have

your decision was correct by pressing, again with your index finger, one of the six confidence buttons at the top. The button at the far left labelled '50' indicates that you were 50% confident that you were correct i.e. you were guessing. The button at the far right labelled '100' indicates that you were 100% confident that you were correct i.e. absolute certainty. The percentages in between indicate different confidence levels. On some occasions you might feel that you pressed the wrong response button. In that case, instead of pressing a confidence button, press the middle button, to indicate that you think you made a mistake. After the confidence rating, the computer screen will then give you feedback as to whether your comparison between the stationary line and the population average was correct. The screen will then be blank waiting for you to press the middle button again to start the next trial.

It is very important that you maintain a high level of accuracy throughout the session. As a result, two types of breaks are included. Experimental trials will be divided into four groups. After each of these blocks, the screen will read 'REST' and then 'press any key to continue'. At REST, I would like you to physically get up and leave the room even if for just a moment. Feel free to take a short walk to stretch out your legs. After returning to the testing room, close the door completely and give your eyes a minute to readapt before you continue. You will get three major breaks between blocks. I would also like you to take momentary breaks every few trials. You can do this by delaying pressing the middle button which as you will recall is used to start each trial. The emphasis in this experiment is on accuracy. Take the time you need to make the best possible decision."

Instructions for the memory condition. A session began with paired-associate training following by binary comparisons.

1. Paired-associate training. " This task will teach you to associate a word with each of four different line lengths. After this task you will then be given another task. On each trial of the second task one of the four names will presented on the screen along with an actual horizontal line. You will then have to recall the line in memory associated with the name and compare its length with the actual lines presented on the screen.

On each trial of this first task, one of four horizontal lines will be presented in the middle of the screen with four nonsense syllables below each encased within a rectangle [draw an example]. The nonsense syllables used in this experiment (we refer to them as CVCs) are three letter words that are pronounceable but have no particular meaning outside of this experiment. Your task is to select the CVC that corresponds to the presented line using the mouse. By moving the mouse from side to side you can highlight the CVC you want, then click on the left mouse button to register your selection. The screen will give you feedback as to whether or not you were correct. At first your responses will be random but you will soon learn which CVC stands for which line.

If your selection is correct, the rectangle containing the selected CVC will remain lit up for 2 seconds and then the next trial will begin. If your selection was incorrect, the computer will highlight the correct CVC for 2 seconds and then go on to the next trial. There is a 5 second time limit for making a response. If you take longer than 5 seconds a message at the bottom of the screen will tell you that time is up. The computer will then highlight the correct CVC for 2 seconds and then go on to the next trial. The time-out should not alarm you but is simply meant to ensure that the task progresses at a constant

rate. Although you will not know at first which CVC corresponds to which line, I would like you to simply select CVCs at random from the first trial just to get you into the swing of things. You will also notice that the CVCs are randomly moved around on each trial but the same CVC will always be associated with the same line.

It is important that you concentrate on remembering the sizes of the lines and their corresponding names because that information will be needed for the second task. The learning phase takes about 10 minutes.

You may place the mouse so that you feel comfortable when you make your selection. I will be here for the first few trials to answer any questions and then leave you to complete the task. When you have finished the learning phase, the program will simply default back to the DOS C: prompt. At that point, please come and get me in room C563. Do you have any questions?

2. Binary comparison task. "In a traditional experiment, one of two words ('SHORTER' or 'LONGER') appears as an instruction at the top of the computer screen followed a moment later by a CVC and a horizontal line just below that [write example on blackboard]. In that kind of task you would be asked to decide which is longer or shorter, the memorised line associated with the CVC or the physical line. This experiment is slightly different. In this case, the line at the bottom is not actually presented. Instead it serves as the true average of a very large population of lines that differ in length. Your task is to compare the CVC with the true average of the population of lines and decide which is longer or which is shorter depending on the instruction.

Not all of the population lines will be presented. Rather a smaller number of lines, called a sample, will be randomly selected from the entire population and presented one

information costs. As a result, it is important that you keep this trade-off in mind throughout the experiment.

I will ask you to avoid the use of strategies such as always using the same number of lines for the entire experiment. Rather, determine the number of lines you need to make the best decision separately on each trial while keeping the trade-off in mind. It is important to note that the length of the CVC line and the length of the population average will vary from trial to trial. Sometimes they will be more similar than other times.

The message you see on the screen, 'press any key to continue' will recur several times during the course of the session in order to give you breaks. The message can be cleared to start the experiment by pressing any letter or number on the keyboard using your non-preferred hand. Otherwise, all responses are made using these the buttons on this response panel. I will ask you to use the index finger of your preferred for indicating all your responses.

Each trial begins by pressing this middle key with your index finger [the middle button sits in a three-button configuration arranged in the shape of an inverted triangle]. Once you have done so, keep your index finger over the middle button until you have made your decision regarding the stimuli presented on the screen. After pressing the middle button, one of two instructions (SHORTER or LONGER) will be presented at the top of computer screen and will remain there until you end the trial by making a primary response. Two seconds after the instruction has appeared, the CVC will be presented. On half the trials, the CVC will be presented in the top position. On the rest of the trials, the CVC will be presented in the bottom position. The CVC will remain on the screen until the end of the trial. One second after the appearance of the CVC the first line of the

sample will be presented for about $\frac{1}{3}$ of a second. The duration of a sample line on the screen is predetermined so keeping the middle button down will not allow you to examine a sample line for longer than $\frac{1}{3}$ of a sec. However, you do have complete control over the number of lines presented and the rate at which they are presented. Each time you press the bottom middle button another sample line will be presented. The speed at which you press the middle button determines the rate at which sample lines are presented. Keep in mind that the sample lines are randomly selected from the population of lines so that the actual lengths are not appearing in any particular order. Also keep in mind that the difference in length between the CVC line and the population average will change being more similar in some cases and more different in others. Also keep in mind the effect of sample size on the accuracy of your judgement as well as the associated costs.

Once you have made your judgement, stop sampling lines, and using the same index finger indicate your decision by pressing the left button if the item at the top is consistent with the instruction or the right button if the item at the bottom is consistent with the instruction. After making this response, you then to indicate how confident you feel that your decision was correct by pressing, again with your index finger, one of the six confidence buttons at the top. The button at the far left labelled '50' indicates that you were 50% confident that you were correct i.e. you were guessing. The button at the far right labelled '100' indicates that you were 100% confident that you were correct i.e. absolute certainty. The numbers in between indicate different confidence levels expressed in percentages. On some occasions you might feel that you pressed the wrong response button. In that case, instead of pressing a confidence button, press the middle button, to

indicate that you think you made a mistake. After the confidence rating the computer screen will then give you feedback as to whether your comparison between the CVC and the population average was correct. The screen will then be blank waiting for you to press the middle button again to start the next trial.

It is very important that you maintain a high level of accuracy throughout the session. As a result, two types of breaks are included. Experimental trials will be divided into four groups. After each of these blocks, the screen will read 'REST' and then 'press any key to continue'. At REST, I would like you to physically get up and leave the room even if for just a moment. Feel free to take a short walk to stretch out your legs. After returning to the testing room, close the door completely and give your eyes a minute to readapt before you continue. You will get three major breaks between blocks. I would also like you to take momentary breaks every few trials. You can do this by delaying pressing the middle button which as you will recall is used to start each trial. The emphasis in this experiment is on accuracy. Take the time you need to make the best possible decision."

Results

All Os were included in the analyses. For each O, sampling times less than 200 ms or greater than 10000 ms were excluded from calculations involving latencies. Each O received 192 trials. Eighty percent of the Os (13/16) had one sample time (less than 200 ms) excluded on .5% to 25% of the experimental trials. Only 25% of the Os had two sample times excluded on one or two trials. Fifty percent of the Os had more than two sample times excluded on one or two trials. Forty four percent of the Os (7/16) had one sample time excluded on .5% to 3.1 % of the experimental trials because they were greater than 10000 ms. Only two Os had two sample times excluded on one trial. Post-summative latencies shorter than 200 ms and than 30000 ms were also excluded. Forty two of the Os (4/16) had one post-summative latency excluded on .5% of the experimental trials (n=192) for because they was greater than 30000 ms. None were shorter than 200 ms.

Table 1 provides a summary of the predictions made by Thurstonian model 1, Thurstonian model 2, semantic coding theory 1, semantic coding theory 2 and Pike's (1971) simple accumulator. Predicted effects confirmed by the results of the experiment are in bold. Note, however, that this table is overview. Several of the observed effects are not represented in this table because they were specific to certain cells of the design. In general, these latter effects were confined to the perceptual condition with low comparison variability suggesting that increased comparison variability and placing the standard in memory often overwhelm the decisional system. As will become evident, Os probably did take the average of the perceived observations, as the instructions required, and stopped sampling when error variability was optimally low. However, with respect to

the decision process, individual observations were probably retrieved from memory and then accumulated after-the-fact in an implicit evidence accrual-like process. Before examining these effects, let us consider the viability of the expanded judgement paradigm as an instrument for investigating psychophysical properties for remembered line lengths.

Viability of the expanded-judgement paradigm for binary comparisons.

1. The effect of memory, comparison variability and serial-position on the discriminability of percepts. The slope of the psychometric function is an index of stimulus discriminability that reflects the change in standardized accuracy as the distance between the comparison stimulus and the standard changes. For a given set of comparison-standard stimuli, a steep slope results from relatively large changes in accuracy and indicates greater discriminability compared to a shallower slope. The psychometric function is used to obtain the just-noticeable-difference (JND) which according to Fechnerian tradition is the minimal change required in a stimulus for judgements to be correct 75% of the time (Fechner, 1860). The Weber fraction is a convenient standardisation of the JND into a percentage difference that permits direct comparisons between stimulus sets. Small WFs represent greater discriminability, i.e. less of a stimulus change is required for judgements to be correct 75% of the time. Slopes were used instead of Weber fractions (WF) in the analyses because they provide the same information as WFs and can be set to zero when there is a complete lack of discriminability. When slopes are negative, zero or close to zero, WFs are either infinite or cannot be computed. For descriptive purposes, however, slopes were converted to WFs

Table 1

Predicted Behaviour of Sample Size (observations), Sampling times and Post-summative Latencies (PS Latencies) for Thurstonian Model 1, Thurstonian Model 2, Semantic Coding Theory 1 (SCT 1), Semantic Coding Theory 2 (SCT 2) and Pike's Simple Accumulator (PIKE) [H=Hard, E=Easy, M=Memory, P=Perception, Hi=High comparison variance, Lo=Lo comparison variance. Predicted effects are in bold]

MEASURE	THURSTONE 1	THURSTONE 2	SCT 1	SCT 2	PIKE
Observations	H = E (for Per/LoVar) H < E (other 3 cells) M < P Hi > Lo	H > E (for LoVar) H < E (for HiVar) M = P Hi > Lo	H > E M vs P (n/a) Hi = Lo	H > E M vs P (n/a) Hi > Lo	H > E M > P Hi > Lo
Sampling times	H = E M = P Hi = Lo	H = E M = P Hi = Lo	H = E M vs P (n/a) Hi = Lo	H > E M vs P (n/a) Hi < Lo	H > E M > P Hi > Lo
PS Latencies	H = E M = P Hi = Lo	H = E M = P Hi = Lo	H = E M vs P (n/a) Hi = Lo	H = E M vs P (n/a) Hi = Lo	H = E M = P Hi = Lo

for use in graphical representations in order to provide a more intuitive representation of discriminability (see Footnote 5 in Baranski & Petrusic, 1992). Discriminative sensitivity was examined by performing a three-way ANOVA on the slopes of the psychometric functions. Standard type (MEM vs PER) and the variability of the comparison distribution (low versus high) were entered as two between-subjects factors. Standard was the within-subjects factor. Greenhouse-Geiser adjusted degrees of freedom were used to calculate significance levels for within-subjects effects, but actual degrees of freedom are reported along with F values and significance levels.

In clear replication of non-expanded judgement experiments using the method of constant stimuli (Baranski & Petrusic, 1992; Petrusic, Baranski & Harrison, 1993), perceptual standards were significantly more discriminable than remembered standards [$F(1,12)=13.88$, $p=.003$] as shown by the smaller mean Weber fraction for perception compared to memory (.06 and .10, respectively).

Compared to high comparison variability ($WF=.10$), low comparison variability ($WF=.06$) resulted in significantly improved discriminability [$F(1,12)=18.58$, $p=.001$]. A significant two-way interaction [$F(1,12)=5.71$, $p=.034$] indicated that the improvement in discriminability for perception over memory was nearly twofold when comparison variability was low (WFs of .05 and .09), but minimal when comparison variability was high (.08 vs .11, for perception and memory respectively). Note that the WF for memory with low comparison variability (.08) was nearly identical to that for perception with high comparison variability (.09). Note also that the marginal WF for high comparison variability (.10) was identical to the marginal WF for remembered standards, and that the marginal WFs for low comparison variability and for perceptual standards were also

identical (.06 each). Reduced discriminability as a result of increased comparison variability is particularly meaningful in that it provides indirect evidence for the view that placing a percept in memory affects discriminability by increasing the variability of the underlying representation (Holyoak & Patterson, 1981; Johnson, 1991). Comparison variability makes perception look like memory.

A significant two-way interaction between standards and standard type indicated the presence of a serial-position effect (SPE) in memory but none in perception [$F(3,36)=4.58, p=.0214$]. Figure 6 provides insight into the nature of this effect by displaying changes in WFs across standards for each of the four cells of the standard type by comparison variability combination. Simple interaction effects between comparison variability and standard were non-significant for perception but significant for memory [$F(3,18)=5.06, p=.02$]. Trend analyses over standards performed for each of the standard type by comparison cells found a significant cubic trend for the memory condition [$F(1,3)=103.34, p=.002$] with low comparison variability but no significant trends for the other three cells, showing that the serial-position effect for memory comes entirely from the low comparison variability condition. The lack of a SPE for memory judgements with high comparison variability was initially surprising

Figure 6. Discriminability of standards (Weber fractions) for each of the four standard type by comparison variability combinations.

since the distinctiveness of the outer standards should always lead to improved discriminability relative to the inner standards (Johnson, 1991). One would expect an additive effect of comparison variability. Note however that all four WFs are equally bad suggesting that they are either infinite or cannot be computed. For descriptive purposes, however, slopes were converted to WFs for use in graphical representations in order to provide a more intuitive representation of discriminability.

The point of subjective equality (PSE) provides the value on the underlying length continuum at which both the standard and the comparison stimulus are subjectively equal. The difference between the PSE, which is the subjective standard, and the actual physical standard provides an index of the location of the subjective magnitude on the underlying continuum. A signed standardised index was obtained in order to make comparisons between standards (see Footnote 5 in Baranski & Petrusic, 1992). A negative value results from an underestimation of the physical standard. A positive value from an overestimation of the physical standard. An ANOVA on the absolute standardised values revealed no significant differences between perception and memory or between any of the other factors. In fact the mean absolute deviations were only slightly larger for memory than for perception (1.1 versus 1.7). An inspection of figure 7 reveals a tendency for PSEs to be more variable for memory than for perception. One subject in the memory condition had to be excluded because the slope of the psychometric function was 0. This result is anomalous in that PSEs for memory are typically more variable than those for perception (Baranski & Petrusic, 1992) suggesting that the underlying memory representation is more variable than its perceptual counterpart.

Taken together, these data provide evidence that the expanded judgement paradigm is a viable means of investigating the processes involved in judgements with remembered magnitudes. The same effects observed under traditional psychophysical conditions are also replicated in this expanded judgement version of the method of constant stimuli. First, remembered standards are less discriminable than their perceptual counterparts. Second, Weber fractions for perception with low comparison variability (.05) were very similar to those reported by Baranski and Petrusic (1992) for set size three. In their Experiment 1, WFs ranges from .03 to .051 for an average of about .045. Third, explicitly manipulating the variability of the comparison distribution makes discriminability memory-like in that changes in WFs directly induced by manipulating comparison variability mirror overall differences in WFs between perceptual and remembered counterparts of the same physical stimuli. Finally, the discriminability of expanded judgements with remembered standards exhibits a SPE in replication of the same effect that has been well-documented in studies using the method of constant stimuli (Baranski & Petrusic, 1992; Petrusic, Baranski & Harrison, 1993). The lack of a difference in PSE variability between perception and memory is surprising however, since this effect has been so robustly demonstrated in psychophysical experiments. Increased variability in PSEs for memory suggests a more variable internal representation, an argument that is typically invoked to explain reduced discriminability for remembered stimuli.

Figure 7. Standardised differences between PSEs and physical standards for each subject under perceptual and remembered judgements.

2. The effect of memory, variability and serial-position on the time to make expanded comparative judgements and on the number of observations taken. Latencies for the first sample (RT1), the last sample (lastRT) and samples on average (mnrt), as well as post-summative latencies (psrt), number of observations (obs) and accuracy (correct) were analysed separately in 4-way ANOVAs with standard type and comparison variability entered as between-subjects factors. Distance and standard were entered as within-subjects factors. Greenhouse-Geiser adjusted degrees of freedom were used to compute significance levels, however, actual degrees of freedom are reported. Effects are reported in separate sections.

There were no significant main effects of standard type, comparison variability or serial-position on sampling times, post-summative latencies or on the number of observations sampled. The absence of main effects is unfortunate since these are necessary to differentiate between models. At this main effect level of analysis, nothing seems to work. If explicit sampling reflects an implicit accrual process, then the memory-perception effect should have revealed itself in increased sampling for comparisons with remembered standards relative to perceptual standards. The reverse should have been true if a Thurstonian model were at work. Increased comparison variance should also have led to increased sampling. The lack of a SPE seems counterintuitive given the well-documented effect of discriminability on response times in the context of the method of constant stimuli. Kennedy (1990) and Leach (1992) showed that the shape of response times across standards tracks changes in WFs. Specifically, RTs across remembered, but not perceptual, standards form an inverted-U. The lack of a SPE for observations in this experiment may have resulted from the

crudeness of this measure. Explicit observations may therefore not behave like implicit observations hypothesised to exist in the evidence accrual process. Alternatively, Os may not be processing these observations as seriously as they would if real monetary gains and losses were associated with the accuracy of judgements and more importantly, if costs were associated with each observation. However, paralleling Kennedy's (1990) findings, sampling times and post-summative latencies should have been sensitive enough to reflect the relative difficulty with which stimulus information for the middle standards must be retrieved for remembered judgements to occur. The SPE in this experiment clearly demonstrates the presence of differential discriminability across standards.

Distance effects in accuracy, sample size and sampling times.

The distance factor was created by manipulating the physical difference between the expectation of the comparison distribution and the standard but keeping these constant between comparison distributions. For simplicity, results for comparison distributions 1 and 5 were collapsed since their expectations were equidistant from the standard. Likewise, results for distributions 2 and 4 were collapsed. Results for the comparison distribution equal to the standard were removed since these are subject to non-systematic biases. Comparison distributions 1 and 5 were twice as far from the standard than were distributions 2 and 4. The resulting distance factor had two levels that can be labelled as

"short" and "long". Table 2 summarizes the significant effects. Only those relevant to the distance effect are reported in this section.

1. Distance effects on accuracy. A significant distance effect emerged for accuracy [$F(1, 12)=118.1, p=.0001$]. Accuracy was greater for comparison means 1 (79.9%) and 5 (75.6%) that were further away from the standard than for comparison means 2 (67%) and 4 (64.4%). However, the effect of distance interacted significantly with comparison variability [$F(1,12)=6.84, p=.0225$]. Panel A of figure 8 illustrates the difficulty effect for accuracy. The distance codes are effect codes and are appropriate in this case since comparison means differed from the standard in steps of 5% of the length of the standard. Starting on the left of the graph, note the decrease in accuracy for both plots as the length of the comparison mean increases towards that of the standard from a distance code of -2 to -1. Starting on the right side of the graph, a similar drop in accuracy occurs as the length of the comparison mean decreases towards that of the standard from a distance of 2 to 1. The interaction between distance and comparison variability is explained by the average increase in accuracy from the nearest to the farthest comparison means which was 13.8% for low comparison variability, but only 10% for high comparison variability. This effect is the same for perception and for memory. Clearly high comparison variability impairs performance accuracy by introducing error into the system. Comparison variability impairs the ability to retrieve perceptual information thus reducing the effect of distance when comparison variability is high.

2. Distance effects on the number of observations. The number of observations was reliably sensitive to distance [$F(1,12)=4.69, p=.0512$]. Panel B of figure 8 illustrates that changes in the number of observations tracked those expected for response times

measures. Petrusic (1993) provides a short review of studies showing that response times are very sensitive to physical differences and ratios between physical stimuli. The number of explicit observations selected parallels effects seen in response time experiments and provides some evidence for the accrual of evidence associated with implicit observations. Thurstonian model 1 is disconfirmed because it predicted no distance effect for perceptual comparisons with low comparison variability, and a reversed distance effect for the other 3 standard-type by variability combinations. Thurstonian model 2 is disconfirmed because it predicted a distance effect for low comparison variability and a reversed distance effect for high comparison variability. Pike's accumulator and both SCT models are the only ones to receive confirming evidence. Memory did not enhance this effect however as would have been expected on the basis of traditional memory psychophysics where comparisons with remembered stimuli have typically accentuated the effect of distance (Moyer & Bayer, 1976) presumably because they require more accruals (Baranski & Petrusic, 1992; Petrusic, 1993).

3. Distance effects on sampling times and post-summative latencies. The main effect of distance was not reliable but the interaction of distance with comparison variability was [$F(1,12)=5.30$, $p=.04$]. The accumulation of explicit evidence should have resulted in clear distance effects for latencies throughout the sampling process if each comparison line were in fact compared individually to the standard. This would not be the case, however, if the O were calculating a subjective average as is suggested by the Thurstonian models. In the latter scenario, a distance effect would emerge only when the subjective average is compared to the standard in the post-sampling phase. The findings

in this experiment seem to agree with the Thurstonian models. Distance effects were completely absent from the sampling process and did not begin to emerge until the end but only for the judgements with low comparison variability. This suggests that the decisional system began to calculate the mean observation length towards the end of the sampling process. Panel C of figure 8 clearly shows that the distance effect was completely absent from judgements with high comparison variability but evident for low comparison variability. Under low comparison variability, latencies were on average 1030.7 ms for the closer comparison means and 942.3 for the farther.

Finally, the magnitude of the distance effect for post-summative latencies depended on both standard type and comparison variability. Figure 9 shows that this significant three-way interaction [$F(1,12)=16.18$, $p=.0017$] is explained by the massive distance effect for perception with low comparison variability but its total absence from the other three cells. The presence of a distance effect for post-summative latencies is consistent with the Thurstonian models although this conclusion must be tempered by its presence in only one of the four cells of the design. The other models, however, predict no distance effect whatsoever because post-summative latencies simply represent a constant output time.

Semantic congruity with sample size, sampling time and accuracy.

Semantic congruity effects failed to emerge for the number of observations, sampling times and accuracy. Congruity should occur at the macro level if the basis for the effect is the location of line stimuli on the same length continuum (Petrusic, 1992). After collapsing across comparison means and calculating the value of the dependent

Figure 8. Distance effects for accuracy (panel A), number of observations (panel B) and the last sampling time (panel C) across comparison-standard differences represented by effect coding.

A) PERCEPTION - LOW VARIANCE

B) MEMORY - LOW VARIANCE

C) PERCEPTION - HIGH VARIANCE

D) MEMORY - HIGH VARIANCE

Figure 9. Distance effects for post-summative latencies for each of the four standard type by comparison variability combinations.

measure for each standard with each instruction, congruity should emerge for standard 1, which is relatively short, and for standard 4 which is relatively long. At the micro level, values of the dependent measure are calculated for each standard-comparison pair for each instruction. If congruity occurs at the micro level, then sub-ranges are being defined on the length continuum that are bounded by the values of the upper and lower comparison means for each standard-comparison set. As a result, congruity should manifest itself in the smallest and largest comparison means within each set. None of the above cases showed congruity. This is surprising given the well-documented occurrence of the congruity effect especially with remembered stimuli (Banks & Flora, 1977; Petrusic, 1992; Petrusic et al., 1995).

Discussion

This experiment provides qualified support for the expanded judgements as a viable psychophysical paradigm. In replication of findings using the method of constant stimuli, discriminability was poorer for remembered standards relative to their perceptual counterparts, and a serial-position effect occurred in memory but none in perception. The serial-position effect occurred for remembered standards in the low comparison variability condition and not in the high because the standards in the latter condition were equally poorly discriminable. The increase in the variability of the comparison distribution magnified the presence of error in the decisional system and overwhelmed the comparison process. This result is important because it shows limits to the decisional systems ability to handle noise. When enough noise is infused into the perceptual system, characteristics that normally differentiate between perceptual and remembered

representations are no longer measurable. Attesting to the importance of variability, increasing the variance of the comparison distribution made perceptual discriminability look like memory. Some findings were anomalous. Specifically, no serial-position effect was found for sampling or post-sampling latencies. PSEs were not significantly more variable in memory compared to perception although descriptively there seemed to be a difference.

Distance effects emerged for accuracy, for the number of observations and for the tail end of the sampling process. These results were somewhat different than expected. Clearly in an evidence accrual model, the comparison process takes place during the accrual process both in terms of the number of accruals taken and the time with which they are sampled (Petrusic et al., 1995). The accrual process seems to reveal itself differently in expanded judgements. Specifically, it would seem that the observer collects a number of observations, stops, and then recollects them in an evidence accrual-like fashion for comparison with the standard. This is reflected in the presence of a distance effect for post-summative latencies as well as for the number of observations.

The distance effect has been extensively documented in perceptual psychophysics (for a review see Petrusic, 1993) and in memory psychophysics where it has been labelled as the “symbolic” distance effect (Moyer, 1976). As two stimuli approach each other in value, judgements become slower and more prone to error. The distance effect for response times is more sensitive than for percent correct and continues to mirror differences and ratios even for supraliminal stimulus pairs (Munsterberg, 1894, Petrusic & Jamieson, 1979). The distance effect for judgements with remembered stimuli is typically amplified compared to the effect obtained with their perceptual counterparts

(Moyer & Bayer, 1976). The presence of distance effects provides partial vindication for the viability of the expanded judgment paradigm by providing yet another parallel with traditional non-expanded judgement psychophysics and with accrual theories that predict these effects (Petrusic, 1992; Vickers, 1970, 1979). The distance effect suggests that decisional processing involves comparative judgements of perceptual information that is analogous to the physical referents and that perceptual information is stored in memory.

Memory and comparison variability had no effect on the number of observations requested. This again is surprising given the robust finding that response times in the method of constant stimuli are very sensitive to the form of the standard. Judgements with remembered standards should lead to longer response times, a finding predicted on the basis of accrual models and typically explained as a result of increased variability in the memory representation. Increased sampling as a result of a threefold increase in the variability of the comparison distribution did not emerge.

In summary, qualified support is provided for the involvement of evidence accrual in expanded comparative judgements. Conspicuous by their absence were the serial position effect for sampling times and the semantic congruity effect. In addition, the distance effect was not as pervasive as expected. The equivocal nature of these findings may have been due to the difficulty of the task. In Experiment 2, the same design was used but this time real money was used for payoffs. The use of real payoffs was expected to give subjects more of an incentive to take the task seriously.

Experiment 2

Method

Five subjects were randomly assigned to each of the four combinations of the standard type by comparison variability conditions. The stimuli and procedure were exactly the same in this experiment as they were in Experiment 1 except that real money was used for the payoffs. Specifically, the base value of a correct response was 50 cents and the base cost of an error was also 50 cents. Each observation cost 1 cent. The total cost for the number of observations was added to the value of the primary decision.

Results

All Os were included in the analyses. For each O, sampling times less than 200 ms or greater than 10000 ms were excluded from calculations involving latencies. Each O received 192 trials. Seventy five percent of the Os (15/20) had one sample time (less than 200 ms) excluded on .4% to 8.2% of the experimental trials. Only 10% of the Os had two sample times excluded on one or two trials. Twenty percent of the Os had more than two sample times excluded on one or two trials. Eighty percent of the Os (16/20) had one sample time excluded on .4% to 6.4 % of the experimental trials for because they were greater than 10000 ms. Post-summative latencies shorter than 200 ms and than 30000 ms were also excluded. Thirty two percent of the Os (8/20) had one post-summative latency excluded on .4% to 5.5% of the experimental trials ($n=192$) because they were greater than 30000 ms. None were shorter than 200 ms. In all, subjects tended to take longer to make decision in Experiment 2 than they did in Experiment 1.

Viability of the expanded-judgement paradigm for binary comparisons.

1. The effect of memory, comparison variability and serial-position on the discriminability percepts. Discriminative sensitivity was examined by performing a three-way ANOVA on the slopes of the psychometric functions. Standard type (MEM vs PER) and the variability of the comparison distribution (low versus high) were entered as two between-subjects factors. Standard was the within-subjects factor. Greenhouse-Geiser adjusted degrees of freedom were used to calculate significance levels for within-subjects effects, but actual degrees of freedom are reported along with the adjusted F values and significance levels.

As expected, placing a standard in memory [$F(1,16)=22.56, p=.0002$] or increasing the variability of the comparison distribution [$F(1,16)=17.44, p=.0007$] massively impaired discriminability. The WFs for perception and memory were .06 and .12, respectively, and are virtually identical to the .06 and .10 obtained in Experiment 1. The WFs for low comparison variability and for high comparison variability were .06 and .11, respectively, and are identical to the WFs obtained in Experiment 1 (.06 and .10). Note, that as was the case in Experiment 1, increasing the variability of the comparison distribution makes perception look like memory. Specifically, the WFs for perception and low comparison variability are identical, and more to the point, so are the WFs for memory and high comparison variability. However, the effect of comparison variability was differentially affected by the nature of the standard (perceptual vs remembered). Although the discriminability of judgements with perceptual standards exhibited large improvements when the variability of the comparison distribution was reduced (WFs were .09 and .04, respectively), the impact of comparison variability was minimal for

remembered standards (.14 and .11). This significant two-way interaction [$F(1,16)=5.50$, $p=.0323$] suggests that the discriminability and precision of standards stored in memory is greatly impoverished compared to their perceptual counterparts and, as a result, decreasing comparison variability does not necessarily result in visible improvement in performance.

An examination of the plots in figure 10 suggested that the discriminability of the standards may interact with the two between-subjects factors, perception/memory and low/high comparison variability. In replication of Experiment 1 and consistent with the literature in memory psychophysics, the discriminability of the four remembered standards under low comparison variability (right panel, filled squares), displayed a u-shaped serial-position effect (SPE). Perception did not exhibit this effect. The lack of an SPE under high comparison variability results from uniformly low discriminability for all of those four standards. Note that all four Weber fractions are as poor as those displayed by the inner remembered standards under low comparison variability suggesting that increasing comparison variability three-fold under this memory condition has pushed discriminative sensitivity to its limit. The relationship between standards and variability differed reliably between perception and memory as confirmed in a significant three-way interaction [$F(3,48)=3.31$, $p=.0362$].

Trend analyses were performed over standards for each of the four standard type by comparison variability conditions. A significant quadratic trend emerged for perceptual comparisons with high variability [$F(1,4)=7.37$, $p=.0533$]. Figure 10 shows that this trend can be accounted for by the sudden drop in Weber fractions from the third standard down to the fourth standard. The cubic trend for remembered judgements with low comparison

Figure 10. Discriminability of perceptual (left panel) and remembered (right panel) standards under low comparison variability (filled symbols) and under high comparison variability (unfilled symbols).

Figure 11. Standardised differences between PSEs and physical standards for each subject under perceptual and remembered judgements.

variability was only marginally significant [$F(1,4)=4.68$, $p=.09$] because of large standard errors for the inner standards. The shortest standard had a WF of .09 under low comparison variability and .14 under high comparison variability indicating that increasing the variability of the comparison distribution three-fold increased the minimal difference required for a just noticeable difference by 5 percentage points. The longest standard had a WF of .08 under low comparison variability and .16 under high comparison variability. In the latter case, a three-fold increase in variability led to a two-fold increase in the required minimal difference i.e. an added 8 percent (see Appendix A for individual differences).

In Baranski and Petrusic (1992), Weber fractions determined for two experiments with set size three resulted in values ranging from .03 to .05 for perception, and from .04 to .08 for memory. Set size has been shown to have a significant effect on discriminability for remembered standards but not for perceptual standards (Leach, 1992). For set size 6, Leach (1992) found that the WFs ranged from .04 to .23. The perceptual WFs obtained in this experiment are comparable to those obtained by Baranski and Petrusic (1992). The marginal WF for perception was .06 and the cell WF for perception under low comparison variability was .04. The WFs obtained for memory (.12), for memory under low comparison variability (.11), and for memory under high comparison variability (.14) are between the setsize 3 values obtained by Baranski and Petrusic (1992) and the set size 6 values obtained by Leach (1992).

The precision location of the underlying subjective magnitudes were examined using PSEs as was done in Experiment 1. This time PSEs for memory were significantly more variable than their perceptual counterparts [$F(1,16)=5.44$, $p=.03$] suggesting a more

variable underlying remembered representation as has been demonstrated in research using the method of constant stimuli (Baranski & Petrusic, 1992). Figure 11 illustrates that PSEs for each subject separately for each standard. PSEs for perceptual standards tend to cluster more closely around the corresponding physical value while PSEs for memory are generally more deviant. One subject's PSE could not be computed because of a discriminability slope of 0. Note in Table 3 below that the average PSE for perception with low comparison variability is nearly identical to their respective physical standards, but PSEs for memory deviate more. With the exception of standard 120 for perception, the same pattern emerged under high comparison variability; however, increasing comparison variability increases the PSE deviations. The effect of comparison variability was not significant. These results also confirmed Baranski and Petrusic's finding (1992) that the variability of PSEs did not follow a particular pattern, but rather represented idiosyncratic individual differences. Some have suggested that stimuli in memory tend to drift towards the geometric mean of all stimuli (Helson, 1947; Hollingworth, 1910). The geometric mean in this case would be 132.8 pixels. Although standards 1 and 3 under low comparison variability display the appropriate drift towards the geometric mean, standards 2 and 4 go in the opposite direction. Under high comparison variability standards 2 and 3 exhibit the correct drift but standards 1 and 4 do not.

Table 3

Mean PSEs and Absolute Deviations (in parentheses) of PSEs from Their Respective Physical Standards for Each Standard Under Each Comparison Variability by Standard Type Combination

Variability	Standard Type	Standard (pixels)			
		60	120	180	240
Low	Perception	60.2(0.2)	118.7(0.3)	180.6(0.2)	240.0(0.1)
	Memory	61.3(0.9)	185.8(14.7)	173.2(2.3)	248.0(1.0)
High	Perception	63.9(1.3)	128.5(1.5)	186.1(0.7)	238.0(0.4)
	Memory	50.6(3.4)	121.9(5.8)	168.6(2.9)	272.0(2.8)

2. *The effect of memory, variability and serial-position on the time to make expanded comparative judgements.* Latencies for the first sample (RT1), the last sample (lastRT) and samples on average (mnrt), as well as post-summative latencies (psrt), number of observations (obs) and accuracy (correct) were analysed separately in 4-way ANOVAs with standard type and comparison variability entered as between-subjects factors. Distance and standard were entered as within-subjects factors. Greenhouse-Geiser adjusted degrees of freedom were used to compute significance levels, however, actual degrees of freedom are reported (see Table 4 for effects). Effects are discussed in separate sections.

The number of observations required for remembered judgements tended to be slightly less than that required with perceptual standards (6.2 versus 8.7). Nearly twice as many observations were required when comparison information was more variable (9.4 versus 5.4), although again this difference was not statistically significant. The difference between memory and perception was slightly larger with low comparison variability (5 samples) than with high comparison variability (3 samples), although not significantly so. These findings confirm exactly the predictions made by Thurstonian model 1 (see Table 1) and show the effect of variability on the propensity to continue accumulating observations. As predicted by this model, placing a standard in memory overwhelms the decisional system such that the cost of requesting more information is too high for the minimal improvement in accuracy that one would expect. However, when comparison variability is manipulated, increased sampling does result in a meaningful improvement in performance, and as a result, O's are more willing to continue sampling.

These findings are counter-intuitive when interpreted in the context of increased decision times for memory compared to perception obtained in traditional psychophysical experiments (Petrusic & Baranski, 1994). Accrual models suggest that increased sampling should occur due to the increased variability of the percept held in memory (Petrusic et al., 1995). One possible explanation is that subjects are making use of a strategy to set a ceiling on the costs associated with any decision. If the difficulty of a judgement is such that errors are highly likely, then less observations will be sampled because costs associated with sampling will increase the overall loss that results from making an error. This result is exactly what Thurstonian model 1 predicts. Although the cost of observations seems to be relatively low (1 point), when combined with the overall low discriminability, expected value predicts that improvements as a result of increased sample size will be quickly overwhelmed by the cost associated with accumulating information.

This dissociation between the memory/perception manipulation and the comparison variability manipulation is intriguing. Increased sampling in the high variance condition replicates findings by Irwin, Smith and Mayfield (1965) and Irwin and Smith (1967) that increasing distributional variability leads to increased sampling. These results are also consistent with increased response times for judgements with remembered standards in traditional psychophysical experiments where it has been assumed that placing a stimulus in memory increases the variability of the internal representation relative to the internal representation of its perceptual counterpart.

Anecdotally, about 75% of the subjects reported an attempt to maintain a ceiling on the number of observations in order to keep their information costs down, despite

knowing that this strategy might compromise decision accuracy and result in increased losses. This idea is reminiscent of fixed sample models of comparative judgements (for a review see Vickers, 1980) where subjects pre-select a number of observations to maintain a level of accuracy. This number of samples is adjusted at the beginning of each block depending on the difficulty of the discrimination. As reviewed in the literature, however, fixed-sample models cannot account for variations in decision time when difficulty is unpredictably varied within blocks. Subjects in Experiment 2 were not using a fixed sample model since, as will be discussed below, difficulty was varied unpredictably from trial-to-trial within blocks.

Perhaps the observed differences in the length of observations sequences would turn out to be significant if the memory/perception and variability manipulations were within subjects variables. Subjects' comments regarding their reluctance to sample more than absolutely necessary is indicative of an overall strategy to minimize losses that is not directly tied to the regulation of the actual comparison process. This conservatism may act to mitigate some of the effects by restricting the range of values for the dependent measure.

Another possibility is that having both the standard in memory and the comparison stimulus available only as a distribution of observations, overwhelms the decisional system such that further sampling is counter-productive. Recall, however, that the discriminability of perceptual standards when comparison variability was high was as poor as that of remembered standards with low variability. Since increasing variability leads to comparable impairments in discriminability then the high variability condition should also have led to decreased sampling compared to low comparison variability. In

this case, however, Weber fractions may in fact be experiencing a floor effect such that their ability to differentiate between the impact of comparison variability and that of long-term memory on the comparison process is weakened. Experiment 3 will pursue this issue by using successive presentations of the standard and the comparison distribution according to the method of constant stimuli. This design allows the opportunity to explore the differential impact of the variability of the standard and the variability of the comparison distribution on the sampling process.

The number of observations requested showed absolutely no evidence of a serial-position effect even when the variability of the comparison distribution and the difficulty of the comparison (distance effect) were taken into account. At first, this finding might suggest that discriminability, which has been consistently shown to be impaired for remembered stimuli, has no impact on a comparative decisions that involve some form of aggregation of information. This interpretation should be made with caution since other aspects of this experiment also manipulated difficulty and resulted in reliable changes in observation sequence length. As will be shown below, sampling time and post-summative latencies associated with post-sampling decision time clearly show the serial-position effect suggesting that in fact memory-based differences in discriminability do in fact impact on the decisional process.

3. Mean sampling time and the effects of memory and of comparison variability.

Memory judgements did not reliably differ from perceptual, nor low comparison variability judgements from high, with respect to the mean time taken to request the next observation. Observations for memory judgements were sampled on average 282 ms

slower than their perceptual counterparts. Recall however that memory judgements required less observations than perception. Also of importance is that observations sampled from the low variability distribution were requested 195 ms more slowly on average than were those from the high variability distribution. Recall that the high variability condition required more observations than the low variability condition. The inverse relation between the sequence length and sampling speed suggests a trade-off between information acquisition and the speed with which information is sampled. A series of Pearson product-moment correlations were performed between the number of samples and each of the latency measures, i.e. time to request the first sample, time to request the last sample, mean sampling times, and post-summative latencies (see Table 4). The most dramatic effect shows a positive relationship between the number of observations taken and post-summative latencies for each of the 20 subjects. This result is consistent with the view that the post-sampling latency serves as the locus of the decisional process. Subjects may be recalling the observations in an evidence-accrual comparative judgement process. The more explicit observations sampled then the longer the decisional process because more evidence has to be recalled and processed decisionally. Secondly, a negative relationship between the number of observations and mean sampling time was evident for 17 of the 20 subjects (but 5 were not significant). On average, faster sampling times tended to be accompanied by increases in the number of observations requested. No clear patterns emerged for the first or last samplings.

The increased speed required under increased comparison variability may be due to limits in the persistence of perceptual information after the initial stimulus has been removed (Fechner, 1882/1987) and the role of subsequent stimuli in interfering with this

information (Massaro, 1970a, 1970b). As the variability of the comparison stimuli is increased, subjects sample faster in order to increase the clarity of their estimate of the true comparison mean and to compensate for retroactive interference of previous stimuli (Massaro, 1970a, 1970b). The possibility of retroactive interference has never been examined in an expanded judgement paradigm. This potential relationship between observation length and sampling time is complicated by differences in discriminability between conditions. A comparison of two conditions with equivalent Weber fractions affords useful insight into this issue. The high-variability perceptual condition exhibited a Weber fraction that was very similar to that obtained for the low-variability memory condition (.09 versus .11). Sampling was faster in the former compared to the latter by 477 ms. But the perception paid for this speed by requiring 6.5 more samples than the memory condition. Past research has shown that the comparison process is slower when one of the stimuli is in memory (Banks, 1977; Baranski & Petrusic, 1992). This form of speed-sampling trade-off complicates any comparisons made between the memory and the comparison variability manipulations.

4. *First and last sampling times and the effects of memory and of comparison variability.* As was the case for mean sampling times, both the time to take the second (first sampling time) and last samples were longer for comparisons with remembered standards than with perceptual standards. In addition, the first and last samplings took longer for judgements with more variable comparison stimuli than for judgements with less variable comparison stimuli. These effects were not reliable, however, but show that the differences observed for average sampling times are present during the entire sampling process.

Table 5

Correlations Between the Number of Observations and the First Sample Time (First), the Last Sample Time (Last), the Mean Sample Time (Mean) and Post-summative Latencies (Post)

MEM/PER	SD	SUBJ	# OBS	FIRST		LAST		MEAN		POST	
				r	p	r	p	r	p	r	p
PER	LOW	3	4.1	-.10	ns	.18	.011	.13	ns	.77	.001
PER	LOW	6	2.9	-.07	ns	.15	.060	-.07	ns	.65	.001
PER	LOW	13	4.0	.18	ns	-.05	ns	.03	ns	.95	.001
PER	LOW	15	4.6	-.06	ns	.16	.027	.04	ns	.92	.001
PER	LOW	18	15.5	.01	ns	-.02	ns	-.27	.001	.78	.001
PER	HIGH	1	8.6	-.16	.031	-.13	.064	-.41	.001	.78	.001
PER	HIGH	5	9.4	-.28	.001	-.34	.001	-.46	.001	.20	.005
PER	HIGH	11	11.9	-.06	ns	.06	ns	-.19	.008	.97	.001
PER	HIGH	12	5.4	-.13	.078	.02	ns	-.33	.001	.81	.001
PER	HIGH	21	20.6	-.02	ns	-.16	.024	-.35	.001	.72	.001
MEM	LOW	8	3.3	.06	ns	-.10	ns	-.29	.016	.40	.001
MEM	LOW	10	6.0	-.04	ns	-.03	ns	-.19	.010	.55	.001
MEM	LOW	14	3.7	-.19	.011	-.15	.047	-.28	.001	.34	.001
MEM	LOW	17	5.7	-.18	.016	.33	.001	-.07	ns	.84	.001
MEM	LOW	20	4.3	.10	ns	.17	.027	-.10	ns	.67	.001
MEM	HIGH	2	3.9	.03	ns	.04	ns	-.10	ns	.40	.001
MEM	HIGH	4	4.9	.08	ns	-.09	ns	-.06	ns	.50	.001
MEM	HIGH	9	21.2	.06	ns	-.33	.001	-.35	.001	.58	.001
MEM	HIGH	16	4.5	-.01	ns	-.15	.045	-.26	.001	.61	.001
MEM	HIGH	19	4.0	-.14	.056	-.02	ns	-.29	.001	.45	.001

The serial-position effect over remembered standards was present throughout the sampling process as shown by reliable standard by standard-type interactions for the first [$F(3,48)=3.22$, $p=.0518$] and for the last sampling times [$F(3,48)=4.41$, $p=.0149$] (see figure 12, top two panels). The effect is much stronger by the end of the sampling process. Note the significant main effect of standard for the last sampling time [$F(3,48)=3.83$, $p=.0251$] suggesting that the SPE in memory was strong enough to overwhelm any mitigating effects of the flatness of the perceptual curve. The main effect of standard was not significant for the first sampling time. The differential impact of the SPE at the end of the sampling process compared to the beginning can be accounted for by a more general reduction in the variability of sampling times by the end of the sampling process (see error bars in top left panel compared to the top right panel of figure 12).

5. Mean sampling time and the serial-position effect. As shown in figure 12 (bottom left panel), the average time to request the next observation for comparisons with remembered standards was slightly longer overall than for their perceptual counterparts, although not significantly so. Mean sampling times in the memory condition were slowed-down for the two middle standards compared to the outer standards, but were constant across perceptual standards. The serial-position effect evident in the reliable interaction between standard and standard type [$F(3,48)=5.10$, $p=.013$] clearly replicates the differences between memory and perception so consistently demonstrated for decision times (Kennedy, 1990). A main effect of standard [$F(3,48)=3.52$, $p=.0433$]

showed that the serial-position effect exhibited by the remembered standards was strong enough to overwhelm the flatness of sampling times across percepts.

6. Post-sampling time and the effects of memory and of comparison variability. As was the case for sampling times, post-summative latencies (psrt) took longer when the variability of the comparison distribution was increased. Interestingly, psrts show effects opposite those exhibited during the sampling process. Specifically, times are shorter for comparisons with remembered standards than with perceptual standards, but they are longer for judgements with highly variable comparison stimuli relative to judgements with less variable comparison stimuli. This would make sense if the comparison process takes place after observations have been accumulated. Shorter psrt's for memory are consistent with the fewer observations requested compared to perception. Conversely, longer psrt's for the high variance comparison condition are consistent with the greater number of observations requested in that condition.

The serial-position effect occurs in memory but not in perception although the effect just misses significance [$F(1,48)=2.57$, $p=.0785$] because of large overlapping error variability between judgements with perceptual and remembered standards. This SPE is not typical however because latencies for perceptual standards should be equivalent to or faster than the latencies for the remembered outer standards. In this case, figure 12 clearly shows that latencies for perception are as bad as those of the inner remembered standards.

Distance effects in accuracy, sample size and sampling times.

Memory/perception and variance were entered as between subjects factors. Distance and standard were entered as within-subjects factors in this 4-way ANOVA. The distance

Figure 12. Serial position effects for the first sampling time (top left panel), the last sampling time (top right panel), mean sampling times (bottom left panel) and post summative latencies (bottom right panel).

effect should be enhanced in memory compared to perception (Moyer and Bayer, 1976). Increased comparison variance is also expected to amplify the distance effect. Dependent variables for each univariate ANOVA will be the arcsine transformed percent correct, n (sample size) and RT (sampling time).

1. Distance effects on accuracy. Distance significantly affected accuracy [$F(1, 16)=38.80, p<.0001$] with the longer distances exhibiting improved accuracy (.88) compared to the shorter (.77). The magnitude of the distance effect depended on the nature of the standard however [$F(1, 16)=5.26, p=.0357$]. The top panel of figure 13 shows that the distance effect was enhanced in perception compared to memory (the overall differences in accuracy between the short and long distances were .11 and .07 respectively). The bottom panel shows that the magnitude of the distance effect showed a tendency to be depend on the variability of the comparison distribution although this interaction was not quite reliable [$F(1, 16)=3.87, p=.0667$]. Specifically, the distance effect was attenuated by increasing comparison variability so that the overall differences in accuracy were .11 for the short distances and .08 for the long. Interestingly, the distance effect was of exactly the same magnitude in the memory condition as it was in the high variability condition. Likewise, both perception and low comparison variability resulted in distance effects of the same magnitude. Comparability in the magnitude of the distance effects in the memory manipulation in and the variability manipulation suggests that placing a standard in memory acts by increasing the variability of the internal representation relative to the actual percept.

MEMORY vs PERCEPTION

LOW vs HIGH COMPARISON VARIANCE

Figure 13. Distance effects on accuracy for memory and perception (top panel) and for low and high comparison variability (bottom panel).

Distance effects on the number of observations. Distance significantly affected the number of observations [$F(1, 16)=7.36, p=.0154$] with the longer distances requiring fewer observations (7) than the shorter ones (7.7). The magnitude of the distance effect depended on the nature of the standard however [$F(1, 16)=5.39, p=.0338$]. As illustrated in figure 14, the effect was larger for perceptual judgements (9.3 vs 8.1) and nearly non-existent in memory (6.2 vs 6.1).

Again we find peculiarities of the SCON expanded-judgement paradigm. Traditional psychophysical comparative judgements behave differently in that the distance effect is enhanced in memory relative to perception (Moyer & Bayer, 1976). The distance effect seems equivalent in memory and in perception as illustrated in Figure 14. One possibility is that the variability of the internal representation is so increased that measurement of the stimuli becomes very difficult.

Distance effects on sampling times and on post-summative latencies. As can be seen in figure 15 and as confirmed in a series of ANOVAs, the distance effect was completely absent from the beginning of the sampling process and from sampling times on average (top left and bottom left panels). The last sampling time did approach significance [$F(1,16)=3.87, p=.0666$] suggesting an emergence of the comparison process as sampling progresses (top right panel). This suggests that although the comparison process is not present overall during sampling, it does emerge over time. Post-summative latencies showed a significant distance effect [$F(1,16)=17.47, p=.0007$] confirming the finding in Experiment 1 that subjects recollect and accumulate observation in an evidence

accrual-like fashion after explicit sampling is terminated (bottom right panel). On average, psrt's were 804 ms longer for the closer comparison pairs.

Semantic Congruity for sample size, sampling time and accuracy.

As was the case for Experiment 1, semantic congruity failed to emerge for any of the dependent measures. This lack of an effect will be discussed in the context of peculiarities of the expanded judgement paradigm.

Figure 14. Distance effects on the number of observations for memory and perception.

Figure 15. Distance effects on overall sampling times and post-summativ latencies.

Discussion

Experiment 2 provided a better test of the viability of the expanded judgement paradigm than did Experiment 1 by providing subjects with real monetary incentives for optimal performance. Characteristic differences between perception and memory in terms of the discriminability and the variability of the underlying memory representation confirmed the results of traditional psychophysical experiments. In replication of Experiment 1, judgements with remembered stimuli exhibited poorer discriminability (Weber fractions) than judgements with perceptual standards. Also in replication of Experiment 1, high comparison variability resulted in poorer discriminability of the standards compared to low comparison variability. This latter difference in discriminability was attenuated for judgements with remembered standards suggesting that the efficiency of decisional processes is impeded by the compound effect of the noisy representation of a stimulus in memory and the noisiness of comparison variability. As was the case in Experiment 1, the serial-position effect occurred for memory judgements with low comparison variability. Increasing the variability of comparison observations introduced so much noise into the decisional system that discriminability was as bad for all standards as it was for the inner standards in the low comparison variability condition. Sampling times and post-summative latencies exhibited serial-position effects for remembered standards but not for perception. These findings suggest that a subject-controlled expanded judgement paradigm provides a viable means of psychophysical investigation. As shown by PSEs in Experiment 2, the representations of subjective magnitudes held in memory are more variable than their perceptual counterparts. The

deviations of PSEs from their corresponding physical values did not follow any particular pattern, but rather, represented individual differences (Baranski & Petrusic, 1992).

As was the case for Experiment 1, memory and comparison variability had no significant impact on the number of observations. The lack of significance could have been due to small sample sizes and small variability in the dependent measure due to a reluctance on the part of observers to vary sample size. On average subjects sampled 7 observations in Experiment 2 and 10 in Experiment 1 suggesting that imposing monetary costs may have led to a reluctance to spend money on information when the task is considered difficult. Unfortunately, Vickers et al. (1985) does not report the overall mean number of samples. They do present a graph (see their figure 1) suggesting that all subjects in the low discriminability condition sampled at least 19 observations while subjects in the high discriminability condition sampled 17. In the current Experiment 2, mean samples per subject varied between 4 and 23, but the variability within each subject was quite small.

This two-alternative forced choice expanded judgement task was more difficult than Vickers et al.'s (1985) choice reaction time task. In the expended judgement card task, Irwin and Smith (1956) found that subjects requested between 20 and 45 observations when no payoffs were used. Irwin and Smith (1957) showed that manipulating the amount of real money associated with correct and incorrect responses had a significant impact on the number of observations requested. In their experiment, as in the current Experiment 2, each observation requested cost money. The higher the cost, the more reluctant observers were to sample. Although using payoffs may have helped observers to

take their judgements more seriously, the resulting reluctance to sample may be responsible for the lack of effects for sample size in the current experiment.

Despite the lack of significance for sample size, a descriptive examination of the number of observations for memory/perceptual judgements in conjunction with mean sampling times and post-summative latencies revealed theoretically important results. Fewer observations were sampled for remembered than for perceptual judgements. This is counter to evidence accrual theories and empirical psychophysical data. However, sampling times were negatively correlated with the number of observations. The same relationship between sample size and sampling times was also observed between high and low comparison variability manipulations. Although more samples were requested when comparison variability was high, observations were also sampled more quickly. This trade-off between number of observations and sampling times suggests that care must be taken when interpreting changes in sample size in an expanded judgement task. These results suggest that the number of observations required for a decision is adjusted in tandem with the speed of comparison stimulus inspection in order to achieve an optimal level of stimulus information storage. Faster sampling leads to more observations, while longer inspections times result in reaching optimal decisional capacity with fewer observations. These results suggest two main ideas. First, explicit observations do not behave like implicit observations. But secondly, some unknown factor causes the decisional system to slow down sampling when observations are compared to a remembered standard, and when comparison observations are less variable for either perceptual or remembered standards. Increased sampling for more highly variable comparison observations, but reduced sampling for memory judgements, appear

contradictory if one accepts the view that a memory representation leads to slower response times because of increased variability in the stimulus representation. Accrual models have suggested that differences in response times are due to differences in the number of implicit accruals. These expanded judgement data suggest that both the number of accruals and the time taken to sample and judge each accrual is also important. On the other hand, there may be fundamental differences between perceptual and remembered stimuli, other than representational variability, that account for differences in decision latencies.

Vickers et al.'s (1985) finding that post-summative latencies reflect decisional processing has been replicated. Post-summative latencies were highly positively related to the number of observations suggesting that after sampling has ended, subjects recollect the observations and process them in an evidence-accrual like fashion. Importantly, a large distance effect was present for post-summative latencies suggesting that indeed post-summative processing of stimuli is decisional in nature. However, the presence of a distance effect towards the end of the sampling process suggests that decisional processing of an evidence-accrual like nature is involved in the decision to stop sampling. The late emergence of the distance effect in the sampling process may be an artifact of the small number of observations requested by observers. In a difficult expanded judgement task, the decisional system may need to tune itself up on each trial before it can begin to make comparisons.

The delay in the comparison process during sampling may also be due to the instructions that require the observer to calculate a mean and then compare it to the standard. Note that the differences in sample size between memory/perception and

low/high comparison variability were as predicted on the basis of the Thurstonian model suggesting that observers make a decision by forming some sort of aggregate difference judged in light of distributional variability. That explicit observations might be analogous to implicit accruals is dubious, since a clear distance effect for observations under perceptual judgements was not paralleled by a distance effect in memory. In fact, the distance effect was restricted to perceptual judgements under low comparison variability. Non-expanded judgement pair comparisons (Moyer, & Bayer, 1976; Petrusic et al., 1994) show amplified distance effects in memory compared to perception. Explicit observations in a subject-controlled indeed do not behave like implicit accruals. As a caveat, the nature of the comparison processes used in this experiment may have been susceptible to a conflict between the natural tendency to process information in an accrual-like fashion and the instructions to use the average of the observations to make the comparative decision. Another experiment would have to explicitly instruct observers to compare each observation to the standard to arrive at a decision. Under such conditions, evidence-accrual like behaviour might emerge for explicit observations.

Experiment 3

Introduction

Poorer judgements with remembered stimuli, compared to perceptually available stimuli, are assumed to result from increased variability in the remembered representation. A long research tradition shows that the representations of various types of stimuli in short-term memory degrade over time and that this degradation is affected by a multiplicity of factors such as the number of items to be remembered (Miller, 1956), the depth with which stimuli were originally processed (Craik & Lockhart, 1972), and the quality of verbal encoding and rehearsal (Peterson & Peterson, 1959). The E-J paradigm provides a potentially fruitful opportunity to examine the effects of variability on performance directly through the manipulation of the variability of the comparison distribution and indirectly through the degradation of the standard over time.

Diffusion Theory Versus Interference Theory

Judgements regarding successively presented stimuli necessarily require that the first stimulus be held in short-term memory while the second stimulus is perceived. The efficiency with which the decisional system compares these stimuli depends on the extent to which the first stimulus is forgotten. Several studies have provided evidence that forgetting can be described as a time-bound increase in the variability of the representation for the first stimulus through the process of diffusion (Kinchla & Allan, 1969; Kinchla & Smyzer, 1967; Palmer, 1990). Kinchla & Smyzer (1967) showed that

the variability of the internal representation increases as a linear function of time elapsed since acquisition of the study stimulus.

Alternatively, interference theory (Massaro, 1970 a, b, c), focuses on the role of attentional capacity on memory. Posner (1969) and Posner and Konick (1966) found that the maintenance of visual codes in memory during a task with successive stimuli requires attentional resources. Since refreshing or rehearsal is a form of division of attention that increases demands on a finite amount of attentional capacity, the rate at which a study stimulus is forgotten should increase as the number of intervening stimuli (comparison stimuli) increases. Massaro (1970 a, b, c) showed that both time since acquisition and the number of intervening items after acquisition are significantly related to forgetting. Using d' as a measure of discriminability, diffusion theory predicts an inverse relationship with the variance of the study stimulus over time. Interference theory predicts an exponential decrease in d' with the number of stimuli presented.

Palmer (1990) presented subjects with one study line length that was accompanied by either two or four distracter line lengths. After a variable ISI, a line length was presented for comparative judgement. Palmer's failure to obtain an interaction between diffusion rate and the number of study items may have resulted from his use of only one standard to which subjects received extensive prior exposure. The standard may have been stored in long-term memory thus completely circumventing the effect of forgetting. Using naive subjects and a wider range of study line lengths and of ISIs, Leth-Steensen (1994) found an interaction between diffusion rate and the number of study items. The slope of the linear relation between variability and time increased as the number of study stimuli increased suggesting that the locus of forgetting involves attentional capacity

limitations. This result was equivocal however since the effect was not consistently demonstrated in all subjects.

Diffusion-based Expanded Judgement Model

In the diffusion model, as in the SDT model (Green & Swets, 1966), the internal representations of the study stimulus at acquisition time and of the comparison stimulus at time t are assumed to be built up from normally distributed fluctuations in inputs from their actual respective physical stimulus values. The diffusion model⁸ presents itself as a more general model than SDT since it describes a situation where the input sensory process of the study stimulus undergoes an increase in variability over time. As a result, diffusion theory can be applied to an expanded judgement paradigm where the underlying representations of both the estimates of the standard and of the comparison mean undergo changes in variability as a function of sample size.

The implications of diffusion theory (Kinchla & Smyzer, 1967) for evidence accrual models are clear (Audley & Pike, 1965; Petrusic, 1992; Vickers, 1970). Diffusion increases the variability of the line stimulus held in memory to the extent that the time separating the acquisition of the first stimulus from the sensory input of the second forces the storage of one stimulus in short-term memory while the other stimulus is perceptually analysed. This occurs for successively presented stimuli, and for simultaneously presented stimuli to the extent that scanning between lines requires that one line be held in short-term memory while the other is scanned. Under memory conditions, an evidence accrual model would predict an increase in the number of accruals due to the variability-based increase in the difficulty of the judgement (Vickers, 1980). However, consideration

to diffusion over time. Finally, increasing the variability of the comparison distribution of explicit observations should exacerbate the above effects.

Experiment 3 is an experimenter-controlled expanded judgement task that leads to a more general diffusion model which simultaneously assesses the trade-off between on the one hand, the increased discriminability achieved by increasing the sample size, and on the other, decreased discriminability due to a time-bound diffusion of the representation of the study stimulus.

The Effect of Diffusion and of Comparison Variability on Discriminability

Since a particular standard-comparison pair is only presented in one order, then SDT cannot apply. In the case of a 2AFC task, the sensitivity index of SDT, d' , is the standardised distance between two distributions of sensory differences, one corresponding to the sensory event shorter-longer, the other to the event longer-shorter (Macmillan & Creelman, 1991) where the standard and comparison stimuli are presented in both orders. In the current paradigm, the decisional space is based on only one presentation order for each stimulus pair thus precluding the measurement of both d' and β , the bias criterion.

The E-J ECON paradigm must rely on a Thurstonian representation which permits the use of the Gaussian distribution to make explicit and testable predictions concerning the effect of comparison variability and sample size on discriminability. The O is assumed to compute the difference between a single observation sampled from the representation of the study stimulus and the average value of the comparison sequence drawn from a sampling distribution with a parent population $N(\mu, \sigma^2)$. According to the

algebra of expectations (Hays, 1994, p.916) and assuming independence,

$\sigma_{(x-r)}^2 = \sigma_x^2 + \sigma_r^2$. In this case, the variance of the differences between the study and comparison stimuli can be expressed as the sum of two sampling distributions, where $n=1$ for the study stimulus and $n \geq 1$ for the comparison stimulus.

In the case of a Thurstonian distribution of differences, the criterion represents the point at which the subject is unable to discriminate between the magnitudes of the two stimuli. Differences to the right of the criterion are positive and reflect the input sensory event, first stimulus is longer. Differences to the left of the criterion are negative and reflect the input sensory event, first stimulus is shorter. The distance of the criterion from the expectation of the discriminial dispersion tells us the extent to which the stimuli are discriminable. When the $E(x)=0$, the expectation equals the criterion and the first stimulus is judged longer 50% of the time. As the criterion moves left, an increasing proportion of stimulus pairs are judged as positive thus reflecting increased discriminability. According to this view, the discriminability index is proportional to the difference between the criterion and the expectation of the underlying distribution of sensory differences after n observations, standardised by the variance of the differences.

The discriminability index in the unbiased case, D_n can be defined as

$$D_n = \frac{0 - E(m_{t,n} - \bar{x}_n)}{\sqrt{\sigma_{m_{t,n}}^2 + \sigma_{\bar{x}_n}^2}} \quad (6)$$

where $m_{t,n}$ is the subjective value of the study stimulus at time t after n observations, $\bar{x}_n$

is the subjective estimate of the comparison mean μ after n observations, $\sigma_{\bar{x}_n}^2 = \frac{\sigma_\mu^2}{n}$

where σ_μ^2 is the variance of the comparison distribution, and finally, $\sigma_{m_{t,n}}^2$ is the variance

of the distribution associated with the standard at time t after n observations. Since after acquisition the standard is subject to diffusion over time, then $m_{t|n} = m_0 + d_{t|n}$ where m_0 is the subjective value of the standard at the end of acquisition and $d_{t|n}$ represents the effect of diffusion on m_0 following a random walk process that stops at time t after n observations. As a result, $E(m_{t|n} - \bar{x}_n) = E(m_0 - d_{t|n} - \bar{x}_n)$, and after distributing we obtain $E(m_0) - E(d_{t|n}) - E(\bar{x}_n)$. Since diffusion is said to act on the variance and not on the mean, then $E(d_{t|n}) = 0$ and $E(m_{t|n} - \bar{x}_n) = E(m_0) - E(\bar{x}_n)$. Since $E(m_0) = s_0$, the physical value of the standard, and $E(\bar{x}_n) = \mu$, the physical value of the mean of the comparison distribution, then the numerator of equation (1) can be expressed as $0 - E(s_0 - \mu) = -y$ (Kinchla & Smyzer, 1967).

Kinchla & Smyzer's (1967) assumption that $E(m_0) = s_0$ and $E(\bar{x}_n) = \mu$ seems unrealistic however, since it has been shown that the subjective value of a stimulus is affected by the stimulus context in which it is evaluated. The E-J paradigm complicates contextual effects since a trial is comprised of a standard and comparison lines whose lengths vary randomly from trial to trial. Even Thurstonian scale values obtained using a traditional psychophysical 2AFC task could not be accurate when applied to E-J data. As a result, the current experiment used the average absolute difference, y , between the standard and the comparison mean as the input difficulty level despite the realisation that the subjective value of y differed depending on where it is located on the underlying physical continuum.

Equation (6) asserts that the variance of the acquired stimulus undergoes diffusion over n observations such that $\sigma_{m_{t|n}}^2 = \sigma_0^2 + \sigma_{d_{t|n}}^2 = \sigma_0^2 + \phi_{t|n}$, where ϕ is the diffusion

rate of the study stimulus that takes place over time, $t|n$. The notation $t|n$ refers to a time interval specific to the n^{th} observation. Equation (6) can be rewritten as

$$D_n = \frac{-y}{\sqrt{(\sigma_0^2 + \phi_{t|n}) + \frac{\sigma_\mu^2}{n}}} \quad (7).$$

D_n can be converted to a probability correct by using tails' normal deviates i.e. the area to the right of the criterion. The standard error of the differences reflects the two contrasting effects of n , one which increases the precision of the estimate of μ , the mean of the comparison distribution, and the other which increases the magnitude of $\sigma_{m,n}^2$ over time through a diffusion process. Manipulating equation (7) provides a three parameter equation of the form $a + bx + \frac{c}{x}$ for estimating the variance of the differences. After squaring and rearranging,

$$\sigma_0^2 + \phi_{t|n} + \frac{1}{n}\sigma_\mu^2 = \left(\frac{-y}{D_n}\right)^2 \quad (8).$$

The amount of time, t , elapsed since acquisition is monotonically related to n . Since in this experimenter-controlled model n observations are presented at a constant rate, then t can be set to some constant magnitude. The term t can therefore be re-expressed as a single observation, specifically one $n = \delta t$.

Variance estimate for the memory condition

Equation (8) then becomes

$$\sigma_0^2 + \phi(n \times \delta t) + \frac{1}{n} \sigma_\mu^2 = \left(\frac{-y}{D_n} \right)^2 \quad (9)$$

where $\sigma_{y_n}^2$ is the total variance of the differences. Plotting these variance estimates as a function of n should provide a curvilinear relationship with linear and hyperbolic components. The optimal number of observations used to stop the sampling process should be that point at which σ_y^2 , the variance of the differences, levels off before it begins to increase in magnitude. In figure 16, panels A & B provide idealised curves showing how optimal n (arrows) is reached as the total variance changes across n . Panels C to E show that diffusion (ϕ) causes an increase in the variability of the standard (filled diamonds) that counteracts the beneficial effect of increasing sample size on the precision with which the comparison mean can be estimated (unfilled triangles). This trend is evident by comparing perception ($\phi = 0$, panel C) with memory ($\phi > 0$, panels D and E).

Figure 16. Effect of sample size (n) on total variances of the differences (top panels) and on separate variances (bottom panels) for perception ($\theta=0$) and for different levels of diffusion ($\theta=10, 20$) under low ($\sigma_{jk}^2 = 100$) and high ($\sigma_{jk}^2 = 900$) comparison variances.

Figure 17. Effect of sample size (n) on discriminability for perception ($\theta=0$) and for different levels of diffusion ($\theta=10,20$) under low ($\sigma_{jk}^2 = 100$) and high ($\sigma_{jk}^2 = 900$) comparison variances.

Experimentally increasing the variability of the comparison distribution should decrease discriminability (figure 17, panels D versus E) by increasing the total variability of the difference distribution (figure 16, panels A versus B). Increased sampling should also result since each sample leads to relatively larger gains in variance reduction than when the base comparison variance is smaller. Figure 16 provides a conceptualisation of the effect of diffusion (filled circles versus unfilled circles and squares in panels A and B), comparison variability (panel A vs B) and n (arrows in panels A and B) on total variances. Figure 17 provides a conceptualisation of the effect of diffusion (panels A through C), comparison variability (up triangles versus down triangles in panels A through C) and n (arrows in panels A through C) on the Thurstonian discriminability index. Clearly, diffusion reduces overall discriminability but more importantly reduces optimal sample size relative to the perceptual condition where no diffusion occurs. Increasing comparison variability acts in concert with diffusion to lower discriminability even further, but increases the relative gains in discriminability that result from increased sample sizes.

Subtracting equation (5) from equation (4) for each n separately provides $\phi(n \times \delta)$. This value can be estimated by subtracting variance estimates for observed perceptual judgements, $\left(\frac{-y}{\hat{D}_n} \right)^2$, from variance estimates for observed memory judgements.

Regressing these values on n should provide an estimate of the diffusion rate. These values can then be entered into equation (4) as estimates of σ_0^2 and ϕ , and added to the

objective values σ_{μ}^2 to obtain an E-J model prediction for the total variance for sample sizes.

Diffusion estimate using pair comparisons

A proper estimate of the diffusion rate can also be obtained by conducting a traditional pair comparison experiment with the same standards and comparison stimuli used in ECON expanded-judgement task. The pair comparison task was administered after all E-J sessions were completed in order to avoid giving the Os added experience with the line length continuum. One drawback of this strategy is that diffusion rate values may not correspond to those values that would have been obtained by intermixing pair comparison sessions with E-J session.

Method

Since both experimenter-controlled expanded judgement task and then a traditional pair comparison task were administered to each subject the stimulus, design and procedures sections will be explained in two phases.

Phase 1: The Experimenter-Controlled Expanded Judgement Task

Subjects. Three subjects (honours students) participated for pay according to a payoff schedule discussed below in the section entitled 'Payoff Schedule'.

Apparatus. See Experiment 2.

Stimuli and Design. *1. Selection of the standard according to an infinite set paradigm and construction of the comparison mean.* Past research has shown that the traditional method of constant stimuli which exposes the observer to repeated instances of the same perceptual stimuli, provides the subject with the opportunity to gather information about the stimuli in long-term memory (Palmer, 1990). In fact, subjects may store long-term memory information about stimuli, even if the standard is randomly selected from a range of values with a non-uniform and symmetric probability distribution. Over several trials, a Gaussian distribution could conceivably allow the subject to use the value of the expectation of the distribution in order to encode stimulus information in long-term memory (Leth-Steensen, 1994). Ordinal semantic codes could be substituted for analogue information thus precluding the strict perceptual analysis of stimuli and its associated processing and representational properties (Banks, 1977; Banks et al., 1982). In order to avoid such a confound, Experiment 3 made use of an infinite-set paradigm whereby, on

that enter into a judgement. In the long run $E(s_j) = 200$, the median of the uniformly distributed range. Lengths range from 100 to 300 pixels (38.28mm, 114.84mm). As a result, $E(s_j + ks_j) = (190, 210)$ where $k = \pm 0.05$.

This experiment was experimenter-controlled to ensure control over the number of observations and the amount of time between observations. Although a subject-controlled experiment would have permitted examination of the impact of diffusion on the number of observations selected, the lack of control of inter-stimulus latencies would have precluded an analysis of the effect of elapsed time since stimulus acquisition. The number of observations presented was either 3, 6 or 9 based on the range of observations requested by the same subjects in a subject-controlled version of the same task.

2. *Variability of the comparison distribution.* Variability was used as a within subjects factor to produce two different variance-based a priori difficulty levels. On low variability trials $\sigma = 10$ which results in a minimum standardised difference of $z = \frac{100 - 95}{10} = 0.50$ when $s_j = 100$, and a maximum difference of $z = \frac{300 - 285}{10} = 1.5$ when $s_j = 300$. Overall a priori difficulty for the low variability condition is therefore defined by uniformly distributed z scores irrespective of sign that range from 0.50 to 1.50 with a mean value of 1.00. On high variability trials, $\sigma = 10$ was increased threefold to result in $m\sigma = 3 \times 10 = 30$. As a result, the a priori defined difficulty indices for the high variability manipulation ranged from .1667 to .50 with a mean of .33.

3. *Type of trial.* Half of the trials consisted of a standard presented for 2000 ms followed immediately by the first observation of the comparison distribution. The other half consisted of trials with a standard that was present throughout until the primary

response was made. Both types of trials were randomly intermixed within each block of trials.

4. Number of trials. In each block, the base number of trials was determined by counterbalancing two standard-comparison percent differences (k) with two instructions, two standard type conditions (perceptual/memory), two σ 's and three sample sizes (3, 6 and 9) for a subtotal of $2 \times 2 \times 2 \times 2 \times 3 = 48$ base pairs. These stimuli were repeated twice for a total of 96 trials for each of the three blocks that comprised each of the 1 1/2 hour session. Observers 1 and 2 did nine sessions each for a total of 2592 trials. For Observers 1 and 2, each of the 12 perception/memory by σ by sample size cell was based on 216 replicates. Observer 3 did 10 sessions for a total of 2880 trials. For Observers 3, each of the 12 perception/memory by σ by sample size cell was based on 240 replicates. Each session was preceded by 12 practice trials.

Payoff Schedule. A priori importance of the primary decision in general and of the number of observations in particular was determined through monetary gains and losses associated with correct and incorrect responses (see Experiments 1 and 2 for rationale). Experiment 3 involved actual monetary gains and losses.

A point system was used to represent monetary values for each possible outcome. Correct primary responses were worth 50 points. Incorrect responses cost 50 points. Actual subject pay was based on the conversion of these points into monetary values. Due to budgetary restrictions, the intention was to keep subject pay at approximately \$15.00 per session. The basic win/loss amount was 5 cents based on pilot data (Aubin,

1994). The observer was advised that they would make an average of \$15.00 per hour, and that, although unlikely, could lose money if they made too many incorrect responses.

Procedure. Preliminary procedures were the same as those used in Experiments 1. Task procedures were the same as those used in Experiment 2, except that the task was experimenter-controlled, and subjects were not required to provide confidence ratings. The instructions follow.

"In a traditional experiment, one of two words ('SHORTER' or 'LONGER') appears as an instruction at the top of the computer screen followed a moment later by two horizontal lines positioned one above the other [write example on blackboard]. In that kind of task you would be asked to select the longer or the shorter of the two lines. This experiment is slightly different. In this case, the line at the bottom is not actually presented. Instead it serves as the true average of a very large population of lines that differ in length. Your task is to compare the stationary line with the true average of the population of lines and decide which is longer or which is shorter depending on the instruction.

Not all of the population lines will be presented. Rather, a smaller sample of lines will be randomly selected from the entire population and presented one after the other below the stationary line. Using the average of the sample as your best guess for the population average, you will then decide which is longer or shorter, the population average or the stationary line. The population lines will be presented one after another in approximately the same spot just below the stationary line. The centre of the comparison lines will be randomly staggered with respect to the stationary line. As a result, you will

need to compare the entire line lengths not just their edges in order to be correct. The number of lines selected from the population is under the computer's control and will vary randomly from trial to trial. A beep will sound when the last comparison line is presented after which you should indicate your decision.

The message you see on the screen, 'press any key to continue' will recur several times during the course of the session in order to give you breaks. The message can be cleared to start the experiment by pressing any letter or number on the keyboard using your non-preferred hand. Otherwise, all responses are made using these the buttons on this response panel. I will ask you to use the index finger of your preferred for indicating all your responses.

Each trial begins by pressing this middle key with your index finger [the middle button sits in a three-button configuration arranged in the shape of an inverted triangle]. Once you have done so, keep your index finger over the middle button until you have made your decision regarding the stimuli presented on the screen. After pressing the middle button, one of two instructions (SHORTER or LONGER) will be presented at the top of computer screen and remain there until you end the trial by making a primary response. Two seconds after the instruction has appeared, a stationary line will be presented under the instruction. On some trials the stationary line will remain on the screen until the end of the trial. On other trials, and unpredictably, the stationary line will disappear after 2 sec just before the comparison lines are presented. In either trial type, comparison lines are presented 2 seconds after the initial presentation of the stationary line at a rate of three per second.

Once you have made your judgement, and using your index finger, indicate your decision by pressing the left button if the stationary line is consistent with the instruction (shorter/longer), or the right button if the population average is consistent with the instruction. The screen will then be blank waiting for you to press the middle button again to start the next trial.

Accuracy is very important in this experiment because it will determine whether you win or lose money on each trial. A point system represents money. A correct decision wins you 50 points. An error costs you 50 points. Points will be translated into money. You can make on average about \$15 dollars per session. You could make more or less depending on your performance on each trial. You will be advised of the amount won after each session. Although unlikely the possibility exists that you could lose money if you make too many errors.

It is very important that you maintain a high level of accuracy throughout the session. As a result, two types of breaks are included. Experimental trials are divided into three blocks. After each block, the screen will read 'REST' and then 'press any key to continue'. At REST, I would like you to physically get up and leave the room even if for just a moment. Feel free to take a short walk to stretch out your legs. After returning to the testing room, close the door completely and give your eyes a minute to adapt to the change in lighting before you continue. You will get two major breaks. I would also like you to take momentary breaks every few trials by delaying pressing the middle button to start each trial."

Phase 2: The traditional pair comparison task

Subjects. Same three subjects used for phase 1 with same payoff schedule.

Apparatus. See Experiment 2.

Stimuli and Procedures. This task was administered after the observers had completed the expanded judgement sessions to help preclude storing of stimulus information in long-term memory about the distribution of line lengths. Stimuli were those used in the Phase 1 expanded judgement task except that the comparison stimuli were actually presented as traditional pair comparisons. The line stimuli were successively presented one above the other permitting the manipulation of ISIs. The centres of each line within a pair were randomly offset from each other to avoid end-point strategies. Five ISI's were used (150, 240, 360 and 480 ms). Each ISI was presented 24 times within each block for a total of 96 trials per block. Three blocks (288 trials) were given for each of four sessions per subject for a total of 1152 trials for subjects 1 and 2. Subject 3 could only do two sessions (576 trials). Eight practice trials were given. Subjects were given the usual instructions emphasising accuracy but without taking too long to respond.

Regressing the variance estimates for the study stimuli on ISI provided an estimate of the diffusion rate which according to Kinchla & Smyzer (1967) should be a positive linear function (see Endnote 8 for equations). This diffusion rate was then be used in equation 4 to obtain estimates of the memory and perceptual variances for the expanded judgement task.

Results

Diffusion estimates using pair comparisons. After removing the first block of each session to control for practice effects, subjects 1 and 2 had 192 replicates per ISI, and subject 3 had 96 replicates for each ISI. All trials less than 200 ms and greater than 10000 ms were excluded from analyses. This criterion resulted in the exclusion of .4% of total trials for Observer 1, 3.7% for Observer 2, and no exclusions for Observer 3..

Figure 18 provides ISI effects on response time and response accuracy for of the three observers. The most important finding is that accuracy decreases monotonically as ISI increases for Observers 1 and 2. Subject 3 can be excluded from further analysis since these data cannot be fit by any model. This latter individual's high level of accuracy regardless of ISI suggests persistence of the perceptual information even after the study stimulus has been removed for nearly half a second. Observers 1 and 2 exhibit what appears to be a linear decrease in accuracy. Note that overall response times increase with ISI suggesting that decision are becoming more difficult in tandem with decreased accuracy. Diffusion theory provides a source of this difficulty by suggesting that the representation of the study stimulus becomes more variable as ISI increases.

Figure 19 illustrates changes in variance estimates of the study stimulus as ISI increases. It is clear that the diffusion model is rejected on the basis of subject 2 whose variance estimates increase in a positively accelerated fashion. Even in the absence of interfering stimuli, the variability of the study stimulus does not increase linearly as predicted by the diffusion model but rather exhibits positive acceleration.

Table 6

Diffusion Parameter Estimates for Each Subject

Subject	σ_0^2	ϕ	R^2
1	-79.1	1.382	0.919
2	-3969.6	20.840	0.762

Figure 18. The effect of ISI on accuracy and response times for each of the three observers.

Although diffusion theory is disconfirmed, parameter estimates were obtained for comparison with an alternative model, decay theory. Table 6 provides parameter estimates for the diffusion model for each of the subjects. The variance of the study stimulus after acquisition is represented by σ_0^2 and the diffusion rate is ϕ .

The decay model suggests that forgetting increases geometrically over time i.e. a fixed proportion of stored information is forgotten each unit of time (Broadbent and Gregory, 1965; Massaro, 1970b). Specifically, $d' = \alpha\phi^t$, where d' is familiarity (discriminability), α represents stored information concerning the study stimulus, ϕ is a value between 0 and 1 representing the rate of forgetting, and t is time. When $\phi=1$, there is no forgetting and d' is simply α . Regardless of the value of ϕ , at $t=0$, the end of stimulus acquisition, no forgetting has yet occurred since $\phi^0 = 1$. Forgetting is greater for small values of ϕ . The simplest method for obtaining ϕ is to express d' on a semilogarithmic scale such that the log of d' changes linearly with the variable t , $\log d' = \log \alpha + t \log \phi$. The rate of forgetting can be provided by first regressing values of $\log d'$ against t , and then applying an exponential function to $\log \phi$. The amount of stimulus information stored at the end of acquisition can be estimated by setting $t=0$ and then applying an exponential transformation to $\log \alpha$.

A comparison of Table 7 with Table 6 shows that the exponential decay model results in improved fits especially for subject 2. Equation (4) for the expanded judgement diffusion model can be altered to reflect the inverse relation between phi as a diffusion

rate and phi as the decay rate:

$$\sigma_0^2 + \frac{1}{\phi} (n \times \bar{\alpha}) + \frac{1}{n} \sigma_\mu^2 = \left(\frac{-y}{D_n} \right) = \sigma_{y_n}^2.$$

Table 7

According to the Decay Model of Forgetting: the Amount of Stored Stimulus Information (α), the Decay Rate (ϕ) and the Corresponding Diffusion Rate ($1/\phi$) for Each Subject

Subject	σ_0^2	α	ϕ	$\left(\frac{1}{\phi}\right)$	R^2
1	84.87	1.0855	.998	1.002	0.952
2	45.29	1.4859	.995	1.005	0.967

This model provides variance estimates of the study stimulus according to an expanded judgement decay model.

Predicted and observed stimulus variances using diffusion and decay models.

Despite the obvious failures of diffusion, decay and interference theories, observed and predicted stimulus variances according to diffusion and decay theories are provided in figures 20 and 21 for illustration. Predictions for interference theory were not provided since no interference was evident in the expanded judgement data.

For subject 1 in figure 20, both diffusion and decay theories predict approximately the same optimal sample size because the rate of forgetting obtained from the pair comparison data was approximately linear across ISIs. Optimal sample size should however be larger when comparison variability is high because under those conditions, the standard error term in a Thurstonian model is differentially reduced by sample size. The observed curves indicate that optimal sample size is actually much smaller than that predicted under high comparison variability. The fit is better under low comparison variability. The apparently rapid increase in observed stimulus variance as n increases is actually relatively meaningless since accuracy is actually only dropping by only 4% from $n=3$ to $n=9$. Neither model fit the data well.

For subject 2 in figure 21, the fit of the data to the diffusion model is deceptively good. Recall, however, that changes in stimulus variance across ISI were actually quadratic so that fitting a straight line through the data resulted in an inflated diffusion

rate with a bad fit ($R^2=.762$). The exponential decay function resulted in a fit of $R^2=.967$ and indicated that forgetting was actually less than that suggested by diffusion. For this subject, only a proportion of .005 of the original stimulus is forgotten during each time interval. Again the predictions by diffusion can be discounted. Decay theory fares quite well, but again the problem is that there are no appreciable decrements in remembered expanded judgements as n increases.

The accuracy and latency of decisions in the expanded judgement task. An examination of accuracy and response time data for the expanded judgement task for each subject reveals all current models of forgetting fail for expanded judgement. As is clearly evident from the individual subject data in Figure 22 (low comparison variability) and in Figure 23 (high comparison variability), accuracy is lower for memory than for perception in replication of previous empirical findings on differences between judgements with perceptual and remembered stimuli; However, the number of observations presented had little if any impact on accuracy in either of the conditions.

Post-summative latencies are particularly informative. The decrease in post-sampling times as sample size increases suggests that observers are learning rather than forgetting. These results are difficult to understand given that an infinite set paradigm was used. This paradigm precludes the long-term memory storage of stimulus information that could conceivable result from repeated exposure to the same small set of standards. Recall that the standard was randomly selected from a range of 200 possible line lengths. Faster post-summative latencies for memory compared to perception is consistent with findings in Experiment 2 where the O controlled the number of comparison observations. In that experiment, however, faster post-summative latencies for remembered judgements

High Comparison Variance

Low Comparison Variance

Figure 20. Observed and predicted variances of the study stimulus for subject 1.

High Comparison Variance

Low Comparison Variance

Figure 21. Observed and predicted variances of the study stimulus for subject 2.

were paralleled by fewer requested observations suggesting, in conjunction with a massive distance effect, that an evidence accrual-like comparison process took place after observations had been accumulated. In Experiment 3, however, the number of observations was strictly controlled by the E to ensure the same sample sizes between memory and perception. Since the evidence accrual process may be occurring post-decisionally, and the number of explicit observations was constant between memory and perception, it is difficult to understand why post-summative latencies were shorter for memory than for perception. Evidence accrual models predict the opposite effect.

Figure 22. The effect of sample size on accuracy (left panels) and post-summativ latencies (right panels) for perceptual (unfilled symbols) under low comparison variability.

Figure 23. The effect of sample size on accuracy (left panels) and post-summative latencies (right panels) for perceptual (unfilled symbols) under high comparison variability.

Discussion

Both diffusion and decay theories predict a decrease in accuracy over time regardless of the number of intervening stimuli. Without rehearsal, the remembered representation passively decays over time. Diffusion predicts linear forgetting over time. Decay theory predicts exponential forgetting. In the case of Massaro's interference theory, however, the predicted exponential increase in forgetting should only occur because of the nature and number of stimuli that intervene between stimulus acquisition and response, not because of a passive process. According to interference theory, the effect of intervening stimuli may be more critically dependent on the nature of the stimuli involved (Massaro, 1969; Posner, Boies, Eichelman & Taylor, 1969). The greater the similarity between the study stimulus and the intervening stimuli, the greater the interference. In this experiment, diffusion theory was refuted because increases in stimulus variances over time were better described by an exponential distribution. Leth-Steensen (1994) found equivocal support for diffusion using pair-comparisons with variable set-sizes.

According to interference theory, the similarity between the explicit intervening observations and the study stimulus used in the expanded judgement task should have led to competition between the two possible responses. This competition would reflect the active disruption of the remembered stimulus due to the storage of new stimulus information. This large expected decrease in performance for memory judgements over sample size did not emerge. Note that since sample size was experimentally controlled, observers should have exhibited an increase in response accuracy with increasing sample size in the perceptual condition. This monotonic increase in accuracy with sample size would have paralleled the finding in psychophysical experiments that make use of

increasingly liberal response deadlines (Petrusic, 1992). This effect predicted on the basis of LPFs from slow-fast guessing (Petrusic, 1992) and from accumulator theories (Vickers, 1979) has been clearly demonstrated in other expanded judgement tasks with cards (Irwin, Smith & Mayfield, 1956) and with line lengths (Vickers, Burt, Smith, et al., 1985).

In this experiment, however, perceptual judgements were highly accurate under low comparison variability. An accuracy of approximately 90% suggests that the task may not have been sufficiently difficult. However, memory judgements were substantially harder than the perceptual judgements. Consider that observations were presented at a rate of three per second. After 1 second presentation time, in addition to at least one second of post-summative latencies, performance accuracy for memory judgements dropped by nearly 20% percentage points compared to perception. Accuracy was reduced to approximately 70% in the memory condition under low comparison variability, and to the low 70% range for both memory and perception under high comparison variability. Yet, accuracy failed to change in relation to sample size. These results stand in contrast to Vickers et al. (1985) who, in their experimenter-controlled expanded judgement task with lines, found that the improvement in accuracy as a function of sample size when the stimuli were highly discriminable (approximately 94%) was magnified when the task was difficult (approximately 75%).

As a post-hoc explanation, the presentation of comparison stimuli may be promoting rehearsal of the study stimuli because the observer is forced to retrieve the study stimuli from memory for a comparison to occur. As a result, the intervening stimuli are really considered part of a single task and not as a secondary task that promotes division of

attention. The decrease in post-summative latencies as sample size increases suggests that increased stimulus information is making the decision easier. This interpretation is consistent with both decay and diffusion theories but inconsistent with interference theory which states that no forgetting will occur unless interfering stimuli are presented. In a non-expanded judgement pair comparison task, no opportunity exists for rehearsal during an empty ISI interval. In contrast, experiments promoting interference theory include intervening stimuli that compete with the main task for attentional resources by require the subject to perform a competing task. This expanded judgement task, which requires the observer to consider each intervening stimulus as an integral part of a single comparison process, may indeed refresh the representation of the study stimulus by promoting rehearsal. If such a situation is true, then the study stimulus is protected against forgetting. An experiment with a wider range of sample sizes would be required to examine the time course of forgetting when rehearsal is involved. Perhaps a comparison condition could be devised in which the observers would be instructed to consider the same comparison stimuli as interfering stimuli rather than as part of the comparison process. In that condition, a comparison would be made between the first stimulus and the last. Subjects would have to tend to some property of the intervening stimuli as a secondary task, perhaps counting the number of observations.

Overall Summary

The current set of experiments demonstrated a number of parallels between the expanded judgement paradigm and the method of constant stimuli. Both show the same effects of memory on the discriminability and on the variability of a subjective magnitude. Remembered physical stimuli are less discriminable and more variable than their perceptual counterparts. They also display an enhancement of discriminability for end points compared to the inner standards through a serial-position effect that is absent for perceptual standards. Paralleling the behaviour of response times in the psychophysical literature, sampling times and post-summative latencies displayed serial-position effects demonstrating the greater ease of processing for remembered end-point standards compared to the inner remembered standards. No such trend occurs in perception. These findings suggest that a subject-controlled expanded judgement paradigm provides a viable means of psychophysical investigation.

The second set of issues dealt with the viability of the expanded judgement as a means of exploring the accrual-based nature of comparative decisions of stimuli on the length continuum. Previous research (Irwin et al., 1956; Irwin & Smith, 1956, 1957; Vickers et al., 1985) has shown results consistent with response time predictions made by evidence accrual theories (Petrusic, 1992; Vickers, 1970, 1979). In these subject-controlled expanded judgement experiments, the number of observations requested increased as the distance between the comparison values decreased (Irwin et al., 1956; Irwin & Smith, 1956, 1957; Vickers et al., 1985), and as the variability of the information increased (Irwin et al., 1956; Irwin & Smith, 1956, 1957). The current subject-controlled experiments also demonstrated a distance effect and an effect of comparison variability.

These results seem to suggest that explicit observations are like implicit accruals.

Nonetheless, as we now summarize, a number of other findings suggest that such is not the case.

Latencies between observations should have displayed a distance effect if in fact each observation is compared to the standard in pre-decisional moments. A distance effect did not emerge until the end of the sampling process. In addition, if decisional processing takes place during observation sampling, post-sampling latencies should reflect constant response-output time. In fact, the largest distance effect emerged during the post-sampling phase suggesting that the bulk of the comparative process takes place after sampling has stopped. In addition, post-sampling latencies were highly positively correlated with the number of explicit observations. Vickers et al. (1985) also found that post summative latencies increased as the number of requested observations increased.

Taken together, these results suggest that observations inputted during the sampling process are recollected in an evidence accrual-like process after sampling has stopped. Explicit observations may therefore place restrictions on decision time but do not, in and of themselves, function as implicit accruals. No reasonable explanation is available at this point for the isolated presence of a distance effect for perception under low comparison variability only. Certainly, the addition of variability from the comparison distribution significantly impacted on the form of the serial-position curve in memory. Under high comparison variability, the end-point standards could not benefit from distinctiveness. In a more general sense, these results may suggest that psychophysical measurement cannot occur beyond a certain amount of error.

The effect of putting a stimulus in memory was actually opposite to that observed in psychophysical experiments and to that predicted by accrual theories. Less observations were requested for memory judgements than for perceptual. However, it was also found that sampling times were negatively correlated with the number of samples. Although the overall number of observations was greater for high comparison variability than for low, sample times were also negatively related to the number of observations. These results emphasise the importance of examining not only the number of observations but also the speed with which they are sampled. In order to reach optimal discriminability, the speed with which observations are sampled must be traded against the number of observations requested. Faster sampling results in less information input and therefore the need for more observations. This suggests that in determining response times, accrual-theories should consider both sampling time and the number of implicit accruals. Expanded judgement suggests that when sampling times are slowed, the number of observations is also decreased.

Finally, diffusion theory was rejected for these data because changes in the variance of the study stimulus were better described by an exponential decay function. In the expanded judgement task, however, and contrary to predictions, performance accuracy for memory judgements did not decrease with an increase in sample size. This suggests that despite the similarity of the intervening stimuli to the study stimulus, interference did not occur. Rather, observers used the comparison observations to refresh the representation of the study stimulus and therefore maintained performance across sample size.

References

- Algom, D. (1992). Memory psychophysics: An examination of its perceptual and cognitive prospects. In D. Algom (Ed.), *Psychophysical approaches to cognition*. Amsterdam, The Netherlands: Elsevier Science Publishers.
- Algom, D., Wolf, Y., & Bergman, B. (1985). Integration of stimulus dimensions in perception and memory. *Journal of Experimental Psychology: General*, **114**, 451-471.
- Allan, L. G., & Kristofferson, A. B. (1974). Psychophysical theories of durations discrimination. *Perception & Psychophysics*, **16**, 26-34.
- Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review*, **77**(3), 153-170.
- Anderson, N. H. (1976). How functional measurement can yield validated interval scales of mental quantities. *Journal of Applied Psychology*, **61**(6), 677-692.
- Ashby, F. G. (1983). A biased random walk model of two choice reaction times. *Journal of Mathematical Psychology*, **27**, 277-297.
- Aubin, P. (1994). Preliminary pilot data on subject-controlled and experimenter-controlled expanded comparative judgements. Unpublished data. Department of Psychology, Carleton University.
- Audley, R. J. (1960). A stochastic model for individual choice behaviour. *Psychological Review*, **67**, 1-15.
- Audley, R. J., & Pike, A. R. (1965). Some alternative stochastic models of choice. *British Journal of Mathematical and Statistical Psychology*, **18**, 207-225.

- Audley, R. J., & Wallis, C. P. (1964). Response instructions and the speed of relative judgments. I. Some experiments on brightness discrimination. *British Journal of Psychology*, **55**, 59-73.
- Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgement. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11). New York: Academic Press.
- Banks, W. P., Clark, H. H., & Lucy, P. (1975). The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance*, **1**, 35-47.
- Banks, W. P., & Flora, J. (1977). Semantic and perceptual processes in symbolic comparison. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 278-290.
- Banks, W. P., Mermelstein, R., & Yu, H. K. (1982). Discriminations among perceptual and symbolic stimuli. *Memory & Cognition*, **10**(3), 265-278.
- Banks, W. P., & Root, M. (1979). Semantic congruity effects in judgments of loudness. *Perception and Psychophysics*, **26**, 133-142.
- Baranski, J. V., & Petrusic, W. M. (1992). The discriminability of remembered magnitudes. *Memory & Cognition*, **20**(3), 254-270.
- Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*, **55**, 412-428.
- Bjorkman, M., Lundberg, U., and Tarnbloom, S. (1960). On the relation between percept and memory: A psychophysical approach. *Scandinavian Journal of Psychology*, **1**, 136-144.

- Block, H. D., & Marschark, J. (1960). Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics*. Stanford: Stanford University Press.
- Boring, E. G. (1917). A chart of the psychometric function. *American Journal of Psychology*, **28**, 465-470.
- Broadbent, D. E., & Gregory, M. (1965). On the interaction of S-R compatibility with other variables affecting reaction time. *British Journal of Psychology*, **56**, 61-67.
- Busmeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **11**(3), 538-564.
- Cartwright, D., & Festinger, L. (1943). A quantitative theory of decision. *Psychological Review*, **50**, 595-621.
- Cattell, J. M. (1893). On errors of observation. *American Journal of Psychology*, **5**, 285-293.
- Cattell, J. M. (1902). The time of perception as a measure of differences in intensity. *Philosophische Studien*, **19**, 63-68.
- Cech, C. G., & Shoben, E. J. (1985). Context effects in symbolic magnitude comparisons. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 299-315.
- Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, **57**, 145-158.
- Coombs, C. G. (1964). *Theory of data*. New York: Wiley.

- Coombs, C. H., Greenberg, M. G., & Zinnes, J. L. (1968) A double law of comparative judgement for the analysis of preferential choice and similarities data. *Psychometrika*, **26**, 165-171.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, **11**, 671-684.
- Crossman, E. R. F. W. (1955). The measurement of discriminability. *Quarterly Journal of Experimental Psychology*, **7**, 176-195.
- Dember, W. N. (1957). The relation of decision time to similarity. *Journal of Experimental Psychology*, **53**, 68-72.
- Desrochers, A., & Petrusic, W. M. (1984). Comprehension effects in comparative judgments. In J. C. Yuille (Ed.), *Imagery, memory, and cognition* (pp. 131-159). Hillsdale, NJ: Erlbaum.
- Duncan, E. M., & MacFarland, C. E., Jr. (1980). Isolating the effects of symbol distance and semantic congruity in comparative judgments: An additive factors analysis. *Memory & Cognition*, **8**, 612-622.
- Edwards, W. (1954a). Probability preferences among bets with differing expected values. *American Journal of Psychology*, **67**, 56-67.
- Edwards, W. (1954b). The theory of decision making. *Psychological Bulletin*, **51**, 380-417.
- Edwards, W. (1961). Costs and payoffs are instructions. *Psychological Review*, **68**(4), 275-284.
- Edwards, W., & Tversky, A. (1967). *Decision making*. Harmondsworth: Penguin.

- Geller, E. S., & Pitz, G. F. (1968). Confidence and decision speed in the revision of opinion. *Organizational Behaviour and Human Performance*, **3**, 190-201.
- Gescheider, G. A., Wright, J. H., & Evans, M. B. (1968). Reaction time in the detection of vibrotactile signals. *Journal of Experimental Psychology*, **77**, 501-504.
- Gescheider, G. A., Wright, J. H., Weber, B. J., Kirchner, B. M., & Milligan, E. A. (1969). Reaction time as a function of the intensity and probability of occurrence of vibrotactile signals. *Perception & Psychophysics*, **10**, 413-417.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Guilford, J. P. (1954). *Psychometric methods*. (2nd ed.). New York: McGraw-Hill.
- Harrison, D. A., & Petrusic, W. M. (1994). Time-order errors and memory for percepts. Presented at the annual conference for Brain, Behaviour and Cognitive Sciences, June, 1994, Vancouver, B.C., Canada.
- Hays, W. L. (1994). *Statistics*. 5th ed. Fort Worth: Harcourt Brace.
- Hegelmaier, F. (1852). Über des Gedächtnis für Linearanschauungen. *Archiv für Physiologisch Heilkunde*, **11**, 844-853.
- Hellström, A. (1977). Time errors are perceptual: An experimental investigation of duration and a quantitative successive comparison model. *Psychological Research*, **39**, 345-388.
- Hellström, A. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 460-477.
- Hellström, A. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, **97**, 35-61.

- Helson, H. (1947). *Adaptation-level theory*. New York: Harper Row.
- Henderson, J. M., & Wells, A. D. (1985). Symbolic comparisons with and without perceptual referents: Is interval information used? *Memory & Cognition*, **13**, 176-182.
- Henmon, V. A. C. (1906). The time of perception as a measure of differences in sensation. *Archives of Philosophy, Psychology and Scientific Methodology*, **8**, 1-75.
- Henmon, V. A. C. (1911). The relation of the time of a judgement to its accuracy. *Psychological Review*, **18**, 186-201.
- Hilgard, E. R. (1951). Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 517-567). New York: John Wiley & Sons.
- Hollingworth, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology & Scientific Methods*, **7**, 461-469.
- Holyoak, K. J., & Patterson, K. K. (1981). A positional discriminability model of linear-order judgments. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 1283-1302.
- Irwin, F. W., & Smith, A. S. (1956). Further tests of theories of decision in an "expanded judgment" situation. *Journal of Experimental Psychology*, **52**(6), 345-348.
- Irwin, F. W., & Smith, A. S. (1957). Value, cost, and information as determinants of decision. *Journal of Experimental Psychology*, **54**(3), 229-232.
- Irwin, F. W., Smith, A. S., & Mayfield, J. F. (1956). Tests of two theories of decision in an "expanded judgment" situation. *Journal of Experimental Psychology*, **51**(4), 261-268.

- Jamieson, D. G., & Petrusic, W. M. (1975a). The dependence of time order error direction on stimulus range. *Canadian Journal of Psychology*, **29**, 175-182.
- Jamieson, D. G., & Petrusic, W. M. (1975b). Presentation order effects in duration discrimination. *Perception & Psychophysics*, **17**, 197-202.
- Jamieson, D. G., & Petrusic, W. M. (1975c). Relational judgments with remembered stimuli. *Perception & Psychophysics*, **18**(6), 373-378.
- Jamieson, D. G., & Petrusic, W. M. (1976). On a bias induced by the provision of feedback in psychophysical experiments. *Acta Psychologica*, **40**, 199-206.
- Johnson, D. M. (1939). Confidence and speed in the two-category judgment. *Archives of Psychology*, **241**, 1-52.
- Johnson, G. H. (1991). A distinctiveness model of serial learning. *Psychological Review*, **98**, 204-217.
- Kellogg, W. N. (1931). Time of judgment in psychometric measures. *American Journal of Psychology*, **43**, 65-86.
- Kennedy, R. (1990). Psychometric functions with remembered standards: Weber fractions and serial position effects. Thesis, Master of Arts, Psychology Department, Carleton University.
- Kerst, S. M., & Howard, J. H. (1978). Memory psychophysics for visual area and length. *Memory & Cognition*, **6**, 327-335.
- Kinchla, R. A., & Allan, L. G. (1969). A theory of visual movement perception. *Psychological Review*, **76**, 537-558.
- Kinchla, R. A., & Smyzer, F. (1967). A diffusion model of perception memory. *Perception & Psychophysics*, **2**, 219-229.

- Köhler, W. (1923). Zur Analyse des Sukzessivvergleichs und der Zeitfehler [On the analysis of successive comparison and of time error.] *Psychologische Forschung*, **4**, 115-175.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement (Vol. I)*. New York: Academic Press.
- Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral and Brain Sciences*, **12**, 251-320.
- Kruskal, J. B. (1964). Non-metric multidimensional scaling: A numerical method. *Psychometrika*, **29**, 115-129.
- La Berge, D. (1962). A recruitment theory of simple behavior. *Psychometrika*, **27**, 375-396.
- Laming, D. (1968). *Information theory of choice reaction times*. New York: Academic press.
- Laming, D., & Scweiwiller, P. (1985). Retention in perceptual memory: A review of models and data. *Perception & Psychophysics*, **2**, 219-229.
- Leach, L. (1992). Memory psychophysics: Set size and the Weber fraction. Thesis, B. A. with Honours, Psychology Department, Carleton University.
- Leth-Steensen, C. (1994). Memory load and the rate of forgetting of unidimensional perceptual stimuli. Masters Thesis. Department of Psychology, Carleton University.
- Lindsay, P. H., Taylor, M. M., & Forbes, S. M. (1968). Attention and multidimensional discrimination. *Perception & Psychophysics*, **4**, 113-117.

- Link, S. W. (1975). The relative judgment theory of two-choice response times. *Journal of Mathematical Psychology*, **12**, 114-135.
- Link, S. W. (1978). The relative judgment theory of the psychometric function. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum.
- Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, N. J.: Lawrence Erlbaum.
- Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77-105.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Luce, R. D., & Galanter, E. (1963). Discrimination. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology: Vol. I*. New York: Wiley.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- Manz, W. (1970). Experiments on probabilistic information processing. *Acta Psychologica*, **34**, 184-200.
- Marschark, M., & Paivio, A. (1981). Congruity and the perceptual comparison task. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 290-308.
- Massaro, D., W. (1970a). Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, **83**(1), 32-39.
- Massaro, D., W. (1970b). Forgetting: Interference or decay. *Journal of Experimental Psychology*, **83**(2), 238-243.

- Massaro, D. W. (1970c). Consolidation and interference in the perceptual memory system. *Perception & Psychophysics*, **7**(3), 153-156.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.
- Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics*, **13**, 180-184.
- Moyer, R. S., & Bayer, R. H. (1976). Mental comparisons and the symbolic distance effect. *Cognitive Psychology*, **8**, 228-246.
- Moyer, R. S., & Dumais, S. T. (1978). Mental comparison. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 117-155). New York: Academic Press.
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, **215**, 1519-1520.
- Moyer, R. S., Sklarew, and Whiting (1982). Memory psychophysics. In H. G. Geissler and P. Petzold (Eds.), *Psychophysical judgement and the process of perception*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Munsterberg, H. (1894). Studies from the Harvard psychological laboratory: A psychometric investigation of the psycho-physic law. *Psychological Review*, **1**, 45-51.
- Murdock, B. B., Jr. (1960). The distinctiveness of stimuli. *Psychological Review*, **67**, 16-31.
- Murray, D. J. (1993). A perspective for viewing the history of psychophysics. *Behavioral and Brain Sciences*, **16**, 115-186.

- Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited process. *Cognitive Psychology*, **7**, 44-64.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 332-350.
- Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, **58**, 193-198.
- Petrusic, W. M. (1984). Explicit counting and time order errors in duration discrimination. *Timing and Time Perception*, **423**, 630-633.
- Petrusic, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance*, **18**(4), 962-986.
- Petrusic, W. M. (1993). Response time based psychophysics: An added perspective. *Behavioral and Brain Sciences*, **16**(1), 158-159.
- Petrusic, W. M., & Aubin, P. (1996). Similarity comparisons and memory psychophysics: Dember revisited. In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 335-357). Washington, DC: American Psychological Association.
- Petrusic, W. M., & Baranski, J. V. (1989). Semantic congruity effects in perceptual comparisons. *Perception & Psychophysics*, **45**(4), 439-452.

- Petrusic, W. M., Baranski, J. V., & Aubin, P. (1995). Analogue and propositional coding memory psychophysics. Paper presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles, CA, November, 1993.
- Petrusic, W. M., Baranski, J. V., & Harrison, D. (1993). Memory psychophysics: Context and acquisition effects. Paper presented at the 34th Annual Meeting of the Psychonomic Society, Washington, D. C., November 5th, 1993.
- Petrusic, W. M. (1996). Memory psychophysics and fundamental measurement: Similarity comparisons with remembered and perceived magnitudes. Oral presentation given at the Symposium on Context Effects In Comparative Judgments at the Midwestern Psychological Association Meeting, Chicago, Illinois, May 6th, 1995.
- Petrusic, W. M., & Jamieson, D. (1976). Studying preferential choice without stable probability estimates: Temporal transitivity analyses. *Acta Psychologica*, **40**, 375-383.
- Petrusic, W. M., & Jamieson, D. (1978). Relation between probability of preferential choice and time to choose changes with practice. *Journal of Experimental Psychology: Human Perception and Performance*, **4**(3), 471-482.
- Petrusic, W. M., & Jamieson, D. (1979). Resolution time and the coding of arithmetic relations on supraliminally different visual extents. *Journal of Mathematical Psychology*, **19**, 89-107.
- Pike, A., R. (1968). Latency and relative frequency of response in psychophysical discrimination. *Journal of Mathematical and Statistical Psychology*, **21**(2), 161-182.

- Pike, A. R. (1971). The latencies of correct and incorrect responses in discrimination and detection tasks: Their interpretation in terms of a model based on simple counting. *Perception & Psychophysics*, **9**, 455-460.
- Pike, A. R. (1973). Response latency models for signal detection. *Psychological Review*, **80**, 53-68.
- Posner, M. I. (1967). Characteristics of visual and kinesthetic memory codes. *Journal of Experimental Psychology*, **75**, 103-107.
- Posner, M. I., & Konick, A. F. (1966). Short-term retention of visual and kinesthetic information. *Organizational Behaviour and Human Performance*, **1**, 71-86.
- Ratcliffe, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
- Ratcliffe, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, **88**, 552-572.
- Ratcliffe, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review*, **92**, 212-225.
- Shepard, R., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of stares. *Cognitive Psychology*, **1**, 1-17.
- Shipley, W., C., Norris, E. D., & Roberts, N. L. (1945). The effect of changed polarity of set on decision time of affective judgments. *Journal of Experimental Psychology*, **36**, 237-243.
- Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice reaction. *Journal of Mathematical Psychology*, **32**, 135-168.

- Smith, P. L., & Vickers, D. (1989). Modelling evidence accumulation with partial loss in expanded judgment. *Journal of Experimental Psychology: Human Perception and Performance*, **15**(4), 797-815.
- Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**(3), 153-179.
- Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, **25**, 251-260.
- Suppes, P., & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush and Galanter (Eds.), *Handbook of mathematical psychology: Vol 1* (pp. 1-76). New York: Wiley.
- Swensson, R. G. (1972). The elusive tradeoff: Speed vs accuracy in visual discrimination tasks. *Perception and Psychophysics*, **12**, 16-32.
- Swets, J. A., Shipley, E. F., McKey, M. J., & Green, D. M. (1959). Multiple observations of signals in noise. *Journal of the Acoustical Society of America*, **31**, 514-521.
- Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, **61**, 401-409.
- Thomas, E. A. C. (1971). Sufficient conditions for monotone hazard rate: An application to latency-probability curves. *Journal of Mathematical Psychology*, **8**, 303-332.
- Thomson, G. H. (1920). A new point of view in the interpretation of thresholds of measurements in psychophysics. *Psychological Review*, **27**, 300-307.
- Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review*, **34**, 237-286.
- Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, **13**, 37-58.

- Vickers, D. (1978). An adaptive module for simple judgments. In J. Requin (Ed.), *Attention and performance VII* (pp. 599-618). Hillsdale, N.J.: Lawrence Erlbaum.
- Vickers, D. (1979). *Decision processes in visual perception*. London: Academic Press.
- Vickers, D. (1980). Discrimination. In A. T. Welford (Ed.), *Reaction times* (pp.25-71). Toronto: Academic Press.
- Vickers, D., Burt, J., Smith, P., & Brown, M. (1985). Experimental paradigms emphasizing state or process limitations: I. Effects on speed-accuracy tradeoffs. *Acta Psychologica*, **59**, 129-161.
- Vickers, D., Foreman, E. A., Nicholls, M. E. R., Innes, N. J., & Gott, R. E. (1989). Some experimental tests of the application of an interval of uncertainty model to a time-limited expanded judgment task. In D. Vickers and P. L. Smith (Eds.), *Human information processing: Measures, mechanisms and models* (pp. 253-265. North-Holland: Elsevier Science Publishers.
- Vickers, D., & Smith, P. (1985). Accumulator and random-walk models of psychophysical discrimination: A counter-evaluation. *Perception*, **14**, 471-497.
- Vickers, D., Smith, P., Burt, J., & Brown, M. (1985). Experimental paradigms emphasizing state or process limitations: II. Effects on confidence. *Acta Psychologica*, **59**, 163-193.
- Vickers, D., & Packer, J. S. (1982). Effects of alternating set for speed or accuracy on response time., accuracy, and confidence in a unidimensional discrimination task. *Acta Psychologica*, **50**, 179-197.
- Wallis, C. P., & Audley, R. J. (1964). Response instructions and the speed of relative judgments. II. Pitch discriminations. *British Journal of Psychology*, **55**, 133-142.

- Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, **3**, 189-230.
- Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, **5**, 102-122.
- Wickelgren, W. A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, **6**, 13-61.
- Wolf, Y., & Algom, D. (1987). Perceptual and memorial constructs in children's judgments of quantity: A law of cross-representation invariance. *Journal of Experimental Psychology: General*, **116**, 381-397.

Endnotes

1. Decision time plays a crucial role in differentiating between decision theories. Known as the 'second law of psychophysics' (Kellogg, 1931), the distance effect states that the time to complete a discrimination decreases as the physical difference between two stimuli increases (Henmon, 1906; Johnson, 1939; Kellogg, 1931; Munsterberg, 1894; Kellogg, 1931) even for supraliminal stimuli (Petrusic, 1993). As reviewed by Vickers (1980) and as demonstrated by Petrusic and Jamieson (1979), the distance effect can be determined by a number of stimulus characteristics. The additional assumption of latency functions allows strength theories, more specifically signal detection theory (Tanner & Swets, 1954), to account for the effects of discriminability and bias by predicting that response time will decrease as the distance between the observation and the criterion increases (Gescheider, Wright & Evans, 1968; Gescheider, Wright, Weber, Kirchner, & Milligan, 1969; Pike, 1973). Response times will be longer when discriminability is low because observations are more densely clustered around the criterion. The predominance of short observation-criterion distances will result in long overall response times. When the task is easier, distributions are farther apart and distances are farther from the criterion. Latency functions will shift as bias is manipulated due to changes in observations-criterion distances and to changes in response probability (Thomas, 1971). Response times for less probable observations located to the right of a conservative criterion are slower than those for the more probable observations on the left (Espinoza-Varas & Watson, 1994).

2. In replication and extension of Munsterger (1894), Petrusic & Jamieson (1979) showed response times for comparisons with errorless supraliminal stimuli are sensitive

to ratios as well as differences. They used stimuli equally spaced on a log scale such that while adjacent pairs have a common ratio their absolute differences increase and were thus able to show that, if one defines ratios of the larger stimulus to be smaller stimulus, judgements become increasingly difficult and response times slow as the physical ratio approaches 1.00. Petrusic (1993) provides a succinct comment on the much overlooked role of response time in psychophysics.

3. Similarity judgements involving stimuli ordered on an underlying physical continuum must result in an ordering of intervals and, in addition, these intervals must behave such that the concatenation of intervals will result in equivalent intervals as on a single dimension throughout the set of intervals. Krantz et al. (1971) describe the axiomatisation of algebraic-difference structures as requiring two critical testable axioms and two technical but non-testable axioms. Very few studies have made use of similarity-dissimilarity judgements and even less have examined the decision processes involved. Dember (1957) extended Coomb's (1964) theory of data and found evidence for a distance effect where response times decreased as the differences of differences increased (ordinal level). Unfortunately, Dember did not examine quads of the form (ac,bd) and as a result, IDS was not assessed. Petrusic (1995) report two experiments investigating similarity-dissimilarity judgements with perceptual and memory stimuli. Both studies used line lengths, although line lengths in Experiment 1 were vertical while those in Experiment 2 were horizontal. On perceptual trials, subjects chose the 'MORE SIMILAR' of two pairs of line lengths. On the other half, the 'MORE DISSIMILAR' pair was chosen. Memory trials were conducted in exactly the same way, except that stimuli were labels learned during paired-associate training. In both studies, weak

stochastic transitivity was satisfied both in memory and in perception showing that subjects can in fact order intervals in the form of quads whether in perception or in memory. The ordering of pairs based on stochastic dominance matched predictions on the basis of sensitivity to differences. In Experiment 1, ratios of adjacent pairs were held constant but the absolute differences increased predicting that discrimination should become easier as pair members increase in magnitude. And in fact, the order of pairs based on stochastic dominance bears this out. A semantic coding model might criticise Experiment 1 by suggesting that participants implicitly knew that differences increased as stimuli increased in magnitude and, therefore rendered not a similarity-dissimilarity judgement but, rather, a categorisation judgement which required identifying the pair with largest stimulus value as the most dissimilar. However, this strategy would have required the unlikely knowledge of the continuum. Experiment 2 corrected for this possibility by irregularly spacing the stimuli and found exactly the same effects as Experiment 1.

4. Functional Measurement (Anderson, 1970, 1976) is an alternative approach applied to two-dimensional stimuli. In this case each dimension is treated as a factor and each magnitude is a level of that factor e.g. in an experiment which requires subjects to rate individual squares using category scaling techniques, each rectangle has two dimensions and each dimension has several possible values. Treated as a two-way factorial design, the dimensional properties of subjective category ratings are submitted to an analysis of variance to assess the existence of a subjective scale. Significant main effects or interactions prove the existence of subjective scale values at an interval level of measurement and permit a characterisation of the rule by which two-dimensional stimuli

are constructed. Using functional measurement, several authors have outlined some of the crucial properties of perceptual and remembered stimuli (Algom, Wolf & Bergman, 1985; Wolf & Algom, 1987).

5. The two critical testable axioms are a weak ordering and weak monotonicity (Block & Marschark, 1960; Krantz, Luce, Suppes & Tversky, 1971). A ***weak ordering*** guarantees an ordinal scale of measurement and requires 1) comparability (the participant can render a judgement: either $abScd$ or $cdSab$, where $abScd$ denotes that the pair (a,b) is more similar than the pair (c,d)); and 2) transitivity which when satisfied for all triples of pairs (a,b) , (c,d) , (e,f) exhibits the following relations: if $abScd$, and $cdSef$, then $abSef$. The more crucial ***weak monotonicity condition*** is a necessary axiom for and algebraic-difference structures and guarantees an interval level of measurement. This condition ensures the homogeneous concatenation of intervals on the underlying continuum and asserts that for all elements a,b,c,a',b' and c' in the set A :

if $abSa'b'$
 and $bcSb'c'$
 then $acSa'c'$.

In certain situations where not enough stimuli are available, *intradimensional subtractivity* (IDS) can be substituted for weak monotonicity because IDS implies weak monotonicity (Block & Marschark, 1960; Suppes & Zinnes, 1963). For all ordered subsets of stimuli a, b, c, d such that $a < b < c < d$, IDS exists if the following conditions are met:

if $abScd$
 then $acSbd$.

This is the case because the segment (b,c) is common both to (a,c) and to (b,d) . Weak monotonicity implies $acSbd$, since if $abSbc$ and $bcScd$, as monotonicity dictates, then $acSbd$. IDS follows i.e. $abScd$ if and only if $acSbd$. One should note however, that while monotonicity guarantees IDS, the opposite is not necessarily true. One usually tests for all three axioms, transitivity, IDS and monotonicity.

In addition, two technical but non-testable axioms must hold. These include *continuity* (solvability) and the *Archimedean condition*: i) Continuity asserts that intervals are sufficiently dense so that any interval can be 'copied' into any larger interval i.e. intervals can be infinitely small. Support for this more crucial of the non-testable axioms would prove that the memory representation holds information that exists on a continuum analogous to the physical and perceptual stimuli; ii) The Archimedean principle states that no magnitude is so small or so large that it cannot be approximated by a concatenation of smaller magnitudes. It supports the existence of a finite standard

dominance is thus substituted for the strict form of the axioms. Only weak stochastic transitivity (WST) is required to satisfy the axiom of *weak ordering* e.g.

$$\begin{aligned} &\text{if} && P(abScd) > .50 \\ &\text{and} && P(cdSef) > .50 \\ &\text{then} && P(abSef) > .50, \end{aligned}$$

where ' $P(abScd) > .50$ ' denotes that more than half the time pair (a,b) is judged more similar than the pair (c,d). Stochastic IDS for all (a,b) and (c,d) is specified as:

$$P(abScd) > .50 \text{ if and only if } P(acSbc) > .50.$$

6. Response times for the quads with adjacent and disjoint pairs, which are more similar, were faster with the instruction 'MORE SIMILAR' than the instruction 'MORE DISSIMILAR'. Conversely, it took less time to select the "MORE DISSIMILAR" than the 'MORE SIMILAR' of the overlapping pairs which tend to be more dissimilar.

7. Given the linear relation between response time and the differences of differences, an examination of the quads involved in IDS should be consistent with a subtractive decision rule. According to this view, RTs for similarity-dissimilarity judgements should be a monotone decreasing function of the difference of the differences whether perceived or remembered i.e. $RT(abScd) = F[d(a,b) - d(c,d)]$, where $d(a,d)$ denotes the distance between the subjective values corresponding to stimuli a and b, i.e. $d(a,b)=s(a)-s(b)$, where $s(x)$ is the subjective value. F is a strictly monotone decreasing function of the difference of similarities. As an example, response times to the disjoint quad, $\{(1,2),(3,4)\}$, and the overlapping quad, $\{(1,3),(2,4)\}$, should be the same because the interval between stimuli 2 and 3 is common to both quads i.e. $d(1,3)-d(2,4) = [d(1,2)+d(2,3)]-[d(2,3)+d(3,4)] = d(1,2)+d(2,3)-d(2,3)-d(3,4)=d(1,2)-d(3,4)$. The basis for

the similarity judgement is pair (1,2) and (3,4) for both quads since the interval (2,3) is simply subtracted from the overlapping pair. The difference of differences model appeared to be rejected, however, since RTs for the overlapping quads were consistently slower than RTs for the disjoint quads. The RT advantage for the disjoint pairs over the overlapping was greater in memory than in perception.

8. Diffusion theory was originally based on a signal detection model (SDT: Green & Swets, 1966) of the input and decisional processes in order to obtain an estimate of discriminability that is independent of bias effects on performance accuracy. The noise stimulus, s_n , and the signal+noise stimulus, s_{sn} give rise to two normal distributions of subjective values, x_n and x_{sn} , respectively, with equal variances, $\sigma_{x_n}^2 = \sigma_{x_{sn}}^2$, and with means equal to the actual physical stimulus values i.e. $\mu_{x_n} = s_n$ and $\mu_{x_{sn}} = s_{sn}$. As a result, sensitivity is proportional to the absolute difference between the two physical values expressed in units of the standard deviation of the differences between the subjective distributions. According to signal detection theory (SDT) then

$$d' = \frac{|s_n - s_s|}{\sqrt{\sigma_{x_n}^2 + \sigma_{x_{sn}}^2}}.$$

Although not of primary concern in diffusion, STD also posits a

decisional process whereby a response criterion partitions the sensory space into two decisional events, 'Yes' there is a signal and 'No' there is no signal. Diffusion theory adds a third process applicable to cases where the time lapsed between the successive presentation of two stimuli leads to a reduction in sensitivity. Although originally conceptualising diffusion as a random walk process where increases and decreases in the study stimulus are distributed as a Poisson distribution, Kinchla & Smyzer (1967) made the simplifying assumption that the Gaussian approximation can be used if p, the

probability of an increase in value is approximately .5 and the number of steps in the random walk is not too small. An initial formulation of the diffusion theory sensitivity, can be conceptualised as

$$\delta_t = \frac{|x_t - m_t|}{\sqrt{\sigma_{x_t}^2 + \sigma_{m_t}^2}} \quad (1)$$

According to this view, the numerator becomes $|x_t - m_t|$, where x_t is the second perceptually available stimulus presented at time t , and m_t is the study stimulus also at time t . Since m_t is the original value of the perceptual study stimulus after it has undergone some amount of diffusion, then $m_t = x_o + d_t$, where x_o is the value of the perceptual study stimulus at the end of acquisition, thus $t=0$, and d_t represents the effect of diffusion that cumulates up to time t . Since diffusion follows a symmetric random walk process (see Kinchla & Smyzer (1967) for a full mathematical description), then $E(d_t) = 0$. As a result, $E(m_t) = E(x_o) + E(d_t) = E(x_o)$. Since $E(x_o) = s_o$ and $E(x_t) = s_t$, then the difference between the means of the subjective distributions is simply the difference between the study and comparison stimuli, $|s_o - s_t| = y_t$.

Accordingly, diffusion has its effect not on the mean value of the stimulus, but on the variability of the underlying distribution of the study stimulus. Consider from equation (1) that $\sigma_{m_t}^2 = \sigma_{x_o}^2 + \sigma_{d_t}^2$ because of the additivity of variances, and that $\sigma_{d_t}^2 = \phi t$ which is the rate at which variance of d_t changes as a function of ISI. The standard error of the differences in equation (1) becomes $(\sigma_{x_t}^2 + \sigma_{x_o}^2 + \phi t)^{\frac{1}{2}}$. After substitution, squaring and rearranging terms, equation (1) becomes,

$$(\sigma_{x_i}^2 + \sigma_{x_o}^2) + \phi t = \frac{y_t}{\delta_t} \quad (2)$$

The slope of this linear function provides the diffusion rate, ϕ , when the estimate of the variance of the differences is plotted against time, t . When $t = 0$, the intercept, $(\sigma_{x_i}^2 + \sigma_{x_o}^2)$, provides an estimate of the difference variance when both compared stimuli are perceptually available and is therefore equivalent to the denominator of the SDT sensitivity parameter, d' . In such a case, $\delta_t = d'$. Predicted total variance of the differences at time t can be easily obtained since y_t is known and δ_t is estimated by transforming the subjects' hit and false alarm rates using $z(hits) - z(fa)$ (Macmillan & Creelman, 1991).

Appendix

Individual Differences in the serial-position effect for the slopes of the psychometric functions in Experiment 2

An examination of individual differences showed that three of the five subjects (8, 10 and 14) showed clear SPE's under low comparison variability (figure 1). Two of the three exhibited large effects while the third had an attenuated effect. Both of the subjects that showed no SPE exhibited uniformly high discriminability across all four standards. Under high comparison variability (figure 2), two of the five subjects showed attenuated SPE effects. The overall SPE for the latter condition was therefore masked by the poorer performance of the three subjects and the overall negative effect of increasing comparison variability on the precision of comparative judgements.

Low Comparison Variability

Figure 1. Psychometric slopes for each individual subject in the low comparison variability conditions

High Comparison Variability

Figure 2. Psychometric slopes for each individual subject in the low comparison variability conditions

IMAGE EVALUATION TEST TARGET (QA-3)

APPLIED IMAGE, Inc.
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved